

Published in final edited form as:

Phys Med Biol. 2012 March 21; 57(6): 1459–1476. doi:10.1088/0031-9155/57/6/1459.

Joint L^1 and Total Variation Regularization for Fluorescence Molecular Tomography

Joyita Dutta^{1,2}, Sangtae Ahn¹, Changqing Li³, Simon R. Cherry³, and Richard M. Leahy¹

Joyita Dutta: Dutta.Joyita@mgh.harvard.edu; Richard M. Leahy: leahy@sipi.usc.edu

¹Signal and Image Processing Institute, Department of Electrical Engineering -Systems, University of Southern California, Los Angeles, CA 90089, USA

²Division of Nuclear Medicine & Molecular Imaging, Massachusetts General Hospital, Boston, MA 02114, USA

³Department of Biomedical Engineering, University of California, Davis, Davis, CA 95616, USA

Abstract

Fluorescence molecular tomography (FMT) is an imaging modality that exploits the specificity of fluorescent biomarkers to enable 3D visualization of molecular targets and pathways *in vivo* in small animals. Owing to the high degree of absorption and scattering of light through tissue, the FMT inverse problem is inherently illconditioned making image reconstruction highly susceptible to the effects of noise and numerical errors. Appropriate priors or penalties are needed to facilitate reconstruction and to restrict the search space to a specific solution set. Typically, fluorescent probes are locally concentrated within specific areas of interest (e.g., inside tumors). The commonly used L^2 norm penalty generates the minimum energy solution, which tends to be spread out in space. Instead, we present here an approach involving a combination of the L^1 and total variation norm penalties, the former to suppress spurious background signals and enforce sparsity and the latter to preserve local smoothness and piecewise constancy in the reconstructed images. We have developed a surrogate-based optimization method for minimizing the joint penalties. The method was validated using both simulated and experimental data obtained from a mouse-shaped phantom mimicking tissue optical properties and containing two embedded fluorescent sources. Fluorescence data was collected using a 3D FMT setup that uses an EMCCD camera for image acquisition and a conical mirror for full-surface viewing. A range of performance metrics were utilized to evaluate our simulation results and to compare our method with the L^1 , L^2 , and total variation norm penalty based approaches. The experimental results were assessed using Dice similarity coefficients computed after co-registration with a CT image of the phantom.

1. Introduction

Inverse operators for ill-posed (or ill-conditioned) problems tend to be unbounded (or have a very large norm). Regularization is the use of bounded (or low-norm) approximations for such inverse operators that allow us to generate meaningful and numerically stable solutions (Daubechies et al. 2004). This typically involves incorporating some prior knowledge usually in the form of a penalty function that controls some desired property of an unknown deterministic variable or the prior probability distribution of an unknown random variable. In the context of image reconstruction, for example, penalty functions are often chosen to enforce smoothness (Fessler and Hero 1995), preserve edges (Lange 1990, Strong and Chan 2003), promote sparsity (Mohajerani et al. 2007, Cao et al. 2007, Gao and Zhao 2010a), or incorporate anatomical information (Leahy and Yan 1991, Fessler et al. 1992, Hyde et al. 2010).

The most widespread method for handling the ill-conditioning of inverse problems is Tikhonov regularization (Tikhonov and Arsenin 1977), in which the cost function contains an L^2 norm penalty term in addition to the data-fitting term. For the linear system $\mathbf{Ax} = \mathbf{b}$, with system matrix \mathbf{A} , unknown vector \mathbf{x} , and data vector \mathbf{b} , the Tikhonov regularized solution is the minimizer of the cost function:

$$\Phi(\mathbf{x}) = \frac{1}{2} \|\mathbf{Ax} - \mathbf{b}\|_2^2 + \frac{\lambda}{2} \|\Gamma \mathbf{x}\|_2^2, \quad (1)$$

where Γ is an appropriately chosen Tikhonov matrix. In the simplest case, in which $\Gamma = \mathbf{I}$, this regularization technique favors solutions with lower L^2 norms. The Bayesian implication of Tikhonov regularization is the assumption of a multivariate Gaussian probability distribution on the unknown random vector.

The goal of fluorescence molecular tomography (FMT) is to compute the 3D distribution of a fluorescent source inside a volume from the photon density detected on the surface (Dutta et al. 2008, Weissleder and Ntziachristos 2003, Eppstein et al. 2002). The fluorescent source may be a near infrared (NIR) fluorescent dye, an active or activatable fluorescent biomarker, or a fluorescent protein expressed by a reporter gene (Shu et al. 2009, Massoud and Gambhir 2003). FMT is a promising low-cost molecular imaging modality that does not use any ionizing radiation and exploits the availability of a variety of highly specific fluorescent molecular markers. Yet, this technique is confounded by the high degree of absorption and scattering of photons propagating through tissue, making the FMT problem ill-conditioned. Approaches for alleviating this problem and improving source localization include multispectral illumination and/or detection (which exploit the spectral variation of tissue optical properties) (Gardner et al. 2010, Li et al. 2009, Chaudhari et al. 2009, Zacharakis et al. 2005) and the use of multiple spatial patterns of illumination (Dutta et al. 2010, Bélanger et al. 2010, Dutta et al. 2009, Joshi et al. 2006) to increase the information content in the collected data. While these approaches improve the conditioning of the FMT system matrix, image reconstruction continues to be highly susceptible to the effects of noise and numerical errors, necessitating the use of some kind of regularization.

While the simple Tikhonov regularizer generates moderately satisfactory results when applied to FMT reconstruction, we would like to design an improved regularizer that specifically takes into account prior knowledge about FMT images. A variety of regularizers have been reported in the fields of FMT, diffuse optical tomography, and bioluminescence tomography (Hyde et al. 2010, Pogue et al. 1999, Axelsson et al. 2007, Li, Boverman, Zhang, Brooks, Miller, Kilmer, Zhang, Hillman & Boas 2005, Guven et al. 2005). We focus on the fact that NIR probes used in FMT are designed to preferentially accumulate in specific areas of interest, e.g., tumors or cancerous tissue. Therefore, reconstructed FMT images commonly exhibit fluorophore concentrations localized within tumors and the major excretory organs (Bloch et al. 2005, Weissleder et al. 1999). Consequently these images tend to be very sparse with some locally smooth high intensity regions. This inspires us to investigate a combination of regularizers that enforce conditions of sparsity and smoothness on reconstructed FMT images.

The most fundamental sparsity metric is the L^0 norm, which is the total number of non-zero elements in a vector. However, the L^0 norm-penalized inverse problem, at least in the underdetermined case, is NP-hard (Natarajan 1995). Instead, the L^1 norm, which is a convex relaxation of the L^0 norm, is often used to enforce sparsity in images and is particularly popular in the field of compressed sensing (Donoho 2006, Candes and Wakin 2008). The corresponding optimization problem can be formulated in a number of equivalent ways

including basis pursuit, where the L^1 norm appears in the cost function and the L^2 data-fitting term appears in the constraint (Chen et al. 1998), least absolute shrinkage and selection operator (commonly referred to as LASSO), where the L^2 data-fitting term appears in the cost function and the L^1 norm appears in the constraint (Tibshirani 1996), and L^1 -penalized least squares, where the L^1 norm penalty weighed by a regularization parameter is added to the L^2 data-fitting term to construct the cost function. Approaches to solve these problems include pivoting algorithms (Efron et al. 2004), interior-point methods (Kim et al. 2007), and gradient-based techniques (Daubechies et al. 2004, Figueiredo et al. 2007, Beck and Teboulle 2009). In this work, we choose the L^1 penalty for enforcing sparsity in FMT images.

A variety of smoothing priors have been applied to tomographic reconstruction (Hebert and Leahy 1989, Green1990). Amongst these, priors involving quadratic penalties are particularly common, since they are easy to handle. However, they tend to smooth out edges in images. We, therefore, choose to use the total variation (TV) penalty, which promotes smoothness while preserving edges in images (Rudin et al. 1992). In this work, we define the total variation as the L^1 norm of the differences between neighboring pixels. This particular form of the TV penalty enforces sparsity on pixel differences and consequently tends to generate images with piecewise constant regions and sharp boundaries. A variety of methods have been employed for handling the TV penalty, including dual-based and interior point approaches (Chambolle 2004, Vogel and Oman 1998, Huang et al. 2008, Gao and Zhao 2010b). In this work, we derive a method based on the separable paraboloidal surrogates (SPS) algorithm for minimizing the TV penalty.

Section 2 describes a compound approach that uses a combination of the SPS method with the preconditioned conjugate gradient (PCG) algorithm for handling the joint L^1 and TV penalties. We use ordered subsets to accelerate the SPS approach. In section 3, we describe an experimental setup based on a mouse-shaped phantom for testing the joint penalties. We validate our method by applying it to both simulated and experimental data. A discussion and analysis of our simulation and experimental results is presented in section 4.

2. Methods

The cost function we seek to minimize contains three parts – a data-fitting term, a sparsifying penalty term, and a smoothing penalty term. As mentioned before, the sparsifying penalty under consideration is the L^1 norm of the unknown image \mathbf{x} . For smoothing the 3D image \mathbf{x} , we penalize its total variation, defined as follows:

$$TV(\mathbf{x}) = \sum_k |x_{m(k)} - x_{n(k)}|, \quad (2)$$

where $m(k)$ and $n(k)$ are pixel indices corresponding to the k th neighboring pixel pair in the 3D image \mathbf{x} . The resulting optimization problem is as follows:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x} \geq 0} \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda_{L1} \|\mathbf{x}\|_1 + \lambda_{TV} \sum_k |x_{m(k)} - x_{n(k)}|. \quad (3)$$

Here λ_{L1} and λ_{TV} are regularization parameters.

In FMT, the non-negativity constraint on the unknown vector helps us circumvent any complications due to the non-differentiability of the L^1 norm penalty near zero. Let us consider two functions $f_1(x) = g(x) + \|x\|_1$ and $f_2(x) = g(x) + x$. Then, $\min_{x \geq 0} f_1(x) = \min_{x \geq 0}$

$f_2(x)$. This allows us to simplify the L^1 norm penalty to a linear term. To handle the non-differentiability of the TV penalty, we approximate it as follows:

$$TV(\mathbf{x}) \approx \Phi^{TV}(\mathbf{x}) \triangleq \sum_k \sqrt{[\mathbf{C}\mathbf{x}]_k^2 + \delta_{TV}}, \quad (4)$$

where the transformation $\mathbf{v} = \mathbf{C}\mathbf{x}$ computes nearest neighbor differences and $\mathbf{C} \in \mathbb{R}^{n_m \times n_s}$, n_s being the number of pixels in \mathbf{x} and n_m being the number of neighboring pixel pairs. The parameter δ_{TV} tends to round off sharp edges when large but leads to instability as it approaches machine precision (Vogel and Oman 1996, Chan et al. 1996). We set it to a fixed value of 10^{-9} here. The modified cost function is:

$$\Phi(\mathbf{x}) = \frac{1}{2} \|\mathbf{A}\mathbf{x} - \mathbf{b}\|_2^2 + \lambda_{L1} \mathbf{1}' \mathbf{x} + \lambda_{TV} \sum_k \sqrt{(\mathbf{x}_{m(k)} - \mathbf{x}_{n(k)})^2 + \delta_{TV}}, \quad (5)$$

where $\mathbf{1}$ is an $n_s \times 1$ vector of all ones. In the subsequent subsections, we describe two approaches for minimizing this cost function.

2.1. Preconditioned conjugate gradient algorithm

The preconditioned conjugate gradient (PCG) method (Bertsekas 1999) can be used to minimize the cost function in (5). This method uses the gradient given by:

$$\nabla \Phi(\mathbf{x}) = \mathbf{A}'(\mathbf{A}\mathbf{x} - \mathbf{b}) + \lambda_{L1} \mathbf{1} + \lambda_{TV} \mathbf{C}'((\mathbf{C}\mathbf{x}) \circ \mathbf{z}(\mathbf{C}\mathbf{x})). \quad (6)$$

Here the prime ($'$) symbol represents the transpose of a matrix, and (\circ) represents the Hadamard (entrywise) matrix product. We define the vector function $\mathbf{z}(\cdot)$ as:

$$\mathbf{z}(v) = [\zeta(v_1) \ \zeta(v_2) \ \dots \ \zeta(v_m)]', \quad (7)$$

where the notation $\zeta(v)$ is used to represent the scalar function:

$$\zeta(v) = \frac{1}{\sqrt{v^2 + \delta_T}}. \quad (8)$$

To accelerate convergence, we use the reciprocal of the diagonal terms of the Hessian of the data-fitting term, $\mathbf{A}'\mathbf{A}$, for preconditioning. The step size is determined using an Armijo line search (Bertsekas 1999). To enforce non-negativity, every iterate is projected onto the non-negative orthant. This is achieved by computing a feasible direction within the non-negative orthant and an optimal step-size along that direction.

2.2. Separable paraboloidal surrogates algorithm

While the PCG algorithm is straightforward to implement, the computational time per iteration is increased by the expensive line search. As an alternative, we explore a different approach based on the optimization transfer principle, also known as majorization-minimization (MM) (Lange et al. 2000). In every iteration, we generate and minimize a surrogate function which satisfies a set of majorization conditions that guarantee monotonicity. We design surrogates that are separable using the SPS approach. Separable

surrogate functions are easier to minimize and eliminate the need for an expensive line search.

2.2.1. Surrogate function properties—Optimization transfer allows us to replace a cost function $\Phi(\mathbf{x})$ that is difficult to minimize by a surrogate function $\phi(\mathbf{x}; \mathbf{x}^n)$ that is constructed at every iterate, \mathbf{x}^n , and is easier to minimize. The resulting update equation is:

$$\mathbf{x}^{n+1} = \arg \min_{\mathbf{x} \in D} \phi(\mathbf{x}; \mathbf{x}^n). \quad (9)$$

This surrogate function must satisfy the following majorization conditions over the domain, D , of \mathbf{x} :

$$\phi(\mathbf{x}^n; \mathbf{x}^n) = \Phi(\mathbf{x}^n) \quad (10)$$

$$\phi(\mathbf{x}; \mathbf{x}^n) \geq \Phi(\mathbf{x}), \quad \forall \mathbf{x} \in D \quad (11)$$

$$\nabla^{\top} \phi(\mathbf{x}; \mathbf{x}^n) |_{\mathbf{x}=\mathbf{x}^n} = \nabla \Phi(\mathbf{x}) |_{\mathbf{x}=\mathbf{x}^n}, \quad (12)$$

where ∇^{\top} is the column gradient operator with respect to the first argument. Using (10) and (11), we can show that:

$$\Phi(\mathbf{x}^{n+1}) - \Phi(\mathbf{x}^n) \leq \phi(\mathbf{x}^{n+1}; \mathbf{x}^n) - \phi(\mathbf{x}^n; \mathbf{x}^n) \leq 0, \quad \forall \mathbf{x} \in D. \quad (13)$$

Equation (13) ensures that the update in (9) allows $\Phi(\mathbf{x})$ to decrease monotonically.

2.2.2. SPS for the data-fitting term—We first construct a surrogate function that is convex, separable, and tangential to the data-fitting cost function and lies above it. We may write the data-fitting term as (Erdogan and Fessler 1999):

$$\Phi^{DF}(\mathbf{x}) = \frac{1}{2} (\mathbf{Ax} - \mathbf{b})' (\mathbf{Ax} - \mathbf{b}) = \frac{1}{2} \sum_i f_i([\mathbf{Ax}]_i), \quad (14)$$

where

$$f_i([\mathbf{Ax}]_i) \triangleq ([\mathbf{Ax}]_i - b_i)^2. \quad (15)$$

The term $[\mathbf{Ax}]_i$ can also be written as:

$$[\mathbf{Ax}]_i = \sum_j a_{ij} x_j = \sum_j a_{ij} (x_j - x_j^n + x_j^n) = \sum_k a_{ik} x_k^n \sum_j \alpha_{ij} + \sum_j \alpha_{ij} \frac{a_{ij}}{\alpha_{ij}} (x_j - x_j^n) = \sum_j \alpha_{ij} \left(\frac{a_{ij}}{\alpha_{ij}} (x_j - x_j^n) + [\mathbf{Ax}^n]_i \right).$$

The α_{ij} 's are constants introduced here to facilitate the computation of the surrogate and, by definition, satisfy $\sum_j \alpha_{ij} = 1$. By using this constraint on the α_{ij} 's to exploit the convexity of function f_i , we can compute a function that lies above this function:

$$f_i([\mathbf{A}\mathbf{x}]_i) \leq \sum_j \alpha_{ij} f_i \left(\frac{\alpha_{ij}}{\alpha_{ij}} (x_j - x_j^n) + [\mathbf{A}\mathbf{x}^n]_i \right). \quad (17)$$

We pick $a_{ij} = \alpha_{ij} / \sum_k a_{ik}$. This gives us our separable surrogate for the data-fit term:

$$\phi^{DF}(\mathbf{x}; \mathbf{x}^n) = \frac{1}{2} \sum_i \left(\sum_j \frac{a_{ij}}{\sum_k a_{ik}} f_i \left(\frac{a_{ij}}{\alpha_{ij}} (x_j - x_j^n) + [\mathbf{A}\mathbf{x}^n]_i \right) \right). \quad (18)$$

It can be shown that, at the n th iterate, the gradient (which is the same as that for the original function) and Hessian for this surrogate function are as follows:

$$\nabla^{10} \phi^{DF}(\mathbf{x}^n; \mathbf{x}^n) = \mathbf{A}' (\mathbf{A}\mathbf{x}^n - \mathbf{b}) \quad (19)$$

$$\nabla^{20} \phi^{DF}(\mathbf{x}^n; \mathbf{x}^n) = \text{diag}_j \left[\|\mathbf{A}' \mathbf{A} \mathbf{e}^j\|_1 \right], \quad (20)$$

where \mathbf{e}^j is a unit vector with 1 as the j th element and ∇^{20} is the Hessian operator with respect to the first argument.

2.2.3. SPS for the total variation penalty term—The SPS function for the TV penalty is derived in two steps. We first approximate the total variation penalty in (4) by $\sum_k \psi^n(v_k)$, where $v_k = [\mathbf{C}\mathbf{x}]_k = x_{m(k)} - x_{n(k)}$ and $\psi^n(v)$ is a non-separable paraboloidal surrogate function satisfying (10)-(12):

$$\psi^n(v) = \frac{1}{2} \zeta(v^n) v^2 + \frac{1}{2} \zeta(v^n) \left((v^n)^2 + 2\delta_{TV} \right). \quad (21)$$

The function $\zeta(\cdot)$ used here was defined in (8). We then use the convexity of the function $\psi^n(v)$ to derive a separable surrogate as follows (Lange and Fessler 1995):

$$\begin{aligned} & \psi^n(x_{m(k)} - x_{n(k)}) \\ &= \psi^n \left(\frac{1}{2} \left[2x_{m(k)} - x_{m(k)}^n - x_{n(k)}^n \right] + \frac{1}{2} \left[-2x_{n(k)} + x_{m(k)}^n + x_{n(k)}^n \right] \right) \leq \frac{1}{2} \psi^n \left(2x_{m(k)} - x_{m(k)}^n - x_{n(k)}^n \right) \\ & \quad + \frac{1}{2} \psi^n \left(2x_{n(k)} - x_{m(k)}^n - x_{n(k)}^n \right). \end{aligned}$$

The resultant SPS function for the TV penalty is:

$$\phi^{TV}(\mathbf{x}; \mathbf{x}^n) = \sum_k \frac{1}{2} \psi^n \left(2x_{m(k)} - x_{m(k)}^n - x_{n(k)}^n \right) + \sum_k \frac{1}{2} \psi^n \left(2x_{n(k)} - x_{m(k)}^n - x_{n(k)}^n \right). \quad (23)$$

It can be shown that, at the n th iterate, the gradient (which is the same as that for the original function) and Hessian for this surrogate function are as follows:

$$\nabla^{10} \phi^{TV}(\mathbf{x}^n; \mathbf{x}^n) = \mathbf{C}' \left((\mathbf{C}\mathbf{x}^n) \circ z(\mathbf{C}\mathbf{x}^n) \right) \quad (24)$$

$$\nabla^{20} \phi^{TV}(\mathbf{x}^n; \mathbf{x}^n) = \text{diag}_j \left[2(|\mathbf{C}|' z(\mathbf{C}\mathbf{x}^n))_j \right]. \quad (25)$$

Here, $|\mathbf{C}| \in \mathbb{R}^{n_d \times n_s}$ represents a matrix consisting of the absolute values of the elements of \mathbf{C} and should not be confused with the determinant of \mathbf{C} .

2.2.4. Update equation—We obtain the overall surrogate function at an iterate \mathbf{x}^n by replacing the data-fitting and TV terms in the original cost function in (5) by the corresponding SPS functions:

$$\phi(\mathbf{x}; \mathbf{x}^n) = \phi^{DF}(\mathbf{x}; \mathbf{x}^n) + \lambda_{l_1} \|\mathbf{x}\|_1 + \lambda_{TV} \phi^{TV}(\mathbf{x}; \mathbf{x}^n). \quad (26)$$

Owing to the separable nature of this surrogate, we can easily compute its minimizer, \mathbf{x}^{n+1} , over the non-negative orthant in a closed form:

$$\mathbf{x}^{n+1} = \left[\mathbf{x}^n - \mathbf{D}^{-1}(\mathbf{x}^n) \nabla \Phi(\mathbf{x}^n) \right]_+. \quad (27)$$

Here, the notation $[\cdot]_+$ represents projection onto the non-negative orthant, while $\mathbf{D}(\mathbf{x})$, computed using (20) and (25), is given by:

$$\mathbf{D}(\mathbf{x}) = \text{diag}_j \left[\|\mathbf{A}' \mathbf{A} \mathbf{e}^j\|_1 + 2\lambda_{TV} (|\mathbf{C}|' z(\mathbf{C}\mathbf{x}))_j \right]. \quad (28)$$

As per the underlying condition (12), the gradient of this surrogate function can be computed using (6), the original gradient equation.

2.2.5. Ordered subsets implementation—Since the FMT inverse problem typically uses very large data sets, the SPS method can be accelerated using an ordered subsets (OS) (also known as the incremental gradient) approach (Ahn and Fessler 2003, Bertsekas 1999, Erdogan and Fessler 1999). The FMT system matrix is typically very tall and of the form:

$$\mathbf{A} = \begin{bmatrix} \mathbf{A}^{em} \mathbf{D}_1^{ex} \\ \mathbf{A}^{em} \mathbf{D}_2^{ex} \\ \vdots \\ \mathbf{A}^{em} \mathbf{D}_p^{ex} \end{bmatrix}. \quad (29)$$

Here $\mathbf{A}^{em} \in \mathbb{R}^{n_d \times n_s}$ is the emission forward model matrix, where n_d is the number of surface detector nodes and n_s is the number of point source locations distributed inside the volume, and $\mathbf{D}_k^{ex} \in \mathbb{R}^{n_s \times n_s}$ is a diagonal matrix representing the volumetric excitation field corresponding to the k th illumination pattern, p being the total number of illumination patterns (Dutta et al. 2010). For OS implementation, we group the rows of \mathbf{A}^{em} into s blocks or subsets and denote the i th block as \mathbf{A}_i^{em} . Correspondingly, for p illumination patterns, the i th block of the system matrix is:

$$A_i = \begin{bmatrix} A_i^{em} D_1^{ex} \\ A_i^{em} D_2^{ex} \\ \vdots \\ A_i^{em} D_p^{ex} \end{bmatrix}. \quad (30)$$

The gradient for the i th block then is given by:

$$g_i(\mathbf{x}) = A_i' (A_i \mathbf{x} - \mathbf{b}_i) + \lambda_{TV} C' ((C\mathbf{x}) \circ z(C\mathbf{x})) + \lambda_{L^1} \mathbf{1}. \quad (31)$$

Here \mathbf{b}_i represents the block of the data vector corresponding to A_i . The modified update equation for the i th block and for the n th iteration then is:

$$\mathbf{x}^{n,i+1} = \left[\mathbf{x}^{n,i} - \alpha^n D^{-1}(\mathbf{x}^{n,i}) g_i(\mathbf{x}^{n,i}) \right]_+, \quad (32)$$

where α^n is the relaxation parameter or step size.

2.3. Compound approach

We compared the convergence speeds of the SPS algorithm with ordered subsets (OSSPS) with the PCG algorithm using a simulation setup that will be described in section 3. The convergence speed is highly dependent on the initialization. The general observation was that, for starting points very far from the true solution, the OSSPS approach was much faster than PCG. In the vicinity of the solution, however, PCG was significantly faster. This motivated us to use a compound approach where we use a few iterations of OSSPS to initialize the optimization problem and then use PCG to determine the final solution (Li, Ahn & Leahy 2005). Figure 1 demonstrates the speed-up achieved by using the compound approach. We obtained convergence curves for OSSPS-PCG for different transition points as shown in figure 1. Based on these curves, we chose to switch between the methods after the 10th iteration. It should be noted that one complete cycle through all the subsets is considered one iteration here. All convergence plots were obtained for a randomly picked positive starting point.

3. Results

We tested the joint L^1 and TV regularization approach and compared it with the L^1 , TV, and L^2 penalties using simulated and experimental data. In this section, we describe the experimental setup, present some simulations based on a setup similar to the experimental setup, evaluate the different penalties on the basis of some performance metrics, and, finally, present reconstruction results for experimental data.

3.1. Experimental setup

We acquired data using a mouse-shaped phantom. The experimental and computational steps leading up to the inverse problem are described below:

3.1.1. Phantom description—The mouse-shaped phantom (shown in figure 2(a)) was made using a mold created from a euthanized mouse. The phantom material consisted of 1% Intralipid, 2% agar, and 20 μM hemoglobin (Li et al. 2009). Two transparent capillary tubes, each 20 mm long, 1 mm in diameter, and filled with 1 μM DiD solution (D307, Invitrogen Corporation, excitation peak 648 nm, emission peak 669 nm, <http://products.invitrogen.com/ivgn/product/D307>), were embedded inside the mouse-

shaped phantom. The embedded line sources are visible in the CT image, shown in figure 2(b).

3.1.2. Data acquisition—Surface fluorescence data was collected using an FMT imaging setup equipped with an EMCCD camera for image acquisition and a conical mirror for full surface viewing (Li et al. 2009) shown in figure 3(a). The excitation source was a 650 nm laser beam focused to spots of ~1 mm diameter on the surface of the phantom. A pair of motorized mirrors was used to direct the beam to 54 different locations scattered nonuniformly over the surface of the phantom (as shown in figure 3). Fluorescence data was collected at an emission wavelength of 720 nm. MicroCT scans of the phantom were obtained on an Inveon MM system (Siemens Preclinical Solutions). 180 angular projections were acquired over 360° with an X-ray tube voltage of 80 kVp and current of 0.3 mA. Images were reconstructed using the filtered backprojection algorithm with a voxel size of 0.097732 mm.

3.1.3. Data preprocessing—The mapping of the fluorescence data from the CCD camera images to the phantom surface requires a digital volumetric representation of the phantom. We used the Digimouse atlas (<http://neuroimage.usc.edu/Digimouse.html>) (Dogdas et al. 2007, Stout et al. 2002), a labeled atlas, based on co-registered CT and cryosection images of a 28 g normal male nude mouse. We first extracted the surface geometry of the phantom from its CT image. We then used the DigiWarp technique (Joshi et al. 2009, Joshi et al. 2010) to warp the Digimouse atlas to fit the extracted phantom surface. The original and warped Digimouse surfaces are shown in figure 2(c). The fluorescence data was then mapped to the warped atlas surface using a calibrated mapping technique developed for the conical mirror based imaging system (Dutta 2011).

3.1.4. Forward model generation—The warped Digimouse atlas, with 306,771 tetrahedrons and 58,244 tessellation nodes, was used to solve the forward problem. We first used the finite element method to compute the excitation and emission forward model matrices, $A^{ex} \in \mathbb{R}^{n_s \times n_d}$ and $A^{em} \in \mathbb{R}^{n_d \times n_s}$, for $n_d = 6,762$ surface detector nodes and $n_s = 11,036$ volumetric grid points and for excitation and emission wavelengths of 650 nm and 720 nm respectively. The volumetric grid points with a uniform spacing of 1 mm represent the source space for reconstruction. Next, A^{ex} was used to compute the excitation fields, D_k^{ex} , k being the pattern index, for all the 54 patterns. Owing to the large size of the forward model matrix, A , as revealed by (29), it is important to avoid its direct computation. Instead, we use its building blocks – A^{em} and the D_k^{ex} matrices – to efficiently compute forward- and backprojections given by:

$$\mathbf{Ax} = \begin{bmatrix} A^{em} (d_1^{ex} \circ \mathbf{x}) \\ A^{em} (d_2^{ex} \circ \mathbf{x}) \\ \vdots \\ A^{em} (d_p^{ex} \circ \mathbf{x}) \end{bmatrix} \quad (33)$$

$$A' \mathbf{y} = \sum_k d_k^{ex} \circ ((A^{em})' \mathbf{y}_k), \quad (34)$$

where $\mathbf{x} \in \mathbb{R}^{n_s}$ and $\mathbf{y} \in \mathbb{R}^{n_d p}$ are arbitrary vectors, d_k^{ex} is a vector containing the diagonal terms of D_k^{ex} , and \mathbf{y}_k is the k th block of \mathbf{y} such that $\mathbf{y} = [y'_1 \ y'_2 \ \dots \ y'_p]$.

3.2. Simulation results

We simulated two fluorescent line sources inside the warped Digimouse atlas. The excitation and emission wavelengths were assumed to be 650 nm and 720 nm respectively – the same used in the experiment. The same 54 illumination locations from the experiment were used for excitation. White Gaussian noise with an SNR of 5 was added to the fluorescence data to make the simulations realistic. Coronal sections of the warped mouse atlas indicating the positions of the two simulated line sources are shown in figure 4(a). With joint L^1 and TV penalties, there are two regularization parameters to be selected. These parameters were empirically determined by sweeping them over a range of values. The result of this study is presented in figure 5, where a coronal slice from the reconstructed image is shown for different combinations of values for these two parameters. We use values of 10^{-5} , 5×10^{-5} , and 2.5×10^{-4} for λ_{L^1} and 4×10^{-7} , 2×10^{-6} , and 10^{-5} for λ_{TV} . This figure clearly indicates the increasing sparsifying and smoothing effects as we increase the weights on the L^1 and TV penalties respectively. Based on this study, we pick intermediate values of 5×10^{-5} for λ_{L^1} and 2×10^{-6} for λ_{TV} . Figures 4(b), (c), (d), and (e) show full slice-by-slice reconstruction results for L^2 , L^1 , TV, and joint L^1 -TV penalties respectively for indicated regularization parameters.

We used 6 noisy realizations of the described simulation setup to compute a set of metrics for evaluating the different regularization schemes. The fluorescence data was corrupted using Gaussian noise with an SNR of 1. The regularization parameters used are $\lambda_{L^2} = 10^{-3}$, $\lambda_{L^1} = 5 \times 10^{-5}$, and $\lambda_{TV} = 2 \times 10^{-6}$ (same as in figure 4). The computed metrics include the mean squared error (MSE), the mean and standard deviation of the signal over the region of interest (ROI), the mean and standard deviation of the background signal, and, finally, the signal-to-background ratio (SBR), defined as the ratio of the ROI mean to the background mean. The ROI is defined as the region consisting of voxels with a signal level of 1, corresponding to the two line sources in figure 4(a). The remaining voxels with zero signal comprise the background. These performance metrics for the different penalties are shown in figure 6. The joint L^1 -TV penalty was observed to yield the lowest MSE value. The L^1 , TV, and the joint L^1 -TV schemes led to stronger mean signal levels over the ROI than the L^2 . The L^1 and the joint L^1 -TV approaches generated the least mean background signal level, while the joint L^1 -TV approach yielded a lower background standard deviation than the L^1 and TV penalties individually. Lastly, the joint L^1 -TV scheme offered an SBR in between those for the L^1 and TV penalties individually, while all of these three schemes had an SBR significantly better than that for the L^2 penalty.

3.3. Experimental results

We reconstructed the experimental data acquired using the setup described in section 3.1 using the L^2 , L^1 , TV, and joint L^1 -TV penalties. The results are shown in figures 7(a), (b), (c), and (d) respectively for indicated regularization parameters. The result obtained using the joint L^1 -TV regularization scheme is overlaid against the CT image and displayed in figure 8. The two images were registered and displayed using the AMIDE tool for viewing and analyzing medical imaging data (Loening and Gambhir 2003). We observe a reasonable degree of overlap between the reconstructed FMT image and the ground truth revealed by the CT image.

The performance metrics used to evaluate the simulation results utilize quantitative knowledge of the ground truth as well as multiple noise realizations. While the CT image contains information about the location of the capillary tubes, it does not provide the true intensities of the targets. We, therefore, are unable to directly apply the same set of performance metrics to the experimental study. Instead, we resort to the Dice similarity coefficient (Dice 1945), a spatial similarity index, for quantitative evaluation of the

experimental results. The CT image was first blurred using a Gaussian kernel with full width at half maximum of 4 mm to make the resolutions of the two modalities comparable. Intensity based segmentation was then performed on the blurred CT image as well as the reconstructed FMT images. A numeric label $i \in \{1, 2, 3\}$ was assigned to each image voxel, $i = 1$ representing target 1 (the line source to the right in figure 8), $i = 2$ representing target 2 (the line source to the left in figure 8), and $i = 3$ representing the background. The Dice similarity coefficient, d , was then computed as:

$$d = \frac{2 |I_{CT}^i \cap I_{FMT}^i|}{|I_{CT}^i| + |I_{FMT}^i|}, \quad (35)$$

where I_{CT}^i and I_{FMT}^i represent the sets of voxel indices that carry label i in the CT and FMT images respectively and the notation $|\cdot|$ represents the cardinality of a set. The value of d ranges from 0, indicating no spatial overlap between the two sets, to 1, indicating complete overlap. The results, shown in figure 9, indicate that the L^2 and the joint L^1 -TV penalties perform consistently well for all three segments while the relative performance of the individual L^1 and TV penalties varies.

4. Summary and Discussion

We have applied a combination of L^1 and TV penalties to the FMT inverse problem to simultaneously encourage properties of sparsity and smoothness in our reconstructed images. We derived inspiration from the fact that very often fluorescent probes and contrast agents are concentrated within specific areas of interest (e.g., inside localized tumors and excretory organs). We have presented a compound method that uses a combination of OSSPS and PCG algorithms to generate a convergent solution.

Exploiting the non-negativity constraint, we substituted the L^1 norm penalty with a linear term. SPS functions were derived for the data-fitting term and the TV penalty. While the MM algorithm guarantees monotonicity, the OS approach that was incorporated to increase convergence speed does not guarantee convergence. Typically this is handled by using a diminishing step size. But our compound approach, where the final solution is obtained using PCG, guarantees convergence. This approach was inspired by the observation that, far away from the solution, the SPS-OS method has a faster rate of convergence, while PCG is faster in the vicinity of the solution. The advantage of the OSSPS approach over PCG is that it eliminates the need for an expensive line search and requires fewer numbers of flops per iteration. On a 2.93 GHz Intel[®] Core i7 CPU, it took about 40 s to complete one iteration of PCG while it took about 20 s to cycle through all 5 subsets and complete one iteration of OSSPS. 5 data blocks were used for the OS implementation, and the order of the blocks was chosen randomly at the start of the optimization procedure.

We validated the derived procedure using simulated and experimental data based on a mouse-shaped phantom with two embedded line sources. The simulation results showed that the L^1 , TV, and joint L^1 -TV penalties generated lower MSE values, stronger ROI mean signal levels, and higher SBR values than the L^2 regularizer. The joint L^1 -TV approach generated a smooth solution with a sparse background. The joint L^1 -TV result had the least mean background signal level and a lower background standard deviation than the L^1 and TV penalties individually. The L^1 -TV reconstructed image obtained from experimental data was qualitatively validated by overlaying it against a CT image of the phantom. Quantitative evaluation of the experimental results was performed by computing Dice similarity coefficients between the FMT images and the CT image blurred to a similar resolution.

While the simulation and experimental setups appear similar, they are not identical. The only sources of inaccuracy in simulation are noise and numerical errors due to the ill-conditioning of the system matrix. In comparison, the experimental results are susceptible to a number of nonidealities arising from imperfect modeling and registration. These include incorrect estimates of optical properties, inaccuracies in the noise model, geometric and radiometric calibration errors associated with the imaging system, and imperfect alignment of the CT-derived surface in the FMT object space – amongst other factors. Unlike the experimental procedure which is subject to such uncertainties, the simulation setup allows us to focus exclusively on the relative merits of the reconstruction schemes.

Additionally, the illumination setup used consisted of a group of point excitation locations nonuniformly scattered over the surface. The suboptimality of this configuration leads to significant variations in sensitivity over different volumetric locations. A manifestation of this can be found in the marked difference between the reconstruction quality of the two fluorescence targets – in both simulation and experimental results. We may treat the squared L^2 norm of a column of the system matrix corresponding to particular voxel as a measure of the relative sensitivity of that voxel. The overall sensitivity (averaged over constituent voxels) is 3.9957×10^{-5} for target 1 (which appears brighter in the reconstructed images) and 1.1718×10^{-5} for target 2 (which appears fainter). Of the different penalties, the difference in the reconstruction quality between the two targets is perhaps least pronounced in the TV case. This is due to the inherent tendency of the TV penalty to promote piecewise constancy.

Using L^1 or TV regularization, in combination or separately, clearly leads to improvements in localizing fluorescent sources in FMT in cases where the true distribution is consistent with the assumptions implicit in using these regularizers. There is less difference between the L^1 , TV and, joint L^1 -TV schemes than between any of these and L^2 . Qualitatively, the joint L^1 -TV images had the most natural appearance in the simulation and phantom studies we performed, but the quantitative studies reported in figure 6 do not identify a clear winner. As a final note, we observe that even with the use of these regularizers, FMT struggles to produce 3D images with resolutions that are routine in small animal PET and SPECT studies. This is because the FMT inverse problem is inherently more ill-posed, a situation which is exacerbated by uncertainties in the forward model. While the radiative transfer equation (Klose et al. 2005) models photon propagation more accurately than the diffusion equation (used in this work), it adds to the uncertainty by assuming prior knowledge of tissue anisotropies and also poses a significant computational challenge. The use of a larger number of illumination/detection wavelengths or a more optimal set of spatial patterns for illumination may lead to further improvements provided they improve the conditioning of the forward model.

Acknowledgments

We thank Dr. Anand A. Joshi at the University of Southern California for assistance with the DigiWarp tool for atlas warping. This work was supported by the National Cancer Institute under grants R01CA121783 and R44CA138243.

References

- Ahn S, Fessler J. Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms. *Medical Imaging, IEEE Transactions*. 2003; 22(5):613–626.
- Axelsson J, Svensson J, Andersson-Engels S. Spatially varying regularization based on spectrally resolved fluorescence emission in fluorescence molecular tomography. *Optics Express*. 2007; 15(21):13574–13584. [PubMed: 19550625]

- Beck A, Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*. 2009; 2(1):183–202.
- Bélanger S, Abran M, Intes X, Casanova C, Lesage F. Real-time diffuse optical tomography based on structured illumination. *Journal of Biomedical Optics*. 2010; 15:016006. [PubMed: 20210452]
- Bertsekas, DP. *Nonlinear Programming*. Athena Scientific; 1999.
- Bloch S, Lesage F, McIntosh L, Gandjbakhche A, Liang K, Achilefu S. Whole-body fluorescence lifetime imaging of a tumor-targeted near-infrared molecular probe in mice. *Journal of Biomedical Optics*. 2005; 10(5):054003. [PubMed: 16292963]
- Candes EJ, Wakin MB. An introduction to compressive sampling. *Signal Processing Magazine, IEEE*. 2008; 25(2):21–30.
- Cao N, Nehorai A, Jacobs M. Image reconstruction for diffuse optical tomography using sparsity regularization and expectation-maximization algorithm. *Optics Express*. 2007; 15(21):13695–13708. [PubMed: 19550641]
- Chambolle A. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*. 2004; 20:89–97.
- Chan T, Golub G, Mulet P. A nonlinear primal-dual method for total variation-based image restoration. *ICAOS'96*. 1996:241–252.
- Chaudhari AJ, Ahn S, Levenson R, Badawi RD, Cherry SR, Leahy RM. Excitation spectroscopy in multispectral optical fluorescence tomography: methodology, feasibility and computer simulation studies. *Physics in Medicine and Biology*. 2009; 54(15):4687–4704. [PubMed: 19590118]
- Chen SS, Donoho DL, Saunders MA. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*. 1998; 20(1):33–61.
- Daubechies, I.; Frieze, MD.; Mol, CD. *Communications in Pure and Applied Mathematics*. 2004. An iterative thresholding algorithm for linear inverse problems with a sparsity constraint; p. 14131457
- Dice L. Measures of the amount of ecologic association between species. *Ecology*. 1945; 26(3):297–302.
- Dogdas B, Stout D, Chatziioannou A, Leahy RM. Digimouse: A 3D whole body mouse atlas from CT and cryosection data. *Physics in Medicine and Biology*. 2007; 52:577–587. [PubMed: 17228106]
- Donoho DL. Compressed sensing. *Information Theory, IEEE Transactions*. 2006; 52(4)
- Dutta, J. PhD thesis. University of Southern California; Los Angeles, California: 2011. Computational methods for fluorescence molecular tomography.
- Dutta, J.; Ahn, S.; Joshi, AA.; Leahy, RM. 'Biomedical Imaging: From Nano to Macro ISBI '09 IEEE International Symposium on'. 2009. Optimal illumination patterns for fluorescence tomography; p. 1275-1278.
- Dutta J, Ahn S, Joshi AA, Leahy RM. Illumination pattern optimization for fluorescence tomography: theory and simulation studies. *Physics in Medicine and Biology*. 2010; 55(10):2961. [PubMed: 20436232]
- Dutta, J.; Ahn, S.; Li, C.; Chaudhari, AJ.; Cherry, SR.; Leahy, RM. *Medical Imaging 2008: Physics of Medical Imaging*. Vol. 6913. SPIE; 2008. Computationally efficient perturbative forward modeling for 3D multispectral bioluminescence and fluorescence tomography; p. 69130C
- Efron B, Hastie T, Johnstone I, Tibshirani R. Least angle regression. *The Annals of Statistics*. 2004; 32(2):407–499.
- Eppstein MJ, Hawrysz DJ, Godavarty A, Sevick-Muraca EM. Three-dimensional, Bayesian image reconstruction from sparse and noisy data sets: Near-infrared fluorescence tomography. *Proceedings of the National Academy of Sciences*. 2002; 99(15):9619–9624.
- Erdogan H, Fessler JA. Ordered subsets algorithms for transmission tomography. *Physics in Medicine and Biology*. 1999; 44(11):2835–2851. [PubMed: 10588288]
- Fessler JA, Clinthorne NH, Rogers WL. Regularized emission image reconstruction using imperfect side information. *Nuclear Science, IEEE Transactions*. 1992; 39(5):1464–1471.
- Fessler JA, Hero AOI. Penalized maximum-likelihood image reconstruction using space-alternating generalized EM algorithms. *Image Processing, IEEE Transactions*. 1995; 4(10):1417–1429.

- Figueiredo MAT, Nowak RD, Wright SJ. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *Selected Topics in Signal Processing, IEEE Journal*. 2007; 1(4):586–597.
- Gao H, Zhao H. Multilevel bioluminescence tomography based on radiative transfer equation part 1: L1 regularization. *Optics Express*. 2010a; 18(3):1854–1871. [PubMed: 20174013]
- Gao H, Zhao H. Multilevel bioluminescence tomography based on radiative transfer equation part 2: total variation and L1 data fidelity. *Optics Express*. 2010b; 18(3):2894–2912. [PubMed: 20174118]
- Gardner, C.; Dutta, J.; Mitchell, GS.; Ahn, S.; Li, C.; Harvey, P.; Gershman, R.; Sheedy, S.; Mansfield, J.; Cherry, SR.; Leahy, RM.; Levenson, R. *Biomedical Optics, OSA Technical Digest (CD) (Optical Society of America, 2010). OSA: Paper BTuF1; 2010. Improved *in vivo* fluorescence tomography and quantitation in small animals using a novel multiview, multispectral imaging system.*
- Green PJ. Bayesian reconstructions from emission tomography data using a modified EM algorithm. *Medical Imaging, IEEE Transactions*. 1990; 9(1):84–93.
- Guyen M, Yazici B, Intes X, Chance B. Diffuse optical tomography with a priori anatomical information. *Physics in Medicine and Biology*. 2005; 50(12):2837. [PubMed: 15930606]
- Hebert T, Leahy RM. A generalized EM algorithm for 3-D Bayesian reconstruction from Poisson data using Gibbs priors. *Medical Imaging, IEEE Transactions*. 1989; 8(2):194–202.
- Huang Y, Ng MK, Wen YW. A fast total variation minimization method for image restoration. *Multiscale Modeling & Simulation*. 2008; 7(2):774–795.
- Hyde D, Miller EL, Brooks DH, Ntziachristos V. Data specific spatially varying regularization for multimodal fluorescence molecular tomography. *Medical Imaging, IEEE Transactions*. 2010; 29(2):365–374.
- Joshi AA, Chaudhari AJ, Li C, Dutta J, Cherry SR, Shattuck DW, Toga AW, Leahy RM. DigiWarp: a method for deformable mouse atlas warping to surface topographic data. *Physics in Medicine and Biology*. 2010; 55(20):6197. [PubMed: 20885019]
- Joshi, AA.; Chaudhari, AJ.; Li, C.; Shattuck, D.; Dutta, J.; Leahy, RM.; Toga, AW. *Biomedical Imaging: From Nano to Macro, 2009 ISBI '09 IEEE International Symposium on. 2009. Posture matching and elastic registration of a mouse atlas to surface topography range data; p. 366-369.*
- Joshi A, Bangerth W, Sevcik-Muraca EM. Non-contact fluorescence optical tomography with scanning patterned illumination. *Optics Express*. 2006; 14(14):6516–6534. [PubMed: 19516829]
- Kim SJ, Koh K, Lustig M, Boyd S, Gorinevsky D. An interior-point method for large-scale l1-regularized least squares. *Selected Topics in Signal Processing, IEEE Journal*. 2007; 1(4):606–617.
- Klose A, Ntziachristos V, Hielscher A. The inverse source problem based on the radiative transfer equation in optical molecular imaging. *Journal of Computational Physics*. 2005; 202(1):323–345.
- Lange K. Convergence of em image reconstruction algorithms with Gibbs smoothing. *Medical Imaging, IEEE Transactions*. 1990; 9(4):439–446.
- Lange K, Fessler JA. Globally convergent algorithms for maximum a posteriori transmission tomography. *Image Processing, IEEE Transactions*. 1995; 4(10):1430–1438.
- Lange K, Hunter DR, Yang I. Optimization transfer using surrogate objective functions. *Journal of Computational and Graphical Statistics*. 2000; 9(1):1–20.
- Leahy, RM.; Yan, X. *Information Processing in Medical Imaging*. In: Colchester, A.; Hawkes, D., editors. *Lecture Notes in Computer Science*. Vol. 511. Springer Berlin; Heidelberg: 1991. p. 105-120.
- Li A, Boverman G, Zhang Y, Brooks D, Miller EL, Kilmer ME, Zhang Q, Hillman EMC, Boas DA. Optimal linear inverse solution with multiple priors in diffuse optical tomography. *Applied Optics*. 2005; 44(10):1948–1956. [PubMed: 15813531]
- Li C, Mitchell GS, Dutta J, Ahn S, Leahy RM, Cherry SR. A three-dimensional multispectral fluorescence optical tomography imaging system for small animals based on a conical mirror design. *Optics Express*. 2009; 17(9):7571–7585. [PubMed: 19399136]
- Li Q, Ahn S, Leahy RM. Fast hybrid algorithms for PET image reconstruction. *Nuclear Science Symposium Conference Record, 2005 IEEE*. 2005; 4:390–393.

- Loening A, Gambhir S. AMIDE: a free software tool for multimodality medical image analysis. *Molecular Imaging*. 2003; 2(3):131–137. [PubMed: 14649056]
- Massoud TF, Gambhir SS. Molecular imaging in living subjects: seeing fundamental biological processes in a new light. *Genes & Development*. 2003; 17(5):545–580. [PubMed: 12629038]
- Mohajerani P, Eftekhari AA, Huang J, Adibi A. Optimal sparse solution for fluorescent diffuse optical tomography: theory and phantom experimental results. *Applied Optics*. 2007; 46(10):1679–1685. [PubMed: 17356610]
- Natarajan BK. Sparse approximate solutions to linear systems. *SIAM Journal on Computing*. 1995; 24(2):227–234.
- Pogue BW, McBride TO, Prewitt J, Österberg UL, Paulsen KD. Spatially variant regularization improves diffuse optical tomography. *Applied Optics*. 1999; 38(13):2950–2961. [PubMed: 18319877]
- Rudin LI, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*. 1992; 60(1–4):259–268.
- Shu X, Royant A, Lin MZ, Aguilera TA, Lev-Ram V, Steinbach PA, Tsien RY. Mammalian Expression of Infrared Fluorescent Proteins Engineered from a Bacterial Phytochrome. *Science*. 2009; 324(5928):804–807. [PubMed: 19423828]
- Stout D, Chow P, Silverman R, Leahy RM, Lewis X, Gambhir S, Chatziioannou A. Creating a whole body digital mouse atlas with PET, CT and cryosection images. *Molecular Imaging and Biology*. 2002; 4(4):S27.
- Strong D, Chan T. Edge-preserving and scale-dependent properties of total variation regularization. *Inverse Problems*. 2003; 19(6):S165.
- Tibshirani R. Regression shrinkage and selection via the Lasso *Journal of the Royal Statistical Society. Series B (Methodological)*. 1996; 58(1):267–288.
- Tikhonov, AN.; Arsenin, VY. *Solutions of Ill-Posed Problems*. John Wiley & Sons; Washington, D.C: 1977.
- Vogel C, Oman M. Iterative methods for total variation denoising. *SIAM Journal on Scientific Computing*. 1996; 17(1):227–238.
- Vogel CR, Oman ME. Fast, robust total variation-based reconstruction of noisy, blurred images. *Image Processing, IEEE Transactions*. 1998; 7(6):813–824.
- Weissleder R, Ntziachristos V. Shedding light onto live molecular targets. *Nature Medicine*. 2003; 9:123–128.
- Weissleder R, Tung CH, Mahmood U, Bogdanov A. In vivo imaging of tumors with protease-activated near-infrared fluorescent probes. *Nature Biotechnology*. 1999; 17(4):375–378.
- Zacharakis G, Kambara H, Shih H, Ripoll J, Grimm J, Saeki Y, Weissleder R, Ntziachristos V. Volumetric tomography of fluorescent proteins through small animals *in vivo*. *Proceedings of the National Academy of Sciences*. 2005; 102(51):18252–18257.

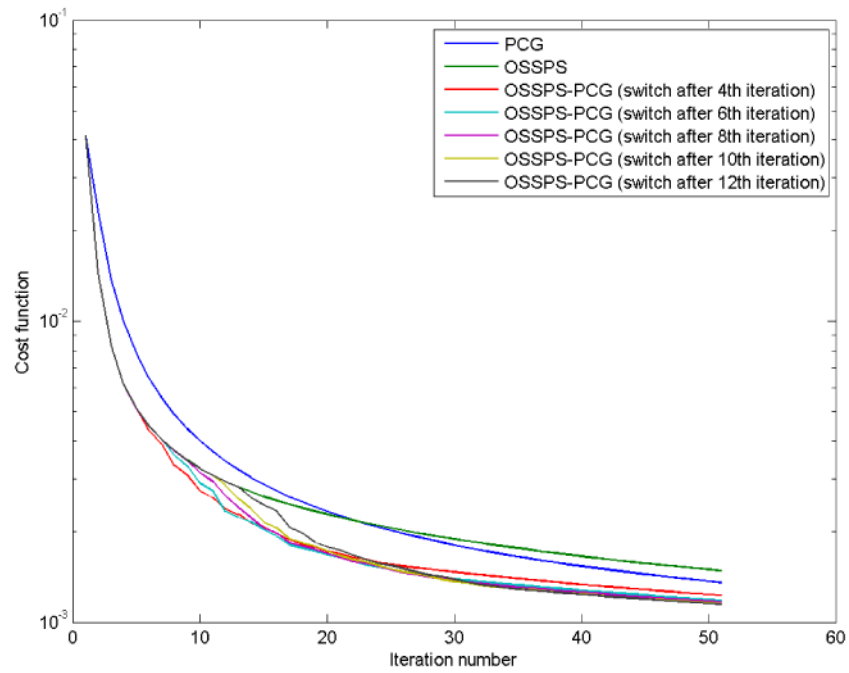


Figure 1. Comparison of convergence curves of PCG, OSSPS, and OSSPS-PCG with the transition from OSSPS to PCG occurring after iteration numbers 4, 6, 8, 10, and 12. These curves were obtained for the L^1 -TV penalty and for a random initialization. For OSSPS, one complete cycle through all the subsets is considered one iteration.

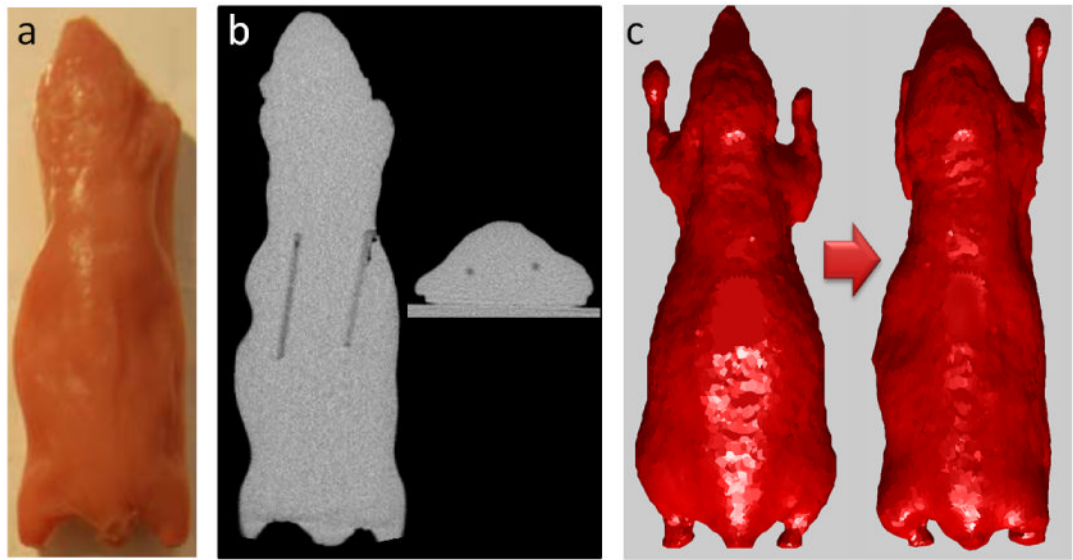


Figure 2. Mouse-shaped phantom. (a) Full-light photograph. (b) A coronal and a transverse section from the CT image showing the two tubes filled with fluorescent DiD dye. (c) Original and warped Digimouse surfaces.

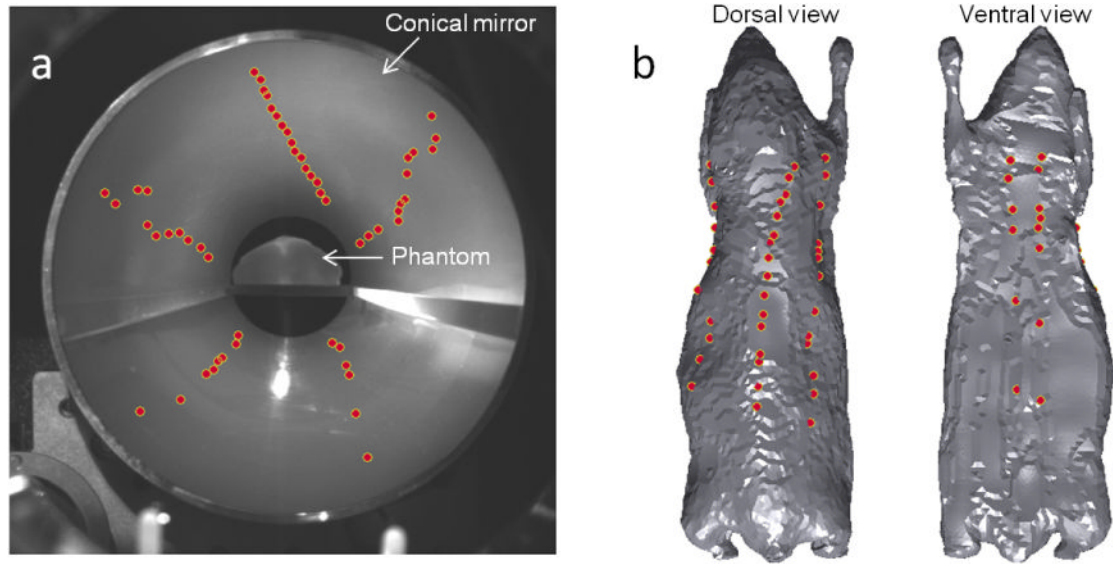


Figure 3.

(a) Full-light EMCCD camera image showing the mouse phantom seated on a stage with its full surface visible in the conical mirror. The dots represent the 54 illumination locations as seen in the image space. (b) Dorsal and ventral views of the tessellated mouse phantom showing the illumination points mapped to the object space.

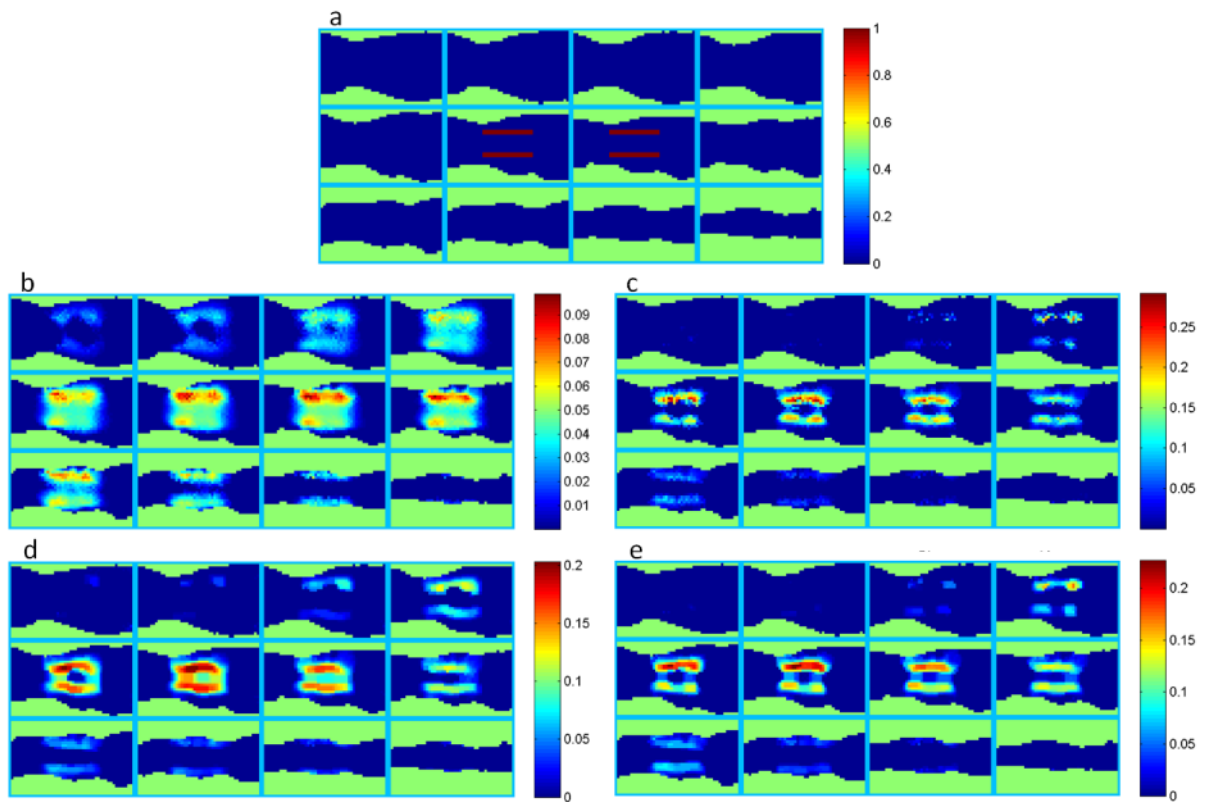


Figure 4.

Simulated mouse phantom: coronal sections of the mouse phantom showing (a) the two simulated line sources and the reconstruction result for (b) the L^2 penalty with a regularization parameter of 10^{-3} , (c) the L^1 penalty with a regularization parameter of 5×10^{-5} , (d) the TV penalty with a regularization parameter of 2×10^{-6} , and (e) both L^1 and TV penalties with regularization parameters of 5×10^{-5} and 2×10^{-6} respectively. The slices are ordered from bottom to top starting top left in the image and ending bottom right.

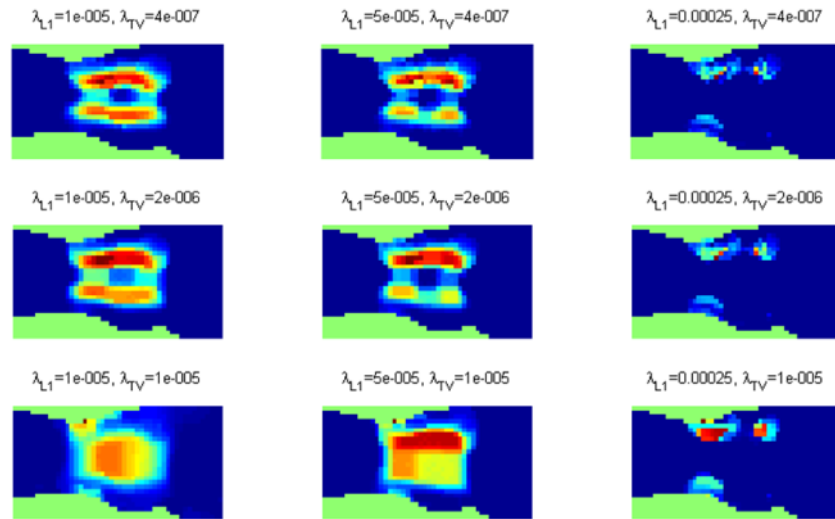


Figure 5. Reconstructed source distribution for different combinations of the L^1 and TV regularization parameters shown for a central coronal section from the phantom. These images demonstrate the sparsifying and smoothing effects of the L^1 and TV terms respectively.

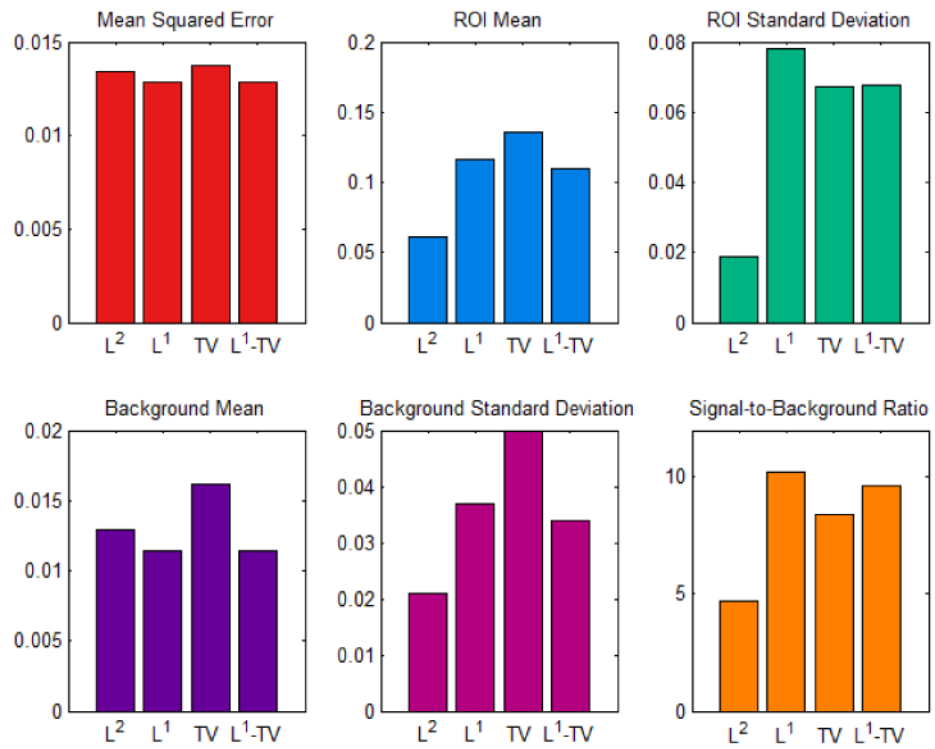


Figure 6. Performance metrics for comparison of the L^2 , L^1 , TV, and joint L^1 and TV penalty functions.

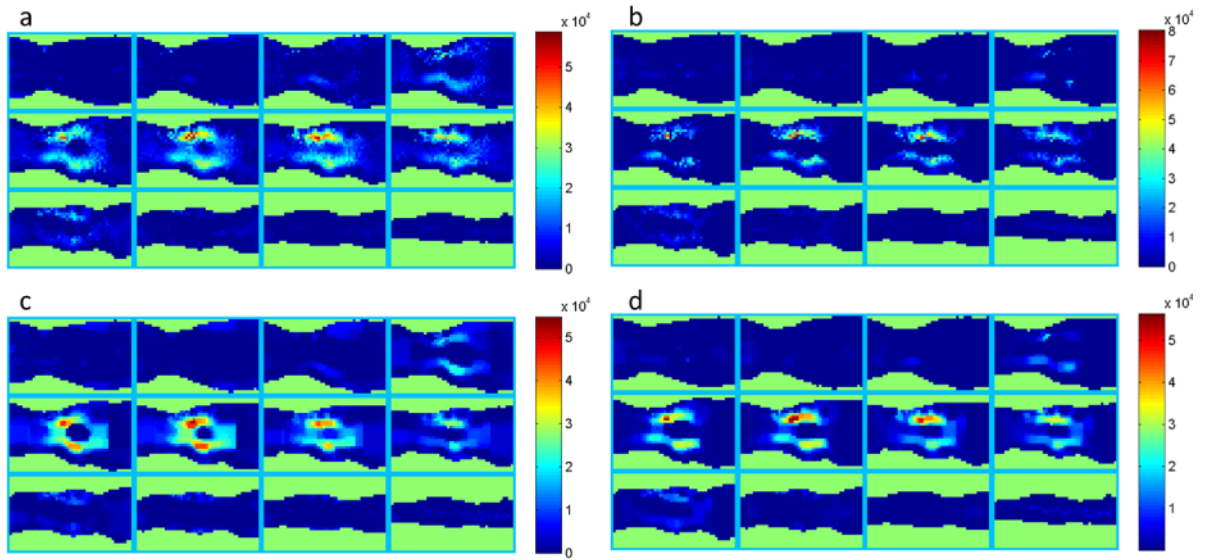


Figure 7.

Experimental mouse phantom: coronal sections of the mouse phantom showing the reconstruction of two embedded line sources using (a) the L^2 penalty with a regularization parameter of 10^{-4} , (b) the L^1 penalty with a regularization parameter of 20, (c) the TV penalty with a regularization parameter of 0.5, and (d) both L^1 and TV penalties with regularization parameters of 10 and 1 respectively.

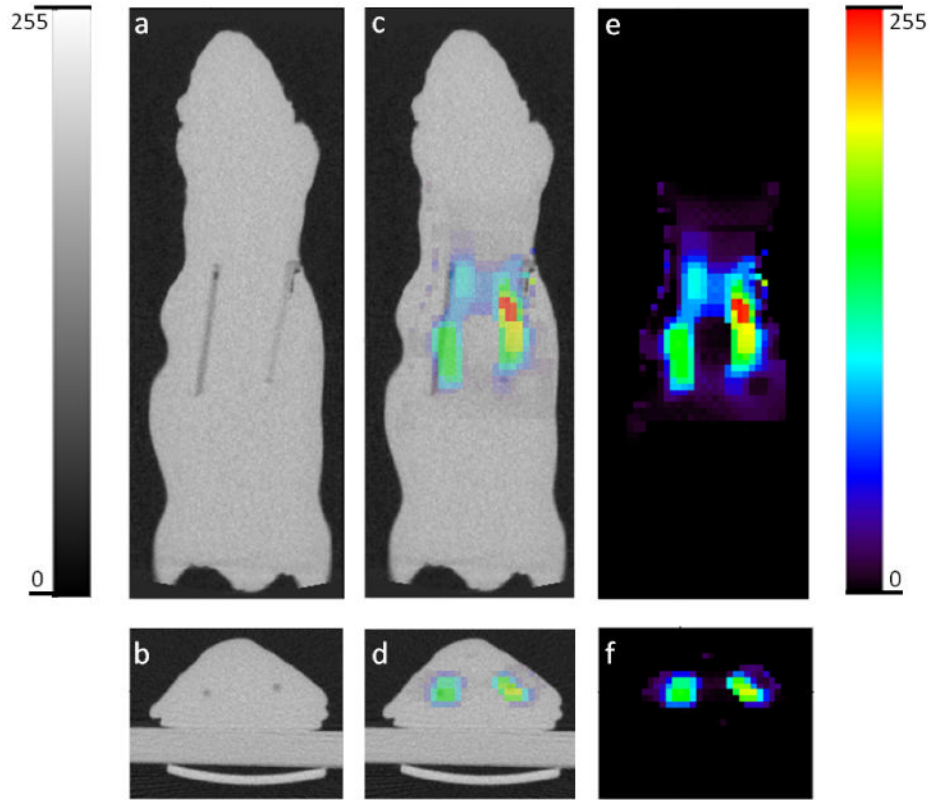


Figure 8.

(a) Coronal and (b) transverse sections of the CT image of the mouse-shaped phantom showing the two embedded fluorescent line sources. (c) Coronal and (d) transverse overlay of CT and FMT images. (e) Coronal and (f) transverse sections of the FMT image showing the two fluorescent line sources reconstructed using both L^1 and TV penalties with regularization parameters of 10 and 1 respectively.

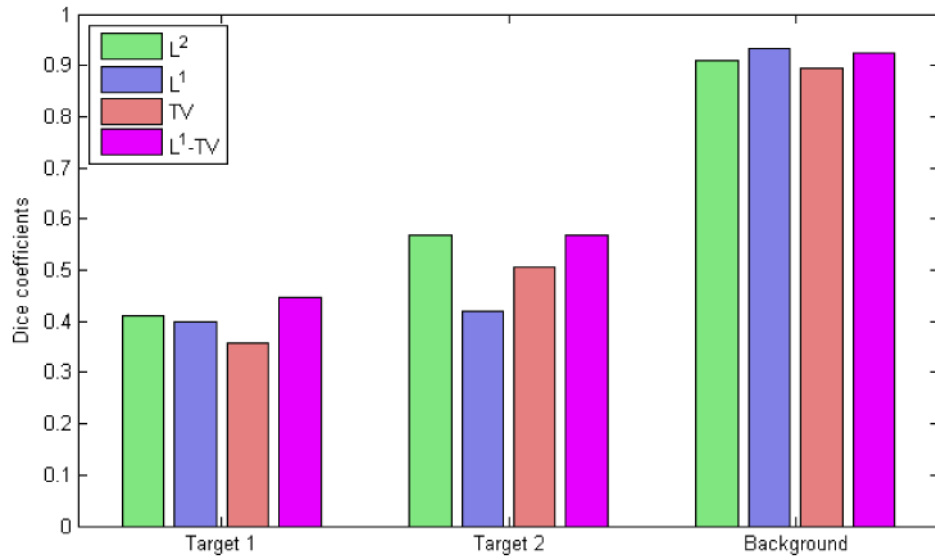


Figure 9. Dice coefficients representing the similarity between the CT image and the reconstructed FMT images for different penalties. Target 1 represents the line source to the right in all the subimages in figure 8 while target 2 represents the line source to the left.