Title

Nonconvex regularizations in fluorescence molecular tomography for sparsity enhancement

Journal

Physics in Medicine and Biology, 59(12)

ISSN

0031-9155

Authors

Zhu, Dianwen

Li, Changqing

Publication Date

2014-06-21

DOI

10.1088/0031-9155/59/12/2901

Peer reviewed

Nonconvex regularizations in fluorescence molecular tomography for sparsity enhancement

Dianwen Zhu, Changqing Li†

School of Engineering, University of California Merced,
Merced, CA 95343, USA

Abstract. *In vivo* fluorescence imaging has been a popular functional imaging modality in preclinical imaging. Near infrared probes used in fluorescence molecular tomography (FMT) are designed to localize in the targeted tissues, hence sparse solution to the FMT image reconstruction problem is preferred. Nonconvex regularization methods are reported to enhance sparsity in the fields of statistical learning, compressed sensing etc. We investigated such regularization methods in FMT for small animal imaging with numerical simulations and phantom experiments. We adopted a majorization-minimization (MM) algorithm for the iterative reconstruction process and compared the reconstructed images using our proposed nonconvex regularizations with those using the well known L^1 regularization. We found that the proposed nonconvex methods outperform L^1 regularization in accurately recovering sparse targets in FMT.

PACS numbers: 42.20.Wb, 42.30.Wb, 87.75.nf

Submitted to: *Phys. Med. Biol.*

† To whom correspondence should be addressed (cli32@ucmerced.edu)

1. INTRODUCTION

Fluorescence molecular tomography (FMT) is an important functional imaging tool that can monitor cell and molecular activities *in vivo*. Its applications include cancer imaging, drug delivery, stem cell trafficking, enzyme activity monitoring, etc. (Cherry 2004, Ntziachristos et al. 2005). FMT collects near infrared photons propagated to the object boundary from inside tissues, then recovers the internal fluorophore concentrations by minimizing the discrepancy between measurements from the FMT system and predicted values from a physical model—usually referred to as the "forward model", through a least squares model. The forward modeling is based on a diffusion equation that is solved on a finite element mesh (Fedele et al. 2003, Li et al. 2009a).

Due to the intense scattering and absorption of photons and the relatively small amount of collected data, FMT system is usually severely ill-conditioned. Improving the system matrix, of course, is another way to alleviate the ill-conditioning. A detailed list of references in this respect is available from (Dutta et al. 2012). Despite these efforts, due to the noise and numerical errors, regularization techniques are still needed to stabilize the system. Tikhonov or L^2 regularization is a popular choice. It is effectively assuming a multivariate Gaussian prior for the solution from a Bayesian point of view, which undesirably tend to oversmooth the image. In FMT for small animal imaging, the target is usually very small, hence the desired solution of the image reconstruction problem would be sparse. We point out that *in vivo* FMT system may suffer a lot of background signal or noise, which reduces the sparisty of the signal. So a background subtraction approach might be applied before we start the regularization methods (Ale et al. 2013). When incorporating this prior knowledge into the least squares model, L^1 regularization is the natural choice since it is well known for sparsity enhancement, especially in the field of compressed sensing (Candes et al. 2008). For the FMT system, comparison has been done recently on the effects of L^1 with joint L^1 and total variation (TV) regularization, noticing that TV may enhance local smoothness, although no clear winner was found comparing the performances of L^1 with joint L^1 and TV (Dutta et al. 2012). The sparsity promoting effects of L^p regularization for a more general $1 \leq p < 2$ has been investigated as well (Han et al. 2010). A matrix-free strategy using $L^p (1 \leq p < 2)$ regularization was also proposed for FMT (Baritaux et al. 2010). These papers clearly demonstrated the effects of $L^p (1 \leq p < 2)$ regularizations in enhancing sparsity. One feature in common among these efforts was that they were all solving convex optimization problems.

However, nonconvex regularizations have stronger effects in promoting sparse solutions, despite the potential local minima issues (Chartrand 2007). It has been actively studied in the signal processing, image denoising, image restoration fields, etc. (Chartrand 2007, Figueiredo et al. 2007). In paticular, a comprehensive study on different nonconvex regularizations have been done as well (Gasso et al. 2009). Recall that the ideal sparsity regularization would be the inclusion of the L^0 norm, but this minimization problem is combinatorial and is NP-hard to solve in general

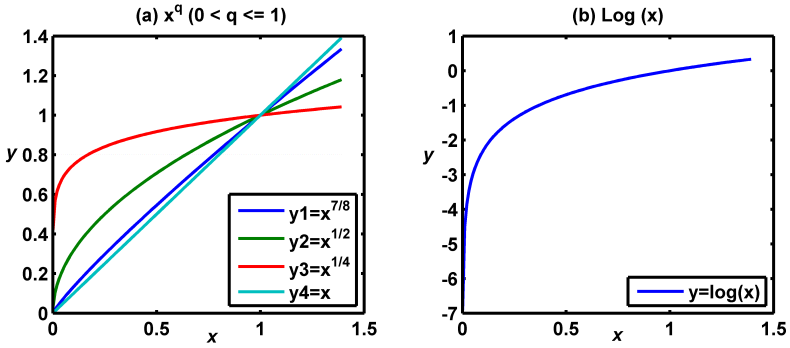

Figure 1. (a) Family of nonconvex functions $L^q\,(0 < q < 1)$ and (b) logarithmic function in one dimension. L^1 is also plotted as a reference.

(Natarajan 1995). L^1 and $L^p\,(1 < p < 2)$ norms are effective approximations, but clearly the L^q pseudo norm, as $q \to 0$, would be an even better approximation to L^0. We can intuitively understand the 1-dimensional version of this phenomenon from figure 1: on interval $x \in (0, 1)$, when we decrease the same amount of y in $L^q : y = x^q\,(0 < q < 1)$ and $L^1 : y = x$, we need to decrease x more significantly in the former, as illustrated in figure 1 (a).

In this paper, we aimed to study the effects of the family of nonconvex $L^q(0 < q < 1)$ regularizations for FMT, and compare their effects in promoting sparsities with the well known L^1 regularization. We also included $Logarithmic(Log)$ regularization as another potential nonconvex candidate for L^0 in promoting sparsity. Similar to the case of $L^q, (0 < q < 1)$, figure 1 (b) shows that to minimize $Log\,(x)$, we need x to rapidly decrease to 0 as well. To the best of our knowledge, effects of nonconvex regularizations have not been studied in FMT although similar efforts have been made in the broader field of medical image reconstruction. For example, a nonconvex reconstruction of 2D CT images was studied (Sidky et al. 2007). $L^q\,(0 < q < 1)$ regularizations was also reported to provide enhanced sparsity and improved spatial resolution for images reconstructed in a diffuse optical tomography (DOT) system (Okawa et al. 2011). However, their system matrix was much smaller in size and there is no non-negativity constraint, so their method can not be directly applied to our FMT model. Also in the field of DOT, it has been shown very recently that $L^q\,(0 < q \leq 1)$ regularizations hold promise in improving the reconstructed image quality (Prakash et al. 2014). Nonetheless, the DOT system is different and their approach is different than ours and can only handle nearly noise-free data.

2. MATERIALS AND METHODS

2.1. Regularized nonlinear least squares models

Image reconstruction in FMT is modeled as a regularized least squares problem with a non-negativity constraint, which can be written into the following form:

$$\hat{\mathbf{x}} = \min_{\mathbf{x}\geq 0} \Phi(\mathbf{x}) := \frac{1}{2}\|A\mathbf{x} - \mathbf{b}\|_{L^2}^2 + \lambda \cdot \mathbf{R}(\mathbf{x}), \tag{1}$$

where $\mathbf{x} = (x_j)_{n\times 1}$ represents the nodes to be reconstructed, $A = (a_{ij})_{m\times n}$ the system matrix, $\mathbf{b} = (b_j)_{m\times 1}$ the measurements, and $\lambda \cdot \mathbf{R}(\mathbf{x})$ the regularization term, reflecting our *a priori* knowledge about the solution. Although other prior information such as structure, shape, location, etc. of the target from other imaging modalities can also be incorporated (Li et al. 2009b, Wang et al. 2011), since our focus of this paper is on the sparsifying effects of regularization terms, we restrict:

$$\lambda \cdot \mathbf{R}(\mathbf{x}) = \lambda_1\|\mathbf{x}\|_{L^1} + \lambda_{\log}\|\mathbf{x}\|_{\log} + \lambda_q\|\mathbf{x}\|_{L^q}, \tag{2}$$

where $0 < q < 1$ and $\lambda_1, \lambda_{\log}$, and λ_q are regularization parameters, among which only one will be non-zero at a time so that we may compare the effects of different penalties.

We have the following two quick remarks:

(i) For the L^1 term, we can safely get rid of the absolute value sign, and replace the L^1 norm by the equivalent linear form: $\lambda_1\mathbf{1^t}\mathbf{x}$, due to the non-negativity constraint, where $\mathbf{1}^t$ is the vector of the same dimension as $\mathbf{x}$ but with all ones (Dutta et al. 2012).

(ii) For the nonconvex terms, we define

$$\|\mathbf{x}\|_{\log} = \sum_{i=1}^{n}[\log(x_j+\delta_{\log}) - \log\delta_{\log}] \quad \text{and} \quad \|\mathbf{x}\|_{L^q}^q = \sum_{i=1}^{n}(x_j+\delta_q)^q, \tag{3}$$

where we added small constants $\delta_{\log}$ and δ_q, (e.g. 1.0E-9) to each x_j in the above definitions to avoid singularities when x_j approaches 0. Note that we again omitted the absolute value symbols due to the non-negativity requirement.

2.2. Preconditioned conjugate gradient

Preconditioned conjugate gradient (PCG) algorithm is the typical way of solving least squares problems and it is usually quite effective. So one may take the first order derivative, or gradient, of $\Phi(\mathbf{x})$ with respect to $\mathbf{x}$, and then solve $\nabla\Phi(\mathbf{x}) = 0$. It is clear that

$$\nabla\Phi(\mathbf{x}) = A^tA\mathbf{x} - (A^t\mathbf{b} - \lambda_1\mathbf{1}_N) + \lambda_{\log}./(\mathbf{x}+\delta_{log}) + q\lambda_q.*(\mathbf{x}+\delta_q)^{q-1}, \tag{4}$$

where $./$ and $.*$ are entry-wise division and multiplication for vectors, respectively. Then one may follow the nonlinear PCG procedure. However, PCG suffers one major issue in our FMT model since the system matrix A is usually very large: a very expensive backtracking line search is needed to enforce the non-negativity constraint.

2.3. MM algorithm

We turn to methods that limits the operations on the system matrix to increase updating speed. MM algorithms—also known as optimization transfer algorithms and its special case, the separable quadratic surrogates (SQS) algorithm (Erdogan & Fessler 1999), has the desired advantages including avoiding matrix inversions, linearizing an optimization problem, dealing gracefully with inequalities, etc. (Hunter & Lange 2004). Recall the definition of a surrogate function $\Phi^{sur}(\mathbf{x})$ in the minimization problem, the following three conditions should hold:

$$\begin{cases} \Phi^{sur}(\mathbf{x}) & \geq \Phi(\mathbf{x}), & \text{for all } \mathbf{x}; \\ \Phi^{sur}(\mathbf{x}^n) & = \Phi(\mathbf{x}^n), & \text{at some point } \mathbf{x}^n; \\ \nabla\Phi^{sur}(\mathbf{x}^{\mathbf{n}}) & = \nabla\Phi(\mathbf{x}^{\mathbf{n}}), & \text{at point } \mathbf{x}^n. \end{cases} \tag{5}$$

There are different ways to implement MM algorithms. For example, one may add a nonnegative term to obtain the surrogate function (Prakash et al. 2014). In the following, we first follow the SQS routine (Erdogan & Fessler 1999) to majorize the least squares fitting term:

$$\begin{aligned} & \frac{1}{2}\|\mathbf{b}-\mathbf{A}\mathbf{x}\|_{L^2}^2 = \frac{1}{2}\sum_{i=1}^{m}(b_i-(A\mathbf{x})_i)^2 \\ \leq\ & \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{n}\beta_{ij}\{b_i-(A\mathbf{x}^k)_i-\frac{a_{ij}}{\beta_{ij}}(x_j-x_j^k)\}^2 \\ =\ & \sum_{j=1}^{n}\{\frac{(x_j-x_j^k)^2}{2}\sum_{i=1}^{m}\frac{a_{ij}^2}{\beta_{ij}} - x_j\sum_{i=1}^{m}a_{ij}(b_i-(A\mathbf{x}^k)_i)+g_j(x_j^k)\} \\ :=\ & \sum_{j=1}^{n}\Phi_j^{ols}(x_j), \end{aligned} \tag{6}$$

where $g_j(x_j^k)$ denotes a function of x_j^k only, and the parameters β_{ij} are non-negative and satisfy that $\sum_{j=1}^{n}\beta_{ij}=1$. It is also helpful to keep in mind that only x_j's are the unknown variables, all the others (e.g. $\mathbf{x}^k, x_j^k$) are known constants. Following (Fessler et al. 1997), we conveniently set $\beta_{ij}=a_{ij}/\sum_j a_{ij}$ in this paper.

For the L^1 regularization, we don't need any surrogate:

$$\lambda_1\|\mathbf{x}\|_{L^1}=\lambda_1\sum_{j=1}^{n}x_j := \sum_{j=1}^{n}\Phi_j^{L^1}(x_j). \tag{7}$$

To handle the complex nonconvex regularizations $\log x_j$ and x_j^q $(0<q<1)$, we use the following:

$$\lambda_{\log}\log(x_j+\delta_{log}) \leq \lambda_{\log}\left(\log(x_j^k+\delta_{log})+(x_j^k+\delta_{log})^{-1}(x_j-x_j^k)\right), \tag{8}$$

$$\lambda_q(x_j+\delta_q)^q \leq \lambda_q\left((x_j^k+\delta_q)^q+q*(x_j^k+\delta_q)^{(q-1)}(x_j-x_j^k)\right). \tag{9}$$

These inequalities are straightforward since the right-hand side terms represent the lines tangent to $\log(x)$ and x^q, respectively, at point x^k, which are above the graphs of these concave functions. We can also quickly see that the 2nd and 3rd requirements in (5) hold. So we take the right-hand sides as our linear surrogates, and denote them as $\Phi_j^{Log}(x_j)$ and $\Phi_j^{L^q}(x_j)$ respectively.

2.4. Iterative update

Now that the surrogate function Φ^{sur} is decoupled, we can derive the iterative update by differentiating each variable separately and setting them to 0:

$$\begin{aligned}
&\frac{\partial \Phi_j^{sur}}{\partial x_j} = 0 \Rightarrow \frac{\partial \Phi_j^{ols}}{\partial x_j} + \frac{\partial \Phi_j^{L^1}}{\partial x_j} + \frac{\partial \Phi_j^{Log}}{\partial x_j} + \frac{\partial \Phi_j^{L^q}}{\partial x_j} = 0 \\
\Rightarrow \quad & (x_j - x_j^k) \sum_{i=1}^{m} \frac{a_{ij}^2}{\beta_{ij}} - \sum_{i=1}^{m} a_{ij}(b_i - (Ax^k)_i) \\
& + \lambda_1 + \frac{\lambda_{log}}{x_j^k + \delta_{log}} + \frac{q\lambda_q}{(x_j^k + \delta_q)^{1-q}} = 0.
\end{aligned} \tag{10}$$

Denoting $\sum_{i=1}^{m} \frac{a_{ij}^2}{\beta_{ij}}$ by κ_j, we derive the non-negative solution for the above equation in two steps:

(i) We obtain the non-negative *ordinary least squares* (OLS) solution, which corresponds to the case when all the regularization parameters disappear,

$$x_{j,OLS}^{k+1} = \left(x_j^k + \frac{\sum_{i=1}^{m} a_{ij}(b_i - (A\mathbf{x}^k)_i)}{\kappa_j} \right)_+, \tag{11}$$

where $u_+ = \max(0, u)$, representing the positive part of any function u. In matrix form, we have

$$\mathbf{x}_{OLS}^{k+1} = \left(\mathbf{x}^k + \frac{1}{\kappa} A^t(\mathbf{b} - A\mathbf{x}^k) \right)_+; \tag{12}$$

where division by κ is in Hadamard sense or entry-wise.

(ii) We obtain the non-negative regularized solution:

$$x_{j,sp}^{k+1} = (x_{OLS}^{k+1} - \frac{\lambda_{sp}}{\kappa_j})_+, \tag{13}$$

where

$$\lambda_{sp} = \begin{cases} \lambda_1, & \text{if } \lambda_1 \neq 0; \\ \frac{\lambda_{log}}{x_j^k + \delta_{log}}, & \text{if } \lambda_{\log} \neq 0; \\ \frac{q\lambda_q}{(x_j^k + \delta_q)^{1-q}}, & \text{if } \lambda_q \neq 0. \end{cases} \tag{14}$$

2.5. Selection of regularization parameters and image quality metrics

For each type of regularization, we identify the best image that can be reconstructed by searching through a range of values for the regularization parameter. There are different criteria in optical tomography for this selection. 11 methods have been compared (Correia et al. 2009) for finding the best Tikhonov regularization parameter in image deblurring—as a first step to optical tomography, and the L-curve method was found to be the best. Instead of relying on any single criterion in comparing image qualities, we base our comparison on a mixture of metrics since our goal is to reconstruct the localized targets, which are expected to be both sparse and correctly positioned. These metrics

include: the volume ratio (VR) (Tian et al. 2009), which measures the ratio of the reconstructed target volume to the true target volume and also indicates the sparsity the reconstructed target; the dice similarity coefficient (Dice) (Dice 1945), which measures the location accuracy of the reconstructed objects; the mean squared error (MSE), which measures the error the reconstructed and the true fluorophore concentrations; and the contrast-to-noise ratio (CNR) (Song et al. 2004), which measures if the reconstructed target can be clearly distinguished from the background. Their definitions are as follows:

$$VR = \frac{|rROI|}{|ROI|}, \tag{15}$$

$$Dice = \frac{2 * |rROI \cap ROI|}{|rROI| + |ROI|}, \tag{16}$$

$$MSE = \frac{1}{N}\sum_{j=1}^{N}(x_j - x_{0j})^2, \tag{17}$$

$$CNR = \frac{Mean(x_{_{ROI}}) - Mean(x_{_{ROB}})}{\sqrt{\omega_{ROI} Var(x_{_{ROI}}) + (1-\omega_{ROI}) Var(x_{_{ROB}})}} \tag{18}$$

where x and x_0 are the reconstructed and true fluorophore concentrations respectively, $rROI$ the reconstructed region of interest that is defined to be the voxels whose concentrations are higher than 50% of the maximum of the reconstructed concentrations, ROI the true region of interest or the true target locations, $\omega_{ROI} = |ROI|/(|ROI| + |ROB|)$, ROB the true background region, $|\cdot|$ number of elements, 'Mean' and 'Var' mean and variance respectively. Generally the closer VR and Dice are to 1, the smaller MSE is, and the larger CNR is, the better. In this paper, we especially focus on the VR and Dice values since they measure sparsity and accuracy of target positions.

2.6. Numerical simulations

To validate our algorithm, we simulated two fluorescent tube sources inside a mouse. The mouse was simulated based on the surface mesh of the Digimouse (Dogdas et al. 2007), from which we used Tetgen (http://tetgen.org) to regenerate a uniform internal mesh with a total of 32,332 nodes and 161,439 tetrahedral elements since the internal tetrahedral elements from Digimouse is not uniform in size. We then simulated two tubes at the center of the trunk of the mouse with diameters 2mm and lengths 20mm. For all the nodes inside the two tubes, we assigned the fluorophore concentration to be 1; and outside, 0. Then we chose 5 rings around the trunk, and selected 12 uniformly distributed nodes from each ring to form a total of 60 laser excitations sources. All the 4020 surface nodes that cover the trunk were set to be detectors. Finally, uniform white noise with a signal-to-noise ratio of 1 was added to each measurement. We assumed the excitation wavelength of 650 nm and the emission wavelength of 700nm. The tissue optical properties were $\mu_a = 0.007\,\mathrm{mm}^{-1}$, $\mu_s' = 0.72\,\mathrm{mm}^{-1}$ at 650nm and $\mu_a = 0.014\,\mathrm{mm}^{-1}$, $\mu_s' = 0.78\,\mathrm{mm}^{-1}$ at 700nm.

Table 1. Metrics of reconstructed images of simulated data under different (a) L^1 and (b) $L^{1/2}$ regularizations, with the best selections highlighted in bold.

(a)

λ_q	VR	Dice	MSE	CNR
6.2E-4	6.95	0.207	3.841E-3	4.25
6.3E-4	6.95	0.204	3.843E-3	4.24
6.4E-4	6.91	0.205	3.845E-3	4.24
6.5E-4	**6.84**	**0.204**	**3.847E-3**	**4.23**
6.6E-4	6.81	0.205	3.849E-3	4.22
6.7E-4	6.77	0.20	3.852E-31	4.21
6.8E-4	6.72	0.19	3.854E-39	4.21

(b)

λ_1	VR	Dice	MSE	CNR
3.2E-5	4.96	0.261	3.51E-3	4.80
3.4E-5	4.33	0.261	3.54E-3	4.65
3.6E-5	4.02	0.260	3.56E-3	4.55
3.8E-5	4.03	0.255	3.60E-3	4.38
4.0E-5	**3.74**	**0.258**	**3.64E-3**	**4.31**
4.2E-5	3.44	0.232	3.77E-3	3.84
4.4E-5	3.20	0.231	3.82E-3	3.65

2.7. Phantom experiment

We further designed a cubic phantom with dimension 32 mm by 32 mm by 29 mm that was composed of 1% intralipid, 2% agar, and water in the background; whereas in the two capillary target tubes with length 12 mm and diameter 1 mm, 6.5 μm DiD fluorescence dye solution was injected. The two target tubes also contained uniformly distributed ^{18}[F]-fluoro-2-deoxy-D-gluocose (FDG) at activity level of 100 μCi for a simultaneous positron emission tomography (PET) scan, whose result will be used to validate our FMT. The FEM mesh consists of 8690 nodes and 47,581 tetrahedral elements. The excitation laser at a wavelength of 650 nm scanned the front surface of the phantom at 20 illumination nodes. The excitation wave length was 700nm. The tissue optical properties were $\mu_a = 0.0022\,\mathrm{mm}^{-1}$, $\mu_s' = 1.10\,\mathrm{mm}^{-1}$ at both 650nm and 700nm wavelengths. Measurements were collected at 1057 detector nodes for the FMT. We threshold the PET images at 20% of the maximum FDG concentrations to identify the positions of the capillary tubes. For more details of the PET imaging, please refer to (Li et al. 2009b).

3. RESULTS

3.1. Simulation results

We first show some examples on how the optimization of our regularization parameters are done. We will only show the results for L^1 and $L^{1/2}$ cases since we are aiming at comparing our nonconvex regularizers—where $L^{1/2}$ is the best as we will show later—with the L^1 regularizer. Other cases in simulations are all done in the same fashion. figure 2 (a) and (b) show how λ_1 =6.5E-4 and λ_q =4E-5 provide the best balance across the four metrics and will be chosen as the best regularization parameter for L^1 and $L^{1/2}$ respectively. We clearly see in table 1 that for both cases, the selected regularization parameters are associated with relatively low VR, high Dice, while the MSE and CNR sit right at their average levels.

The reconstructed images using different regularizations are shown in figure 3. Note

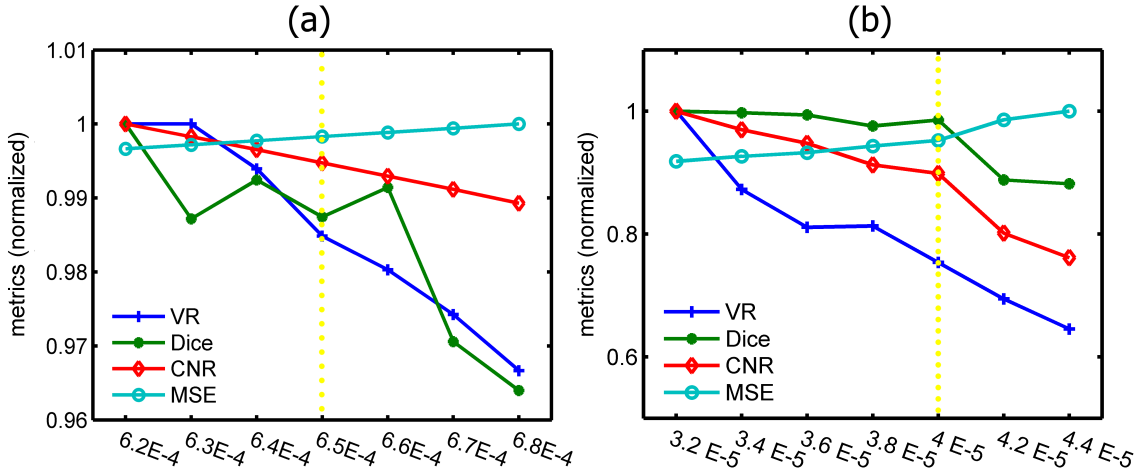

Figure 2. Comparison of effects of different regularization parameters for (a) L^1 and (b) $L^{1/2}$. The metrics are normalized by their maximal values respectively. The true values are listed in Table 1. The best regularization parameters are λ_1 =6.5E-4 and λ_q =4E-5 respectively, as highlighted by the yellow dotted lines.

that result from L^2 regularization is also included as a reference. We find that the images (d)-(g) corresponding to nonconvex regularization methods are more sparse than the images from the well known L^1 (c) and L^2 (b). To further analyze the images quantitatively, the corresponding image quality metrics, VR, Dice, MSE and CNR are calculated and shown in table 2. As expected, L^1 regularization demonstrated a better performance than L^2 in terms of providing a more sparse (smaller VR) and more accurately positioned targets (larger Dice), although it yielded a slightly larger MSE and smaller CNR. More importantly, we can see that $L^q, q = 7/8, 1/2, 1/8$ and *Log* regularizations produce even better results than L^1: the VR's are smaller and Dice's are larger, and the associated MSE and CNR values are also slightly improved.

Table 2. Metrics of the best reconstructed images under different regularizations for simulated data, with the best regularization method highlighted in bold.

Regularizations	VR	Dice	MSE	CNR
L^2, λ_2 =5.0E-5	9.64	0.18	3.77E-3	4.32
L^1, λ_1 = 6.5E-4	6.84	0.20	3.85E-3	4.23
$L^{7/8}, \lambda_q$ =6.1E-4	4.21	0.23	3.83E-3	3.93
$\mathbf{L^{1/2}, \lambda_q}$ **=4.0E-5**	**3.74**	**0.26**	**3.64E-3**	**4.31**
$L^{1/8}, \lambda_q$ =3.0E-6	4.81	0.26	3.59E-3	4.43
Log, λ_{log} =1.3E-5	5.06	0.26	3.63E-3	4.29

Among all the nonconvex regularizations, we found that $L^{1/2}$ produces the best result. It yields the most sparse solution without sacrificing much on Dice, CNR, and MSE values. We don't have a theoretical accounting for this phenomenon yet but we believe this is related to the fact that when q is too small or as in the *Log* case, the regularization emphasizes too much on the sparsity of the solution, whereas the

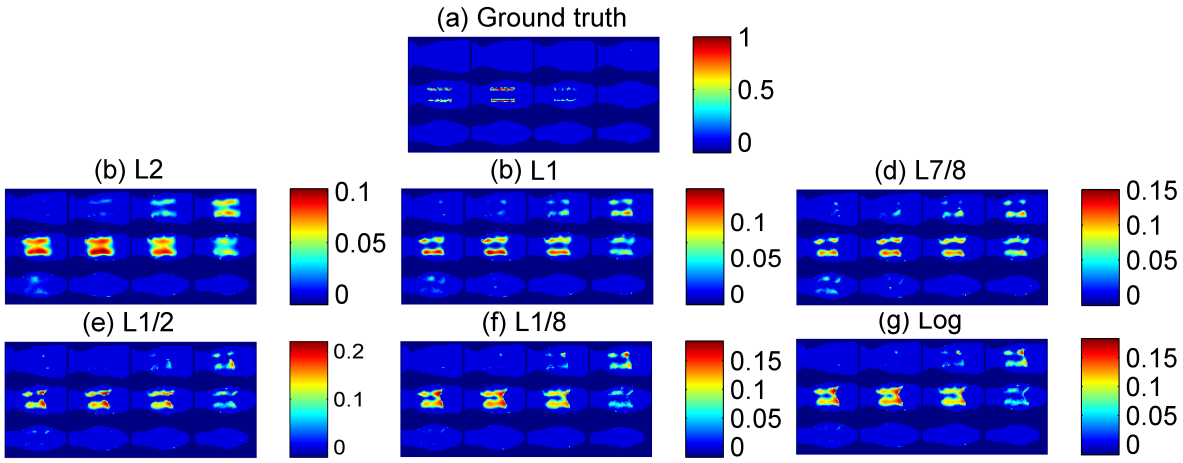

Figure 3. Comparison of (a) the ground truth image, and reconstructed images using MM algorithm with regularization (b) L^2 and λ_2 =5.0E-5, (c) L^1 and λ_1 =6.5E-4, (d) $L^{7/8}$ and $\lambda_{7/8}$ =6.1E-4, (e) $L^{1/2}$ and $\lambda_{1/2}$ =4.0E-5, (f) $L^{1/8}$ and $\lambda_{1/8}$ =3.0E-6, and (g) *Log* and λ_{log} =1.3E-5, respectively.

accuracy of the target locations is compromised. As pointed out earlier, when we select the best regularization parameter for each regularization model, we need to choose the case where the Dice is relatively high, the associated VR of which may also be larger. We also noticed that when q is small, L^q and *Log* performed similarly, which is not surprising since the shape of the curve of *Log* and L^q for small q are quite similar, as seen in figure 1.

3.2. Phantom experiment result

Selection of regularization parameters for phantom experiments are done in a similar way as the numerical simulations. The phantom results are summarized in figure 4 and table 3. We included the image reconstructed from PET as the truth. Since we don't have the true image intensity information available, we can only compute the VR, Dice and CNR metrics. From figure 4, we also clearly see the pattern we identified in the simulation section, that is, (1) for L^q, $0 < q < 1$ and *Log* regularizations, more sparse and accurate images are reconstructed than in L^1 and L^2; and (2) images (e)-(g) are very similar to each other, although (e) corresponding to $L^{1/2}$ regularization has the most sparse structure. Quantitatively, the best performance in terms of VR and Dice are also observed for $L^{1/2}$ as shown in table 3 although $L^{1/8}$ and *Log* yield higher CNR's.

4. DISCUSSION AND CONCLUSION

We introduced nonconvex regularizations into the reconstruction of FMT and validated the new methods with both numerical simulations and phantom experiments. The MM algorithm was adopted for a straightforward iterative update. We started the iterative algorithm from a uniform initial value that can be randomly chosen on interval $(0, 1)$.

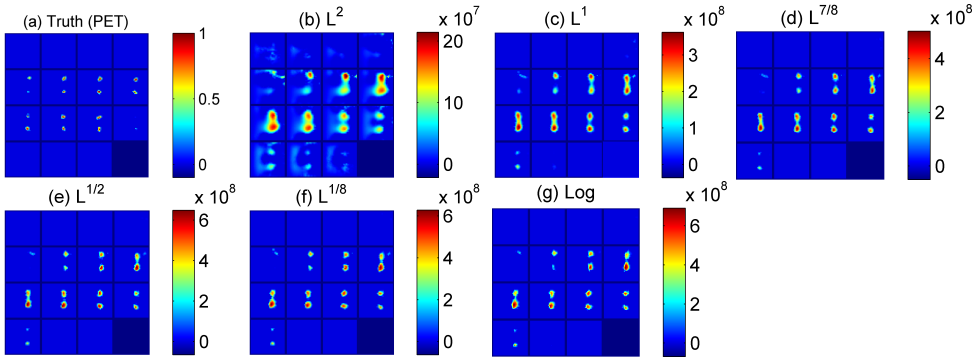

Figure 4. Reconstructed images for phantom experiment: (a) PET result as truth, (b) using MM algorithm with regularization L^2 and λ_2 =3.0E-6, (c) L^1 and λ_1 =9.0E+3, (d) $L^{7/8}$ and $\lambda_{7/8}$ =1.0E+5, (e) $L^{1/2}$ and $\lambda_{1/2}$ =5.0E+7, (f) $L^{1/8}$ and $\lambda_{1/8}$ =1.0E+10, (g) and *Log* and λ_{log} =9.0E+11, respectively.

Table 3. Metrics of the best reconstructed images under different regularizations for phantom data, with the best regularization method highlighted in bold.

Regularizations	VR	Dice	CNR
L^2, λ_2 =3.0E-6	7.64	0.19	4.44
L^1, λ_1 =9.0E+3	4.33	0.26	6.11
$L^{7/8}, \lambda_q$ =1.0E+5	3.17	0.28	6.33
$\mathbf{L^{1/2}, \lambda_q}$ **=5.0E+7**	**2.56**	**0.39**	**5.11**
$L^{1/8}, \lambda_q$ =1.0E+10	3.13	0.35	7.11
Log, λ_{log} =9.0E+11	2.78	0.36	6.88

We ran a total of 2000 steps for both the simulated and phantom data. For the simulated data, each iteration takes about 1.6 seconds to finish on an Intel i5 2400 3.1GHz PC and the incremental update became less than 2E-5 when the iteration stopped. For the phantom, each iteration takes only about 0.09 seconds since its mesh is much smaller. We used four different image quality metrics, VR, Dice, MSE and CNR to analyze the reconstructed images. In both numerical simulations and phantom experiment, the comparison between the nonconvex methods and the well known L^1 and L^2 methods demonstrated that nonconvex L^q $(0 < q < 1)$ and *Log* regularizations can reconstruct more sparse images while keeping targets more accurately positioned, as indicated by the smaller VR and larger Dice values than those from L^1 and L^2 regularizations. The new methods are robust since almost identical reconstruction results were obtained from six simulated noisy data sets, each corrupted by white Gaussian noise with SNR 1.

The MM algorithm that we adopted guarantees monotonicity, yet one of its potential drawback is that it tends to have very slow convergence (Hunter & Lange 2004, Dutta et al. 2012). So the 2000 max iteration steps may still not guarantee that the result we obtained using our convex surrogate be the global minimize, especially when the FMT system is ill-posed and ill conditioned. The reconstructed images look satisfactory overall; however, the VR values for the reconstructed images are still too large (> 3.7 and > 2.5 for simulations and phantom data, respectively), far away from

the ideal value, and the Dice similarity coefficients are also quite low (< 0.3 and < 0.4 for simulations and phantom data, respectively). We noticed that combination of MM and PCG can increase the convergence (Dutta et al. 2012), and different surrogate functions lead to different convergence rates (Zhou et al. 2011) as well. So we may follow these directions and hopefully improve the convergence of our MM algorithm in FMT.

We are also planning to validate our algorithms with *in vivo* experiments, which involve more factors such as different noise patterns and background signals due to autofluorescence etc. (Ale et al. 2013). So we may have to combine other approaches to make sure the targets are indeed sparse before applying our algorithm to the *in vivo* data.

In conclusion, we presented a nonconvex framework for reconstructing localized FMT targets and it is more effective in enhancing sparsity than the well known L^1 regularization method.

Acknowledgments

The authors wish to thank the three anonymous referees for their helpful suggestions in improving the quality of this paper. We would also like to thank Dr. Guobao Wang from the Department of Biomedical Engineering, UC Davis for useful discussions. Finally, this work is supported by the UC Merced Start-up Fund.

References

Ale A, Ermolayev V, Deliolanis N C & Ntziachristos V 2013 Fluorescence background subtraction technique for hybrid fluorescence molecular tomography/x-ray computed tomography imaging of a mouse model of early stage lung cancer *Journal of biomedical optics* **18**(5), 056006–056006.

Baritaux J C, Hassler K & Unser M 2010 An efficient numerical method for general l_p regularization in fluorescence molecular tomography *Medical Imaging, IEEE Transactions on* **29**(4), 1075–1087.

Candes E J, Wakin M B & Boyd S P 2008 Enhancing sparsity by reweighted l_1 minimization *Journal of Fourier Analysis and Applications* **14**(5-6), 877–905.

Chartrand R 2007 Exact reconstruction of sparse signals via nonconvex minimization *Signal Processing Letters, IEEE* **14**(10), 707–710.

Cherry S R 2004 In vivo molecular and genomic imaging: new challenges for imaging physics *Physics in medicine and biology* **49**(3), R13.

Correia T, Gibson A, Schweiger M & Hebden J 2009 Selection of regularization parameter for optical topography *Journal of biomedical optics* **14**(3), 034044–034044.

Dice L R 1945 Measures of the amount of ecologic association between species *Ecology* **26**(3), 297–302.

Dogdas B, Stout D, Chatziioannou A F & Leahy R M 2007 Digimouse: a 3d whole body mouse atlas from ct and cryosection data *Physics in medicine and biology* **52**(3), 577.

Dutta J, Ahn S, Li C, Cherry S R & Leahy R M 2012 Joint l^1 and total variation regularization for fluorescence molecular tomography *Physics in medicine and biology* **57**(6), 1459.

Erdogan H & Fessler J A 1999 Ordered subsets algorithms for transmission tomography *Physics in medicine and biology* **44**(11), 2835.

Fedele F, Laible J & Eppstein M 2003 Coupled complex adjoint sensitivities for frequency-domain fluorescence tomography: theory and vectorized implementation *Journal of Computational Physics* **187**(2), 597–619.

Fessler J A, Ficaro E P, Clinthorne N H & Lange K 1997 Grouped-coordinate ascent algorithms for penalized-likelihood transmission image reconstruction *Medical Imaging, IEEE Transactions on* **16**(2), 166–175.

Figueiredo M A, Bioucas-Dias J M & Nowak R D 2007 Majorization–minimization algorithms for wavelet-based image restoration *Image Processing, IEEE Transactions on* **16**(12), 2980–2991.

Gasso G, Rakotomamonjy A & Canu S 2009 Recovering sparse signals with a certain family of nonconvex penalties and dc programming *Signal Processing, IEEE Transactions on* **57**(12), 4686–4698.

Han D, Tian J, Zhu S, Feng J, Qin C, Zhang B & Yang X 2010 A fast reconstruction algorithm for fluorescence molecular tomography with sparsity regularization *Optics Express* **18**(8), 8630–8646.

Hunter D R & Lange K 2004 A tutorial on mm algorithms *The American Statistician* **58**(1), 30–37.

Li C, Mitchell G S, Dutta J, Ahn S, Leahy R M & Cherry S R 2009a A three-dimensional multispectral fluorescence optical tomography imaging system for small animals based on a conical mirror design *Optics express* **17**, 7571-7585.

Li C, Wang G, Qi J & Cherry S R 2009b Three-dimensional fluorescence optical tomography in small-animal imaging using simultaneous positron-emission-tomography priors *Optics letters* **34**(19), 2933–2935.

Natarajan B K 1995 Sparse approximate solutions to linear systems *SIAM journal on computing* **24**(2), 227–234.

Ntziachristos V, Ripoll J, Wang L V & Weissleder R 2005 Looking and listening to light: the evolution of whole-body photonic imaging *Nature biotechnology* **23**(3), 313–320.

Okawa S, Hoshi Y & Yamada Y 2011 Improvement of image quality of time-domain diffuse optical tomography with l_q sparsity regularization. *Biomedical optics express* **2**(12), 3334.

Prakash J, Shaw C, Manjappa R, Kanhirodan R & Yalavarthy P K 2014 Sparse recovery methods hold promise for diffuse optical tomographic image reconstruction. *IEEE journal of slected topics in quantum electronics (Issue on biophotonics)* **20**, 06800609.

Sidky E Y, Chartrand R & Pan X 2007 Image reconstruction from few views by non-convex optimization *in* 'Nuclear Science Symposium Conference Record, 2007. NSS'07. IEEE' Vol. 5 IEEE pp. 3526–3530.

Song X, Pogue B W, Jiang S, Doyley M M, Dehghani H, Tosteson T D & Paulsen K D 2004 Automated region detection based on the contrast-to-noise ratio in near-infrared tomography *Applied optics* **43**(5), 1053–1062.

Tian F, Alexandrakis G & Liu H 2009 Optimization of probe geometry for diffuse optical brain imaging based on measurement density and distribution *Applied optics* **48**(13), 2496–2504.

Wang G, Li C, Cherry S R & Qi J 2011 Statistical image reconstruction for hybrid fluorescence optical tomography and positron emission tomography *in* 'Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on' IEEE pp. 488–491.

Zhou H, Alexander D & Lange K 2011 A quasi-newton acceleration for high-dimensional optimization algorithms *Statistics and computing* **21**(2), 261–273.