

UC Merced

UC Merced Previously Published Works

Title

Accelerated image reconstruction in fluorescence molecular tomography using a nonuniform updating scheme with momentum and ordered subsets methods

Permalink

<https://escholarship.org/uc/item/7pc2r1xk>

Journal

Journal of Biomedical Optics, 21(1)

ISSN

1083-3668

Authors

Zhu, Dianwen
Li, Changqing

Publication Date

2016-01-13

DOI

10.1117/1.jbo.21.1.016004

Peer reviewed

Accelerated Image Reconstruction in Fluorescence Molecular Tomography using a Non-Uniform Updating Scheme with Momentum and Ordered Subsets Methods

Dianwen Zhu^a, Changqing Li^{a,*}

^aSchool of Engineering, University of California, Merced, 5200 N Lake Rd, Merced, CA, USA

Abstract. Fluorescence molecular tomography (FMT) is a significant preclinical imaging modality that has been actively studied in the past two decades. It remains a challenging task to obtain fast and accurate reconstruction of fluorescent probe distribution in small animals due to the large computational burden and the ill-posed nature of the inverse problem. We have recently studied a non-uniform multiplicative updating algorithm that combines with the ordered subsets (OS) method for fast convergence. However, increasing the number of OS leads to greater approximation errors and the speed gain from larger number of OS is limited. In this paper, we propose to further enhance the convergence speed by incorporating a first order momentum method that uses previous iterations to achieve optimal convergence rate. Using numerical simulations and a cubic phantom experiment, we have systematically compared the effects of the momentum technique, the OS method, and the non-uniform updating scheme in accelerating the FMT reconstruction. We found that the proposed combined method can produce a high quality image using an order of magnitude less time.

Keywords: Fluorescence Molecular Tomography, Iterative Image Reconstruction, Ordered Subsets, Momentum, Non-uniform update, MM algorithm..

*Corresponding Author, cli32@ucmerced.edu

1 Introduction

Fluorescence molecular tomography (FMT) has been an important tool for preclinical imaging in the past two decades and has attracted a lot of research interests.^{1,2} In FMT, fluorescent agents (e.g. fluorophores) are injected into the object, such as small animals. Upon illumination by near-infrared laser beams on the object surface, the fluorophores then become excited and start emitting fluorescence photons, which propagate out of the object surface and are captured by detectors such as charge-coupled device (CCD) cameras. Due to the high scattering effects of photons in tissues, the FMT system matrix is usually ill-conditioned and the reconstruction problem ill-posed. Immense efforts have been devoted to experimentally alleviating the ill-posedness, such as collecting more measurement data by using dense noncontact CCD cameras, employing multi-spectral wavelengths for both excitation and emission,³⁻⁵ and using different illumination patterns etc.⁶⁻⁸ These

indeed improved the image resolutions. However, they led to the processing of large amounts of 3D data, which requires tremendous computational time. From the theoretical aspect, a plethora of nonlinear regularization methods have also been applied,⁹⁻¹⁷ which again have caused additional complexity in computations. The non-negativity constraint in FMT reconstruction is another challenge. This constraint has been typically handled by a backtracking line search strategy,¹⁴ which is time consuming for large-scale measurement data.

To reduce the computational burden, many ideas from other fields have been introduced to FMT. Fourier transform,¹⁸ wavelet transform methods,^{19,20} discrete cosine transform,²¹ and principal component analysis methods²² etc. have been adopted to compress the data and reduce its dimension. The ordered subsets (OS) method,²³ originally introduced to speed up emission/transmission tomography by breaking down large system matrix into smaller blocks, has recently been successfully adopted into FMT by employing the majorization-minimization (MM) framework.^{14,17,24} In contrast with other methods that need a slow line search in handling the non-negativity constraint, the MM framework has the advantage of reducing the optimization into separable one dimensional problems and hence the non-negativity constraint can be enforced straightforwardly in a parallel way.²⁵ In particular, the non-uniform multiplicative MM algorithm with OS acceleration (NUMOS) that we proposed recently has been found to significantly improve the FMT image reconstruction speed and quality.²⁶ Nevertheless, the MM algorithm is a first-order method and typically needs a lot of iterations to converge.

Another direction in reducing the computational cost is to improve the convergence rate of the FMT reconstruction algorithms. Compared with first-order methods, second-order methods such as Newton-type algorithms have faster convergence, however, their computation of Hessian is very demanding. Recently a class of momentum acceleration techniques from Nesterov have

been found to achieve optimal convergence rate for a class of gradient descent based methods whose computational complexities are at linear level.^{27,28} In FMT, Han et al. applied recently¹¹ the fast iterative shrinkage thresholding algorithm (FISTA),²⁹ which is based on the 1983 version of the Nesterov's technique.²⁷ Momentum methods have been combined with the OS techniques for an even faster convergence in computed tomography (CT) very recently.³⁰

In this paper, we propose to include the 2005 version of Nesterov's momentum method²⁸ into our NUMOS algorithm for a faster and stable convergence since the 1983 version was not as stable.³⁰ The concept and some preliminary results of this paper have appeared earlier in a conference paper.³¹ Note that our work is different from the aforementioned work in CT in two ways: First, the regularization they used is an edge-preserving Huber type of function that promotes local smoothness whereas our model for FMT uses L^1 to promote sparsity; second, their surrogate functions are either a spatially uniform type or a non-uniform type that depends on a good estimate of the true solution, which is readily available in CT through the filtered back-projection reconstruction,³⁰ but not in FMT.

The rest of the paper is organized as follows. In Section 2, we first introduce the background of FMT forward modeling, regularized reconstruction, the MM framework, the ordered subsets technique, and the Nesterov's momentum method. Then we propose our fast NUMOS (fNUMOS) algorithm. In Section 3, we present both numerical simulation and phantom experiment results and comparison with other methods. In Section 4, we conclude the paper with summary and some discussions.

2 Methods

2.1 Forward Modeling and Regularized FMT Reconstruction

For FMT in the continuous wave domain, photon transfer is modeled by the following coupled diffusion equations, along with Robin type (mixed) boundary conditions:

$$\left\{ \begin{array}{l} -\nabla \cdot (D_{ex}(\mathbf{r})\nabla\Phi_{ex}(\mathbf{r})) + \mu_{a,ex}(\mathbf{r})\Phi_{ex}(\mathbf{r}) = \Delta_s(\mathbf{r}), \\ \mathbf{n} \cdot (D_{ex}(\mathbf{r})\nabla\Phi_{ex}(\mathbf{r})) + \alpha_{ex}\Phi_{ex}(\mathbf{r}) = 0, \\ -\nabla \cdot (D_{em}(\mathbf{r})\nabla\Phi_{em}(\mathbf{r})) + \mu_{a,em}(\mathbf{r})\Phi_{em}(\mathbf{r}) = \Phi_{ex}(\mathbf{r})\Delta_d(\mathbf{r}), \\ \mathbf{n} \cdot (D_{em}(\mathbf{r})\nabla\Phi_{em}(\mathbf{r})) + \alpha_{em}\Phi_{em}(\mathbf{r}) = 0, \end{array} \right. \quad (1)$$

where ∇ denotes the gradient operator, $D_{ex}(\mathbf{r}) = [3(\mu'_{s,ex}(\mathbf{r}) + \mu_{a,ex}(\mathbf{r}))]^{-1}$ and $D_{em}(\mathbf{r}) = [3(\mu'_{s,em}(\mathbf{r}) + \mu_{a,em}(\mathbf{r}))]^{-1}$, with $\mu_{a,ex}(\mathbf{r})$, $\mu_{a,em}(\mathbf{r})$ being the absorption coefficients and $\mu'_{s,ex}(\mathbf{r})$, $\mu'_{s,em}(\mathbf{r})$ being the reduced scattering coefficients at excitation and emission wavelengths, $\Phi_{ex}(\mathbf{r})$, $\Phi_{em}(\mathbf{r})$ the photon densities, \mathbf{r} the location vector, Δ_s determined by the s^{th} illumination pattern (for example, point sources $\delta_s(\mathbf{r} - \mathbf{r}_s)$), Δ_d determined by the fluorophore yield at the d^{th} detector nodes, \mathbf{n} the outward unit normal vector of the boundary, and α_{ex} , α_{em} the Robin boundary coefficients.

The above equations can be discretized by the finite element method (FEM), leading to linearized equations:^{4,32}

$$\begin{aligned} [K_{ex}]\{\Phi_{ex}\} &= \{\delta_s(\mathbf{r} - \mathbf{r}_s)\} \\ [K_{em}]\{\Phi_{em}\} &= \{\Phi_{ex}\}\{\mathbf{x}\}. \end{aligned} \quad (2)$$

where K_{ex} and K_{em} are the stiffness matrices. Then we obtain

$$[\Phi_{em}] = [K_{em}^{-1}] \otimes [K_{ex}^{-1}][\mathbf{x}], \quad (3)$$

which, upon the removal of equations, corresponds to the unmeasurable nodes gives us

$$A\mathbf{x} = \mathbf{b}, \quad (4)$$

where $A = (a_{ij}) \in \mathbf{R}^{N_m \times N_n}$, $a_{ij} > 0$ is the system matrix, obtained by taking the tensor product \otimes of the sensitivity matrix $[K_{em}^{-1}]$ and the excitation matrix $[K_{ex}^{-1}]$, $\mathbf{x} = (x_j) \in \mathbf{R}^{N_n \times 1}$ the fluorophore distribution to be reconstructed, $\mathbf{b} = (b_i) \in \mathbf{R}^{N_m \times 1}$ the measurements, and N_m, N_n the total number of measurements and FEM mesh nodes, respectively.

A typical solution of (4) is obtained by minimizing the following regularized squared data-measurement misfit under the non-negativity constraint:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}, \mathbf{x} \geq 0} \Psi(\mathbf{x}) := \frac{1}{2} \|A\mathbf{x} - \mathbf{b}\|_2^2 + \lambda \|\mathbf{x}\|_1, \quad (5)$$

where λ is the L^1 regularization parameter and $\|\mathbf{x}\|_1 = \sum_{j=1}^{N_n} x_j$ represents the L^1 regularization when $x \geq 0$. Notice that there are other popular choices of the regularization function $R(x)$, including the L^p (semi-) norm: $R(\mathbf{x}) = \|\mathbf{x}\|_p^p$, $p \geq 0$.

2.2 MM Algorithm

The majorization-minimization algorithm, also known as optimization transfer algorithm, is a general framework for solving minimization problems where an approximation of the objective function, often referred to as the majorization or surrogate function, is minimized at every step. The approximated solution using MM algorithm will converge to the true solution of (5) as the problem is convex.³³ MM algorithm has known advantages in optimization problems including avoiding matrix inversions, linearizing an optimization problem, dealing gracefully with inequalities, etc.³³

Many powerful algorithms can be understood from the MM point of view, especially the gradient based methods such as iterative shrinkage thresholding algorithm (ISTA),³⁴ iteratively reweighted L^1 algorithm (IRL1),³⁵ and iteratively reweighted least squares (IRLS).³⁶

The definition of a surrogate function $\Psi^{sur}(\mathbf{x})$ in the minimization problem includes the following three requirements:

$$\begin{cases} \Psi^{sur}(\mathbf{x}) & \geq \Psi(\mathbf{x}), & \text{for all } \mathbf{x}; \\ \Psi^{sur}(\mathbf{x}^n) & = \Psi(\mathbf{x}^n), & \text{at some point } \mathbf{x}^n; \\ \nabla \Psi^{sur}(\mathbf{x}^n) & = \nabla \Psi(\mathbf{x}^n), & \text{at point } \mathbf{x}^n. \end{cases} \quad (6)$$

It is generally believed that in MM framework there is no optimal way of choosing the surrogate functions.²⁵ In the popular separable quadratic surrogates (SQS) algorithm,^{37,38} the surrogate function $\Psi^k(\mathbf{x}; \mathbf{x}^k)$ was chosen based on a Jensen inequality for least squares, which was first proposed by de Pierro.³⁹ Adding in the L^1 regularization, we have:

$$\begin{aligned} \Psi(\mathbf{x}) &= \frac{1}{2} \|\mathbf{b} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda \|\mathbf{x}\|_1 \\ &\leq \Psi^k(\mathbf{x}; \mathbf{x}^k) := \frac{1}{2} \sum_{i=1}^{N_m} \sum_{j=1}^{N_n} \beta_{ij} \left\{ b_i - (\mathbf{A}\mathbf{x}^k)_i - \frac{a_{ij}}{\beta_{ij}} (x_j - x_j^k) \right\}^2 + \lambda \sum_{j=1}^{N_n} x_j \\ &= \Psi(\mathbf{x}^k) - \sum_{j=1}^{N_n} \left(\sum_{i=1}^{N_m} (b_i - (\mathbf{A}\mathbf{x}^k)_i) * a_{ij} - \lambda \right) * (x_j - x_j^k) + \frac{1}{2} \sum_{j=1}^{N_n} \left(\sum_{i=1}^{N_m} \frac{a_{ij}^2}{\beta_{ij}} \right) (x_j - x_j^k)^2 \\ &= \Psi(\mathbf{x}^k) + \nabla \Psi(\mathbf{x}^k) \cdot (\mathbf{x} - \mathbf{x}^k) + \frac{1}{2} \|\mathbf{x} - \mathbf{x}^k\|_{D^k}^2. \end{aligned} \quad (7)$$

where $\beta_{ij} > 0$ with $\sum_{j=1}^{N_n} \beta_{ij} = 1$, $D^k = \text{diag}(d_j^k)$ with $d_j^k = \sum_{i=1}^{N_m} \frac{a_{ij}^2}{\beta_{ij}}$. The above surrogate

function can be minimized by choosing

$$x_j^{k+1} = (x_j^k - (\nabla \Psi(\mathbf{x}^k))_j / d_j^k)_+ \quad (8)$$

for each component x_j , where $(\cdot)_+ = \max(0, \cdot)$, representing the positive part of any function.¹⁷

This is a gradient descent type algorithm with different step size d_j for each component x_j .

2.3 Non-uniform Weighting Parameter β_{ij}

The choice of the weighting parameter β_{ij} for the Jensen inequality is crucial in determining how well the surrogate approximates the original objective function and also how fast the corresponding iterative algorithm (8) converges. Traditionally, the SQS was chosen in a uniform additive way, which can be precomputed and needs no iterative updates:^{14,17,37,38}

$$\beta_{ij}^A = \frac{a_{ij}}{\sum_{l=1}^{N_n} a_{il}}. \quad (9)$$

The correspondingly updating step size is

$$d_j = (A^t A \mathbf{1}_{N_n})_j, \quad (10)$$

where $\mathbf{1}_{N_n}$ is the N_n -dimensional vector with all entries equal 1, and the update formula (8) becomes:

$$x_{j,L^1}^{A,k+1} = \left(x_j^k + \frac{(A^t \mathbf{b})_j - (A^t A \mathbf{x}^k)_j - \lambda_1}{(A^t A \mathbf{1}_{N_n})_j} \right)_+. \quad (11)$$

Note the uniform additive type of β_{ij} has a more general form as follows:²⁵

$$\beta_{ij}^{gA} = \frac{a_{ij}^q}{\sum_{k=1}^{N_n} a_{ik}^q}, \quad q > 0. \quad (12)$$

Uniform weighting however, ignores the different updating needs at different positions, which could hinder the convergence of iterations to the true solution.⁴⁰ Non-uniform weighting strategy has been reported recently to improve image qualities in CT.^{40,41} In our previous work, we proposed to solve the L^1 regularized least squares optimization problem (5) by using the following non-uniform multiplicative type β_{ij} :^{26,38,42}

$$\beta_{ij}^M = \frac{a_{ij}x_j^k}{(A\mathbf{x}^k)_i}, \quad \text{where } (A\mathbf{x}^k)_i = \sum_{l=1}^{N_n} a_{il}x_l^k. \quad (13)$$

In comparison with the uniform additive β_{ij}^A that only needs to be calculated once, this non-uniform β_{ij} needs to be updated at each iteration. However, we noticed that with the non-uniform type of β_{ij}^M , we have²⁶

$$d_j^k = \frac{(A^t A\mathbf{x}^k)_j}{x_j^k}. \quad (14)$$

and the associated iterative update formula (8) becomes:

$$x_{j,L^1}^{k+1} = x_j^k \frac{((A^t \mathbf{b})_j - \lambda_1)_+}{(A^t A\mathbf{x}^k)_j} \quad (15)$$

which is much simpler and requires less calculations when compared with the additive form (9).

(15) also naturally promotes non-negativity and sparsity, since once $x_j^k = 0$, x_j remains 0.

2.4 Ordered Subsets Acceleration Technique

Due to the large scale of 3D data, tomographic imaging requires a lot of computational time. The ordered subsets technique was proposed in 1994 for emission/transmission tomography to evenly break down large matrices into smaller blocks so that a speedup of the convergence by a factor proportional to nOS , the number of subsets, is possible.^{23,38} In FMT, the OS technique has been successfully applied recently, where the sensitivity matrix has been divided into ordered subsets.^{14,26} In particular, combining the non-uniform update with this OS technique, we have proposed NUMOS (See Algorithm 1) for FMT, which has provided significant speed enhancement over the uniform methods,^{14,17} while maintaining high quality reconstruction results.²⁶

Algorithm 1 NUMOS

Initialization:

$$\mathbf{x}^0 = x_0 * \mathbf{1}_n, B = (A^t b - \lambda/nOS * \mathbf{1}_{N_n})_+$$

for $k = 1$ to N_{max}/nOS **do**

Divide A and B into nOS sub-matrices, $\{A_i\}_{i=1}^{nOS}$, $\{B_i\}_{i=1}^{nOS}$, based on a random partition of the N_d detectors

for $i = 1$ to nOS **do**

$$l = (k - 1) * nOS + i$$

$$\mathbf{x}^{k+1} = B_i .* \mathbf{x}^k ./ (A_i^t A_i \mathbf{x}^k)$$

end for

end for

where $0 < x_0 < 1$ is randomly picked, N_{max} is the number of iterations, nOS number of subsets, $.*$ and $./$ entry-wise multiplication and division respectively.

However, there are a few known issues with the OS technique. One is that the selection of ordered subsets can be complicated when the geometry is complicated. In the study of mouse shaped numerical phantom, we balanced our selection of subsets by randomly generating them at each iteration, which slowed down the acceleration from the OS technique.²⁶ Another issue with this OS technique is that, like other block-iterative methods, increasing the number of OS leads to larger approximation errors, causing the convergence of iterative updates to stop at a limit-

cycle before approaching the minimum.⁴³ Different approaches have been proposed to address this limit-cycle issue, but again at the expense of slowing down the overall convergence.^{44–46}

2.5 Nesterov’s Momentum Acceleration Techniques

Another approach to reduce the computation burden is to design improved algorithms with faster convergence rates. Recently, a series of momentum techniques of Nesterov^{27,28} have emerged, using previous iterations to obtain optimal convergence rates for first order optimization methods. In particular, based on the 1983 version of the Nesterov’s technique, the algorithm FISTA²⁹ was developed, the application of which in FMT will be analyzed in the Discussion section. We chose to focus on the 2005 version of the Nesterov’s technique²⁸ (See Algorithm 2) in this paper, since it provides more stability without requiring much extra calculations than the 1983 version.^{27,30} This 2005 Nesterov’s technique has attracted quite some attention recently.^{30,47} In particular, Kim et al. has considered the combination of this technique with the OS method in CT²⁶ for their separable quadratic surrogate function.³⁰

Algorithm 2 Nesterov’s algorithm (2005)

Initialization: $\mathbf{x}^0 = \mathbf{v}^0 = \mathbf{z}^0$, $t_0 = 1$
for $m = 1$ to N_{max} **do**
 $t^m = (m + 1)/2$
 $\mathbf{x}^m = (\mathbf{z}^{m-1} - \frac{1}{L} \nabla \Psi(\mathbf{z}^{m-1}))_+$
 $\mathbf{v}^m = (\mathbf{z}^0 - \frac{1}{L} \sum_{l=1}^m t^{l-1} * \nabla \Psi(\mathbf{z}^{l-1}))_+$
 $\mathbf{z}^m = (1 - \frac{t^m}{\sum_{l=0}^m t^l}) * \mathbf{x}^m + \frac{t^m}{\sum_{l=0}^m t^l} * \mathbf{v}^m$
end for

where L is the Lipschitz constant for the surrogate function of $\Psi(\mathbf{z})$ at \mathbf{z}^{m-1} .

2.6 Proposed Non-Uniform Multiplicative Updating Scheme Accelerated by Ordered Subsets and Momentum Method

The Nesterov's techniques are based on the gradient method, which from the MM algorithm point of view, is equivalent to using the following surrogate function for $\Psi(\mathbf{x})$:²⁹

$$\Psi^k(\mathbf{x}; \mathbf{x}^k) = \Psi(\mathbf{x}^k) + \nabla \Psi(\mathbf{x}^k) \cdot (\mathbf{x} - \mathbf{x}^k) + \frac{L}{2} \|\mathbf{x} - \mathbf{x}^k\|^2, \quad (16)$$

where L is the Lipschitz constant and is typically chosen to be the largest eigenvalue of the system matrix A , the computation of which is very challenging for large-scale problems.

The above choice of surrogate function (16) is essentially assigning every component x_j the same iterative step size, which does not take into account the different updating needs between target and background locations.⁴⁰ In this paper, we propose to employ the non-uniform update to satisfy the different updating needs and combine with OS method and the 2005 version of the Nesterov's momentum technique for much faster convergence. We refer to it as *fast NUMOS* (fNUMOS). (See Algorithm 3) Note that for fNUMOS, \mathbf{x}_{PM}^m is new term we have to introduce so that the nonuniform technique can be combined with the Nesterov's momentum technique. In addition, we followed the method of Tseng⁴⁸ to choose the weighting parameters t^m , which leads to a faster convergence than the original t^m as used by Nesterov in Algorithm 2.

Algorithm 3 fNUMOS

Initialization: $\mathbf{x}^0 = x_0 * \mathbf{1}_n$, $\mathbf{z}^0 = \mathbf{x}^0$, $t_0 = 1$, $B = A^t b - \lambda/nOS * \mathbf{1}_{N_n}$

for $k = 1$ to N_{max}/nOS **do**

Divide A and B into nOS sub-matrices, $\{A_i\}_{i=1}^{nOS}$, $\{B_i\}_{i=1}^{nOS}$, based on a random partition of the N_d detectors

for $i = 1$ to nOS **do**

$$m = (k - 1) * nOS + i$$

$$t^m = (1 + \sqrt{1 + 4(t^{m-1})^2})/2$$

$$\mathbf{x}_{PM}^m = B_i \cdot * \mathbf{z}^{m-1} ./ (A_i^t A_i \mathbf{z}^{m-1})$$

$$\mathbf{x}^m = (x_{PM})_+$$

$$\mathbf{v}^m = (\mathbf{z}_0 - \sum_{l=1}^m t^{l-1} * (\mathbf{x}_{PM}^l - z^{l-1}))_+$$

$$\mathbf{z}^m = (1 - \frac{t^m}{\sum_{l=0}^m t^l}) * \mathbf{x}^m + \frac{t^m}{\sum_{l=0}^m t^l} * \mathbf{v}^m$$

end for

where $0 < x_0 < 1$ is randomly picked, N_{max} is the number of iterations, nOS number of subsets, $*$ and $./$ entry-wise multiplication and division respectively.

end for

2.7 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 within $(0, \max(A^t b))$. Eleven different criteria in optical tomography have been compared in Correia et al.⁴⁹ and the L-curve method was found to be optimal for finding the best Tikhonov regularization parameter in image deblurring problems. In our study, we employ two criteria: the volume ratio (VR),⁵⁰ which is defined as the ratio of reconstructed target volume to true target volume and is related to the sparsity of the reconstructed targets, and the Dice similarity coefficient (Dice),⁵¹ which measures the location accuracy of the reconstructed objects. These criteria are sufficient to evaluate the sizes and positions of the targets. We also calculate the mean squared error (MSE) for the simulation study when ground truth is available, which measures the difference between reconstructed and true fluorophore concentrations. To access image qualities, we also computed the contrast-to-noise ratio (CNR),⁵² which measures how well the reconstructed target can be distinguished from the back-

ground. Their definitions are as follows:

$$VR = \frac{|rROI|}{|ROI|}, \quad (17)$$

$$Dice = \frac{2 * |rROI \cap ROI|}{|rROI| + |ROI|}, \quad (18)$$

$$MSE = \frac{1}{N} \sum_{j=1}^N (x_j - x_{0j})^2, \quad (19)$$

$$CNR = \frac{Mean(x_{ROI}) - Mean(x_{ROB})}{\sqrt{\omega_{ROI}Var(x_{ROI}) + (1 - \omega_{ROI})Var(x_{ROB})}} \quad (20)$$

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 reconstructed concentrations, ROI the true region of interest or the true target locations, ROB the true background region, $\omega_{ROI} = |ROI|/(|ROI| + |ROB|)$, and $|\cdot|$ number of elements. Generally the closer VR and Dice are to 1, the smaller MSE is the larger the CNR, the better. In this paper, we especially focus on the VR and Dice values since they measure sparsity and accuracy of target positions.

3 Results

3.1 Numerical Simulation

To validate our algorithm for FMT in small animals, we simulated a mouse by first obtaining the surface mesh of the Digimouse,⁵³ then using Tetgen (<http://tetgen.org>) to regenerate a uniform internal mesh with a total of 32,332 nodes and 161,439 tetrahedral elements. We then simulated two capillary tubes at the center of the mouse trunk, each having a diameter of 2 mm and length 20 mm. For simplicity, we assigned the fluorophore concentration to be 1 for all the nodes inside

the two tubes and 0 outside. We selected 60 internal nodes as laser source points, uniformly distributed on five rings around the trunk and we set all the 4020 surface nodes that cover the trunk to be detectors. The simulated tissue optical properties are $\mu_a = 0.007 \text{ mm}^{-1}$, $\mu'_s = 0.72 \text{ mm}^{-1}$ at both the excitation wavelength (650 nm) and the emission wavelength (700 nm). White Gaussian noise was added to the simulated measurement data so that the signal to noise ratio (SNR) of the measurement data was 1.

For the reconstruction, we employed the L^1 regularization since the targets are sparse. We started iterations from the same randomly picked uniform initial point \mathbf{x}_0 and reconstructed the fluorophore distribution using five different algorithms: (a) the uniform update, (b) non-uniform update, (c) NUMOS with $nOS = 24$, (d) fNUMOS with $nOS = 1$ and (e) fNUMOS with $nOS = 24$, so that we can see how each part of the proposed algorithm is making a difference toward a faster convergence, while maintaining a high quality result. For the uniform update, we chose the regularization parameter $\lambda_1 = 6.5\text{E-}4$, and for the non-uniform update, we set the regularization parameter to be $\lambda_1 = 1\text{E-}3$ as before.^{17,26} The maximum number of iterations is set to be 2000.

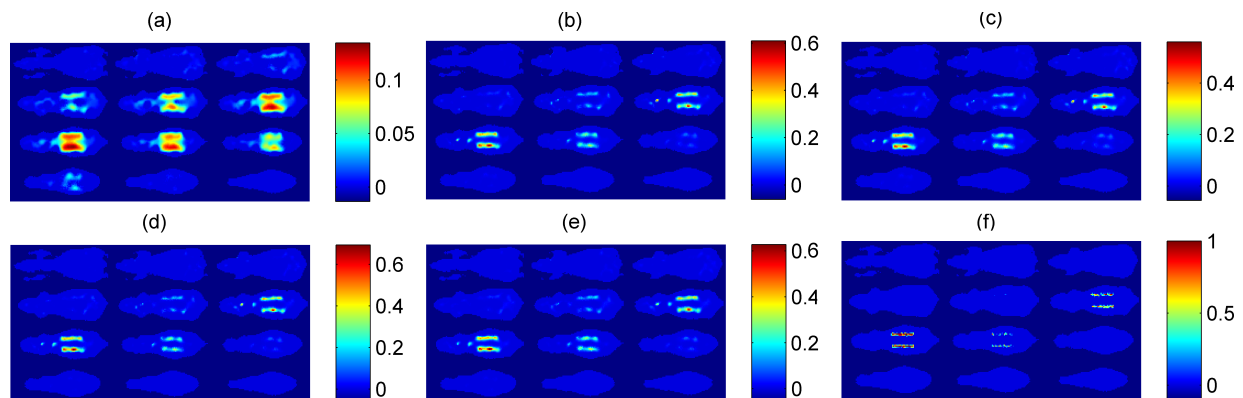


Fig 1 (Selected) Coronal slices of the reconstructed simulated mouse from bottom to top, using (a) Uniform method, Uniform1, at 2000 iterations, (b) NUMOS, $nOS = 1$, at 1310 iterations, (c) NUMOS with $nOS = 24$, at 53 iterations, (d) fNUMOS with $nOS = 1$, at 121 iterations, and (e) fNUMOS with $nOS = 24$ at 5 iterations. The truth image is shown in (f).

In Figure 1(a) and Table 1, we can see that even with 2000 iterations, the Uniform method is still

Table 1 Comparison of VR, Dice, CNR, MSE, Time (seconds) and number of Iterations for the FMT reconstruction of the simulated mouse using different algorithms.

Method	VR	Dice	CNR	MSE	Time (sec)	Iterations
Uniform	6.84	0.20	4.83	2.94E-3	563.46	2000
NUMOS1	1.01	0.58	9.94	1.74E-3	491.28	1310
NUMOS24	1.01	0.58	9.81	1.80E-3	106.97	53
fNUMOS1	1.02	0.59	9.54	1.69E-3	35.13	121
fNUMOS24	1.01	0.59	10.27	1.70E-3	10.02	5

far away from obtaining a close-to-truth result, the reconstructed targets are very large ($VR = 6.84$ and $Dice = 0.20$), with very low image intensities, relatively low CNR, and large MSE. We see a clear boost of image qualities from Figure 1(b) when we use the non-uniform update, i.e. NUMOS with $nOS = 1$. From Table 1, we see that NUMOS1 used only 1310 iterations to produce a much better result with VR about 7x smaller, Dice about 3x higher, CNR 2x higher, and MSE about 2x smaller. When non-uniform update is combined with $nOS = 24$ or the Nesterov’s technique, we see that less computation time is needed to obtain a similar quality image. In addition, we see that fNUMOS1 is around 3x faster than NUMOS24, using approximately 35 seconds and 107 seconds, respectively. This implies that the Nesterov’s momentum acceleration technique is more effective than the OS technique. Finally, when we combine both OS and Nesterov techniques with the non-uniform update, we needed only 5 iterations and 10 seconds to reach an image with high quality as shown in Figure 1(e), which implies a speed up of about 10x than the NUMOS algorithm we previously studied.²⁶

3.2 Phantom Experiment

Next we used a set of data from cubic phantom experiments to validate the acceleration effects of our proposed fNUMOS method for FMT. The cubic phantom was of dimension 32 mm by 32 mm by 29 mm and was composed of 1% intralipid, 2% agar, and water in the background.

We inserted two capillary tubes with length 12 mm and diameter 1 mm as targets, in which both 6.5 μm DiD fluorescence dye solution and uniformly distributed $^{18}\text{[F]}$ -fluoro-2-deoxy-D-glucose (FDG) at activity level of 100 μCi were injected. The cubic surface was extracted first and then the FEM mesh was generated, consisting of 8690 nodes and 47,581 tetrahedral elements.^{5,54} For the optical imaging, laser at a wavelength of 650 nm scanned the front surface of the phantom at 20 illumination nodes. Measurements were collected at 1057 detector nodes by using a conical mirror system and a CCD camera^{4,17} The filtered excitation wavelength was at 700nm. The tissue optical properties were $\mu_a = 0.0022 \text{ mm}^{-1}$, $\mu'_s = 1.10 \text{ mm}^{-1}$ at both 650 nm and 700 nm wavelengths. Details of the simultaneous PET imaging can be found in Li et al.^{4,5,54} We thresholded the PET images at 20% of the maximum FDG concentrations to identify positions of the capillary tubes.

For the regularized reconstruction, we also compared the results from the five different algorithms (as illustrated in Figure 2) and we empirically chose the L^1 regularization parameter λ_1 to be 9E+3 for the uniform methods and 5E+3 for the non-uniform algorithms. We set the maximum number of iterations to be $N_{max} = 768$ for non-uniform algorithms; whereas for the uniform cases, we allowed up to 5000 iterations for better results. We computed the VR, Dice and CNR for the reconstructed images. True locations of targets were obtained from the PET image that was acquired simultaneously. Note the associated intensity information of the true image was not available so we could not compute the MSE here.

From Table 2, we see that when we chose $nOS = 24$, fNUMOS reached a result with VR of 1.02, Dice of 0.42, and CNR of 8.64 after 4 iterations, taking only 0.63 seconds. To obtain an image of similar quality with the Nesterov's momentum technique only, fNUMOS for $nOS = 1$ took 84 iterations and 1.34 seconds; with the ordered subsets acceleration alone, NUMOS for $nOS = 24$ took 32 iterations and 4.85 seconds; without any of those two acceleration techniques, NUMOS

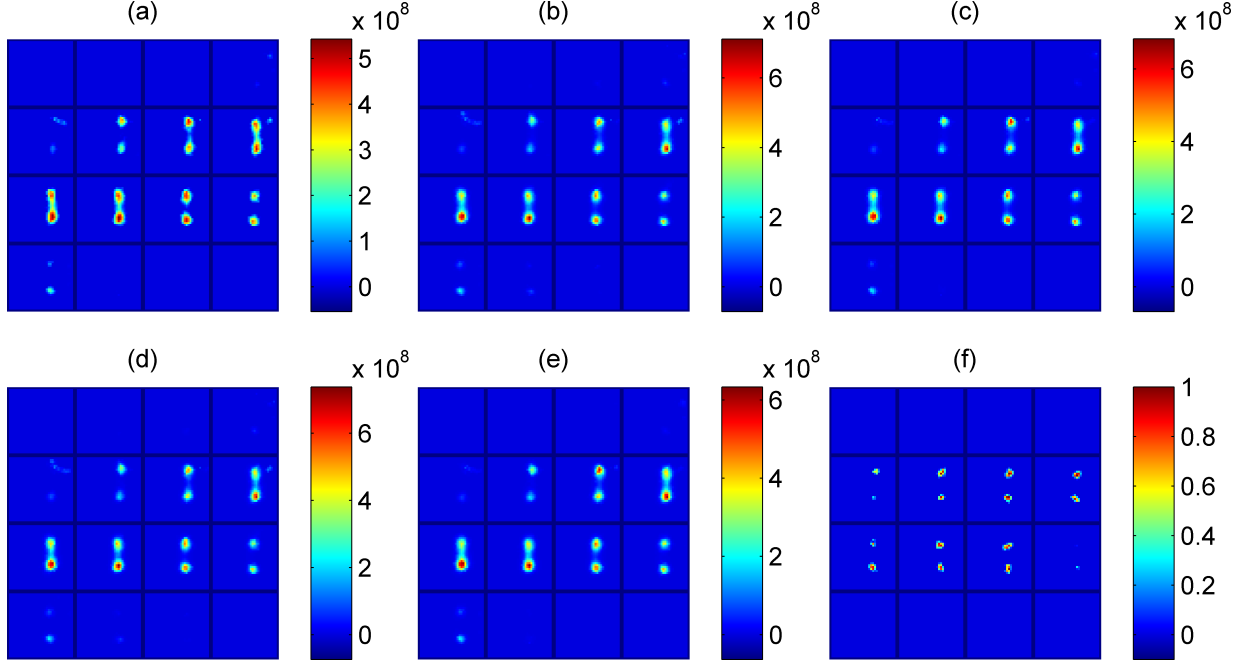


Fig 2 Each slice corresponds to a coronary section of the reconstructed cubic phantom from bottom to top, using (a) Uniform method, Uniform1, at 5000 iterations, (b) NUMOS, $nOS = 1$, at 640 iterations, (c) NUMOS with $nOS = 24$, at 32 iterations, (d) fNUMOS with $nOS = 1$, at 84 iterations and (e) fNUMOS with $nOS = 24$ at 4 iterations. The truth image is shown in (f).

Table 2 Comparison of VR, Dice, CNR, Time (seconds) and number of Iterations for the FMT reconstruction of cubic phantom using different algorithms.

Method	VR	Dice	CNR	Time (sec)	Iterations
Uniform	2.50	0.33	6.62	78.25	5000
NUMOS1	1.13	0.41	7.93	10.42	640
NUMOS24	1.02	0.39	8.20	4.85	32
fNUMOS1	1.06	0.41	8.08	1.34	84
fNUMOS24	1.02	0.42	8.64	0.63	4

for $nOS = 1$ took 640 iterations and 10.42 seconds. So for the phantom experiment, fNUMOS can be about 8x faster than the NUMOS we proposed before.²⁶

4 Discussion & Conclusion

In this paper, we proposed fNUMOS, aiming to solve FMT with L^1 regularization at a very fast speed while maintaining the reconstruction accuracy by combining Nesterov's momentum technique into a non-uniform updating scheme with acceleration from the ordered subsets method

(with a relatively small number of OS). Using both numerical simulation and phantom data, we compared fNUMOS with other algorithms. We found that fNUMOS performed best in the sense that reconstructed targets were more localized (smaller VR, higher Dice, and smaller MSE (only for simulation case) than those from other methods. fNUMOS was found to be robust for high noise levels (i.e. SNR=1) and have a significantly faster speed of convergence. In our numerical simulation study, we found a significant speed gain from the momentum technique at about 10 times in comparison with the NUMOS algorithm we studied before.²⁶ For data of the cubic phantom experiment from a smaller system matrix, the speed gain was about 8 times. This implies that fNUMOS has great potential in obtaining higher spatial resolution when a finer mesh is used.

For a thorough comparison on how the image quality metrics, VR and Dice, change according to the iteration time for uniform and non-uniform algorithm with or without the acceleration techniques of momentum or order subsets, we evaluated all eight scenarios in two groups: (1) Uniform1, f+Uniform1, NUMOS1, fNUMOS1, and (2) Uniform24, f+Uniform24, NUMOS24, fNUMOS24. Results for the simulated mouse data were shown in Figure 3 and 4 for $nOS = 1$ and $nOS = 24$, respectively. In both figures, we clearly see that non-uniform update, NUMOS1, performs much better than the uniform method, Uniform1, (blue/● vs. magenta/o curves), and it is even better than Uniform accelerated by the momentum technique, f+Uniform1, (blue/● vs. green/◇ curves). Across Figures 3 and 4, we can reaffirm that the momentum technique is faster than the ordered subsets technique in accelerating the convergence (red/+ curve in Figure 3 vs. blue/● curve in Figure 4). Lastly when both acceleration techniques are combined together with non-uniform update, the proposed fNUMOS24 can achieve a high quality reconstruction using the shortest amount of time. For the phantom experimental data, the results from comparison are plotted in Figure 5 and 6. We see similar trends as in the simulated case and fNUMOS24 again

performs the best, producing a high quality image within a second.

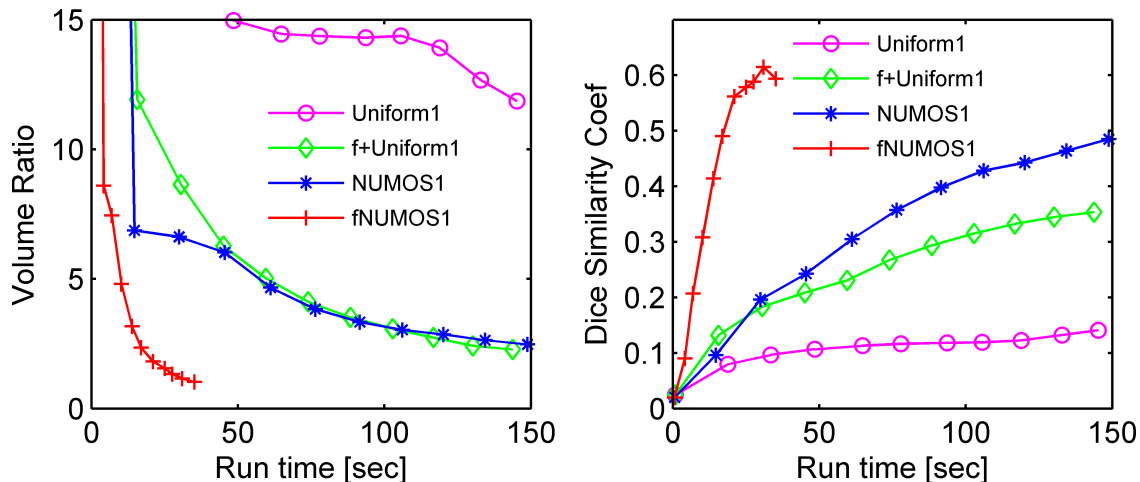


Fig 3 Comparison of Uniform and Non-uniform updates with or without the acceleration of Nesterov's momentum (abbreviated as 'f'), when the number of ordered subsets is fixed to be 1 for the simulated mouse.

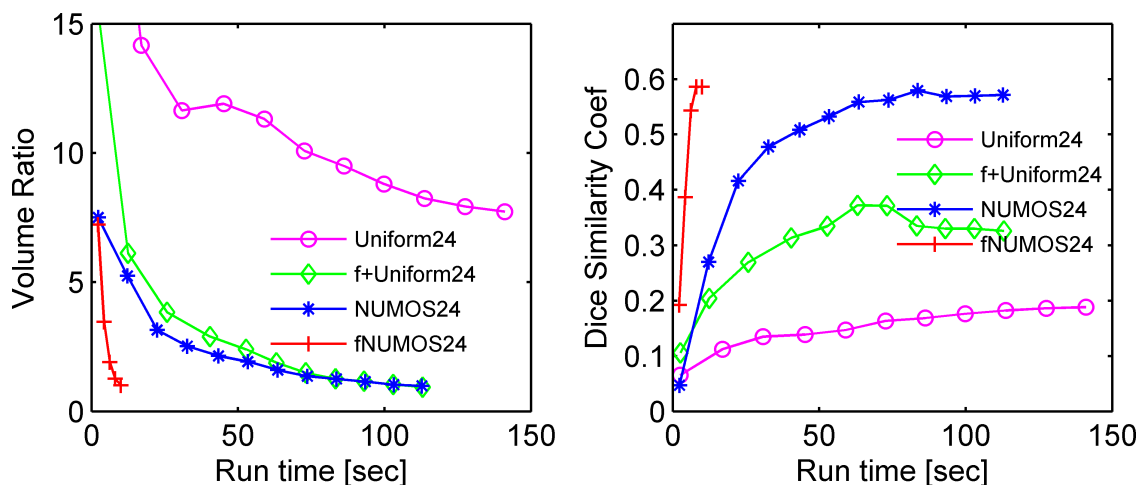


Fig 4 Comparison of Uniform and Non-uniform updates with or without the acceleration of Nesterov's momentum, when the number of ordered subsets is fixed to be 24 for the simulated mouse.

For the popular FISTA algorithm that was adopted in FMT,¹¹ we already compared it with the NUMOS1 algorithm that we proposed in our previous work using data from phantom experiments.²⁶ We found that excluding the computation time for the largest eigenvalue of the system matrix for the L needed in (16), FISTA was still slower than NUMOS and the final reconstruction results were not as good either.²⁶ For the numerically simulated mouse with a large system matrix, we were able to compute the largest eigenvalue with our dual CPU Intel Xeon E5-2680 v2 work-

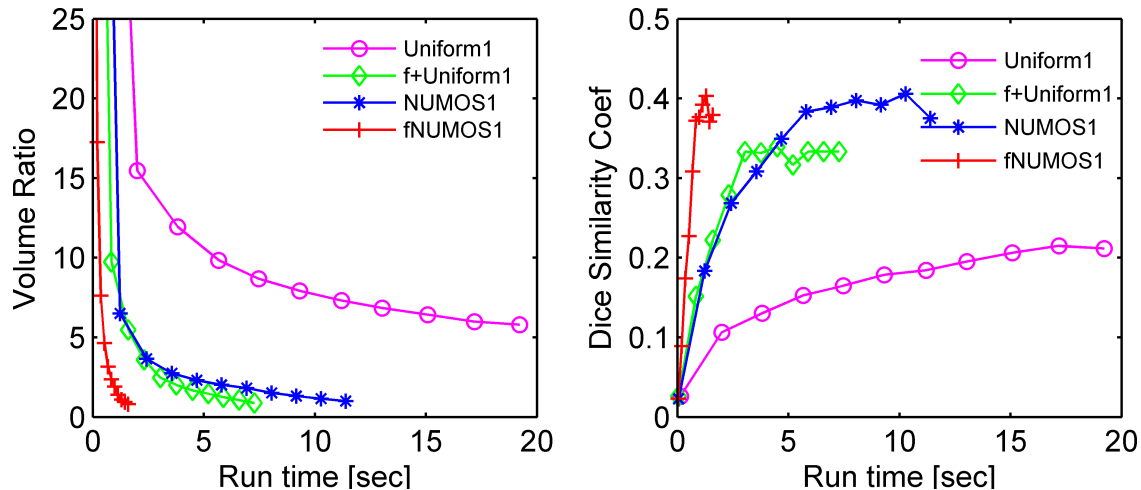


Fig 5 Comparison of Uniform and Non-uniform updates with or without the acceleration of Nesterov’s momentum when the number of ordered subsets is fixed to be 1 for the cubic phantom experiment.

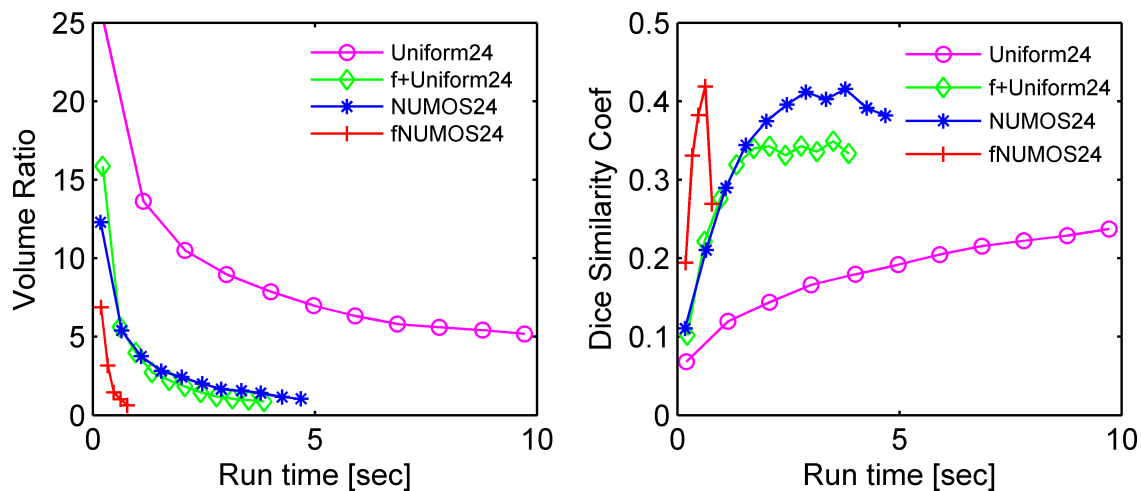


Fig 6 Comparison of Uniform and Non-uniform updates with or without the acceleration of Nesterov’s momentum, when the number of ordered subsets is fixed to be 24 for the cubic phantom experiment.

station with 20 cores and 128GB memory, but this alone cost about 30 minutes. We also explored the backtracking version FISTA²⁹ to estimate the L and we found that it tended to reach a large L value that slows down the convergence. As we pointed out in Section 2.6, the slower convergence and poorer image quality are due to the fact that choosing of L is essentially assigning a uniform updating step size for each location, thus it’s not surprising that FISTA was not as efficient as our non-uniform methods, including the earlier NUMOS and the newly proposed fNUMOS.

We noticed that the OS method (nOS=24) had limited capability increasing the speed of non-

uniform type algorithms by around 4 times for the numerically simulated mouse and around 2 times for the cubic phantom. That limit in speed increase is mainly due to the choice of OS being random at every iteration, which greatly increased the total cost. And we also see that for experimental data of smaller size, the overhead cost from randomization will take a larger portion of the total computation time so that the speed up is not as significant as for the larger simulated data. It would be of great interest to identify an optimal deterministic way of selecting the OS so that we may potentially obtain a speed increase closer to the ideal nOS times.

On the choice of the best L^1 regularization parameter, we swept through a range of values within $(0, \max(A^t b))$. We realized that the L^1 regularization parameter can actually have a range where images of similar qualities (in terms of VR, Dice and CNR) can be reconstructed, although the corresponding numbers of iterations needed will vary accordingly. Due to the significant reduction of computation time, we may further investigate the optimal way to determine an appropriate regularization parameter.

For the stopping criteria for our iterative algorithms, we found that since the convergence speed of different algorithms vary, the stopping criteria need to be selected differently. In addition, our ordered subsets are chosen in a randomized fashion, the errors in between consecutive iterations also possess randomness, so no simple stopping criteria exists. We plan to further investigate the appropriate stopping criterion for each algorithm in the future, especially when we can identify a deterministic selection of ordered subsets.

In summary, we have investigated in this paper the accelerating effects of a Nesterov type momentum technique on the non-uniform updating scheme with or without the OS method. We have obtained high quality images using around 10x less time with the proposed new method than some of the existing cutting-edge methods. Our next step will be applying the proposed new

method to *in vivo* experiments.

Acknowledgments

We would like to thank the support from University of California, Merced, Startup Fund.

References

- 1 S. R. Cherry, “In vivo molecular and genomic imaging: new challenges for imaging physics,” *Physics in Medicine and Biology* **49**(3), R13 (2004).
- 2 V. Ntziachristos, J. Ripoll, L. V. Wang, and R. Weissleder, “Looking and listening to light: the evolution of whole-body photonic imaging,” *Nature Biotechnology* **23**(3), 313–320 (2005).
- 3 S. Ahn, A. J. Chaudhari, F. Darvas, C. A. Bouman, and R. M. Leahy, “Fast iterative image reconstruction methods for fully 3d multispectral bioluminescence tomography,” *Physics in Medicine and Biology* **53**(14), 3921 (2008).
- 4 C. Li, G. S. Mitchell, J. Dutta, S. Ahn, R. M. Leahy, and S. R. Cherry, “A three-dimensional multispectral fluorescence optical tomography imaging system for small animals based on a conical mirror design,” *Optics express* **17**(9), 7571–7585 (2009).
- 5 C. Li, Y. Yang, G. S. Mitchell, and S. R. Cherry, “Simultaneous pet and multispectral 3-dimensional fluorescence optical tomography imaging system,” *Journal of Nuclear Medicine* **52**(8), 1268–1275 (2011).
- 6 V. Lukic, V. A. Markel, and J. C. Schotland, “Optical tomography with structured illumination,” *Optics Letters* **34**(7), 983–985 (2009).

- 7 J. Dutta, S. Ahn, A. A. Joshi, and R. M. Leahy, "Illumination pattern optimization for fluorescence tomography: theory and simulation studies," *Physics in Medicine and Biology* **55**(10), 2961 (2010).
- 8 N. Ducros, A. Bassi, G. Valentini, G. Canti, S. Arridge, and C. DAndrea, "Fluorescence molecular tomography of an animal model using structured light rotating view acquisition," *Journal of Biomedical Optics* **18**(2), 020503–020503 (2013).
- 9 S. R. Arridge and M. Schweiger, "Inverse methods for optical tomography," in *Information Processing in Medical Imaging*, 259–277, Springer (1993).
- 10 D. Han, J. Tian, S. Zhu, J. Feng, C. Qin, B. Zhang, and X. Yang, "A fast reconstruction algorithm for fluorescence molecular tomography with sparsity regularization," *Optics Express* **18**(8), 8630–8646 (2010).
- 11 D. Han, J. Tian, C. Qin, B. Zhang, K. Liu, and X. Ma, "A fast reconstruction method for fluorescence molecular tomography based on improved iterated shrinkage," in *SPIE Medical Imaging*, 79651C–79651C, International Society for Optics and Photonics (2011).
- 12 J. C. Baritiaux, K. Hassler, and M. Unser, "An efficient numerical method for general l_p regularization in fluorescence molecular tomography," *Medical Imaging, IEEE Transactions on* **29**(4), 1075–1087 (2010).
- 13 A. Behrooz, H.-M. Zhou, A. A. Eftekhar, and A. Adibi, "Total variation regularization for 3d reconstruction in fluorescence tomography: experimental phantom studies," *Applied Optics* **51**(34), 8216–8227 (2012).
- 14 J. Dutta, S. Ahn, C. Li, S. R. Cherry, and R. M. Leahy, "Joint l_1 and total variation regulariza-

- tion for fluorescence molecular tomography,” *Physics in Medicine and Biology* **57**(6), 1459 (2012).
- 15 J. Shi, B. Zhang, F. Liu, J. Luo, and J. Bai, “Efficient l1 regularization-based reconstruction for fluorescent molecular tomography using restarted nonlinear conjugate gradient,” *Optics Letters* **38**(18), 3696–3699 (2013).
 - 16 J. Shi, F. Liu, G. Zhang, J. Luo, and J. Bai, “Enhanced spatial resolution in fluorescence molecular tomography using restarted l1-regularized nonlinear conjugate gradient algorithm,” *Journal of Biomedical Optics* **19**(4), 046018 (2014).
 - 17 D. Zhu and C. Li, “Nonconvex regularizations in fluorescence molecular tomography for sparsity enhancement,” *Physics in Medicine and Biology* **59**(12), 2901–2912 (2014).
 - 18 J. Ripoll, “Hybrid fourier-real space method for diffuse optical tomography,” *Optics Letters* **35**(5), 688–690 (2010).
 - 19 N. Ducros, C. D’andrea, G. Valentini, T. Rudge, S. Arridge, and A. Bassi, “Full-wavelet approach for fluorescence diffuse optical tomography with structured illumination,” *Optics Letters* **35**(21), 3676–3678 (2010).
 - 20 T. Correia, T. Rudge, M. Koch, V. Ntziachristos, and S. Arridge, “Wavelet-based data and solution compression for efficient image reconstruction in fluorescence diffuse optical tomography,” *Journal of Biomedical Optics* **18**(8), 086008–086008 (2013).
 - 21 J. Shi, F. Liu, J. Zhang, J. Luo, and J. Bai, “Fluorescence molecular tomography reconstruction via discrete cosine transform-based regularization,” *Journal of Biomedical Optics* **20**(5), 055004–055004 (2015).
 - 22 X. Cao, X. Wang, B. Zhang, F. Liu, J. Luo, and J. Bai, “Accelerated image reconstruction in

- fluorescence molecular tomography using dimension reduction,” *Biomedical Optics Express* **4**(1), 1–14 (2013).
- 23 H. M. Hudson and R. S. Larkin, “Accelerated image reconstruction using ordered subsets of projection data,” *Medical Imaging, IEEE Transactions on* **13**(4), 601–609 (1994).
- 24 D. Zhu, Y. Zhao, R. Baikejiang, Z. Yuan, and C. Li, “Comparison of regularization methods in fluorescence molecular tomography,” *Photonics* **1**(2), 95–109 (2014).
- 25 K. Lange, “The MM algorithm,” in *Optimization*, 185–219, Springer (2013).
- 26 D. Zhu and C. Li, “Nonuniform update for sparse target recovery in fluorescence molecular tomography accelerated by ordered subsets,” *Biomedical Optics Express* **5**(12), 4249–4259 (2014).
- 27 Y. Nesterov, “A method of solving a convex programming problem with convergence rate $O(1/k^2)$,” in *Soviet Mathematics Doklady*, **27**(2), 372–376 (1983).
- 28 Y. Nesterov, “Smooth minimization of non-smooth functions,” *Mathematical Programming* **103**(1), 127–152 (2005).
- 29 A. Beck and M. Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM Journal on Imaging Sciences* **2**(1), 183–202 (2009).
- 30 D. Kim, S. Ramani, J. Fessler, *et al.*, “Combining ordered subsets and momentum for accelerated x-ray ct image reconstruction,” *Medical Imaging, IEEE Transactions on* **34**(1), 167–178 (2015).
- 31 D. Zhu and C. Li, “Accelerating spatially non-uniform update for sparse target recovery in fluorescence molecular tomography by ordered subsets and momentum methods,” in *SPIE BiOS*, 93190U–93190U, International Society for Optics and Photonics (2015).

- 32 F. Fedele, J. Laible, and M. Eppstein, “Coupled complex adjoint sensitivities for frequency-domain fluorescence tomography: theory and vectorized implementation,” *Journal of Computational Physics* **187**(2), 597–619 (2003).
- 33 D. R. Hunter and K. Lange, “A tutorial on MM algorithms,” *The American Statistician* **58**(1), 30–37 (2004).
- 34 I. Daubechies, M. Defrise, and C. De Mol, “An iterative thresholding algorithm for linear inverse problems with a sparsity constraint,” *Communications on Pure and Applied Mathematics* **57**(11), 1413–1457 (2004).
- 35 E. J. Candes, M. B. Wakin, and S. P. Boyd, “Enhancing sparsity by reweighted l_1 minimization,” *Journal of Fourier Analysis and Applications* **14**(5-6), 877–905 (2008).
- 36 I. Daubechies, R. DeVore, M. Fornasier, and C. S. Güntürk, “Iteratively reweighted least squares minimization for sparse recovery,” *Communications on Pure and Applied Mathematics* **63**(1), 1–38 (2010).
- 37 J. A. Fessler, E. P. Ficaro, N. H. Clinthorne, and K. Lange, “Grouped-coordinate ascent algorithms for penalized-likelihood transmission image reconstruction,” *Medical Imaging, IEEE Transactions on* **16**(2), 166–175 (1997).
- 38 H. Erdogan and J. A. Fessler, “Ordered subsets algorithms for transmission tomography,” *Physics in Medicine and Biology* **44**(11), 2835 (1999).
- 39 A. R. De Pierro, “A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography,” *Medical Imaging, IEEE Transactions on* **14**(1), 132 (1995).

- 40 D. Kim, D. Pal, J.-B. Thibault, J. Fessler, *et al.*, “Accelerating ordered subsets image reconstruction for x-ray ct using spatially nonuniform optimization transfer,” *Medical Imaging, IEEE Transactions on* **32**(11), 1965–1978 (2013).
- 41 Z. Yu, J.-B. Thibault, C. A. Bouman, K. D. Sauer, and J. Hsieh, “Fast model-based x-ray ct reconstruction using spatially nonhomogeneous icd optimization,” *Image Processing, IEEE Transactions on* **20**(1), 161–175 (2011).
- 42 A. R. De Pierro, “On the relation between the isra and the em algorithm for positron emission tomography,” *Medical Imaging, IEEE Transactions on* **12**(2), 328–333 (1993).
- 43 Z.-Q. Luo, “On the convergence of the lms algorithm with adaptive learning rate for linear feedforward networks,” *Neural Computation* **3**(2), 226–245 (1991).
- 44 J. Browne and A. R. De Pierro, “A row-action alternative to the em algorithm for maximizing likelihood in emission tomography,” *Medical Imaging, IEEE Transactions on* **15**(5), 687–699 (1996).
- 45 S. Ahn, J. Fessler, *et al.*, “Globally convergent image reconstruction for emission tomography using relaxed ordered subsets algorithms,” *Medical Imaging, IEEE Transactions on* **22**(5), 613–626 (2003).
- 46 S. Ahn, J. Fessler, D. Blatt, A. O. Hero, *et al.*, “Convergent incremental optimization transfer algorithms: Application to tomography,” *Medical Imaging, IEEE Transactions on* **25**(3), 283–296 (2006).
- 47 S. Becker, J. Bobin, and E. J. Candès, “Nesta: a fast and accurate first-order method for sparse recovery,” *SIAM Journal on Imaging Sciences* **4**(1), 1–39 (2011).

- 48 P. Tseng, “On accelerated proximal gradient methods for convex-concave optimization,” *Submitted to SIAM Journal on Optimization* (2008).
- 49 T. Correia, A. Gibson, M. Schweiger, and J. Hebden, “Selection of regularization parameter for optical topography,” *Journal of Biomedical Optics* **14**(3), 034044–034044 (2009).
- 50 F. Tian, G. Alexandrakis, and H. Liu, “Optimization of probe geometry for diffuse optical brain imaging based on measurement density and distribution,” *Applied Optics* **48**(13), 2496–2504 (2009).
- 51 L. R. Dice, “Measures of the amount of ecologic association between species,” *Ecology* **26**(3), 297–302 (1945).
- 52 X. Song, B. W. Pogue, S. Jiang, M. M. Doyley, H. Dehghani, T. D. Tosteson, and K. D. Paulsen, “Automated region detection based on the contrast-to-noise ratio in near-infrared tomography,” *Applied Optics* **43**(5), 1053–1062 (2004).
- 53 B. Dogdas, D. Stout, A. F. Chatziioannou, and R. M. Leahy, “Digimouse: a 3d whole body mouse atlas from ct and cryosection data,” *Physics in Medicine and Biology* **52**(3), 577 (2007).
- 54 C. Li, G. Wang, J. Qi, and S. R. Cherry, “Three-dimensional fluorescence optical tomography in small-animal imaging using simultaneous positron-emission-tomography priors,” *Optics Letters* **34**(19), 2933–2935 (2009).

List of Figures

- 1 (Selected) Coronal slices of the reconstructed simulated mouse from bottom to top, using (a) Uniform method, Uniform1, at 2000 iterations, (b) NUMOS, $nOS = 1$, at 1310 iterations, (c) NUMOS with $nOS = 24$, at 53 iterations, (d) fNUMOS with $nOS = 1$, at 121 iterations, and (e) fNUMOS with $nOS = 24$ at 5 iterations. The truth image is shown in (f).
- 2 Each slice corresponds to a coronary section of the reconstructed cubic phantom from bottom to top, using (a) Uniform method, Uniform1, at 5000 iterations, (b) NUMOS, $nOS = 1$, at 640 iterations, (c) NUMOS with $nOS = 24$, at 32 iterations, (d) fNUMOS with $nOS = 1$, at 84 iterations and (e) fNUMOS with $nOS = 24$ at 4 iterations. The truth image is shown in (f).
- 3 Comparison of Uniform and Non-uniform updates with or without the acceleration of Nesterov's momentum (abbreviated as 'f'), when the number of ordered subsets is fixed to be 1 for the simulated mouse.
- 4 Comparison of Uniform and Non-uniform updates with or without the acceleration of Nesterov's momentum, when the number of ordered subsets is fixed to be 24 for the simulated mouse.
- 5 Comparison of Uniform and Non-uniform updates with or without the acceleration of Nesterov's momentum when the number of ordered subsets is fixed to be 1 for the cubic phantom experiment.

- 6 Comparison of Uniform and Non-uniform updates with or without the acceleration of Nesterov's momentum, when the number of ordered subsets is fixed to be 24 for the cubic phantom experiment.

List of Tables

- 1 Comparison of VR, Dice, CNR, MSE, Time (seconds) and number of Iterations for the FMT reconstruction of the simulated mouse using different algorithms.
- 2 Comparison of VR, Dice, CNR, Time (seconds) and number of Iterations for the FMT reconstruction of cubic phantom using different algorithms.