Nonuniform update for sparse target recovery in fluorescence molecular tomography accelerated by ordered subsets

Dianwen Zhu and Changqing Li[*]

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

[*]*cli32@ucmerced.edu*

Abstract: Fluorescence molecular tomography (FMT) is a promising imaging modality and has been actively studied in the past two decades since it can locate the specific tumor position three-dimensionally in small animals. However, it remains a challenging task to obtain fast, robust and accurate reconstruction of fluorescent probe distribution in small animals due to the large computational burden, the noisy measurement and the ill-posed nature of the inverse problem. In this paper we propose a nonuniform preconditioning method in combination with L^1 regularization and ordered subsets technique (NUMOS) to take care of the different updating needs at different pixels, to enhance sparsity and suppress noise, and to further boost convergence of approximate solutions for fluorescence molecular tomography. Using both simulated data and phantom experiment, we found that the proposed nonuniform updating method outperforms its popular uniform counterpart by obtaining a more localized, less noisy, more accurate image. The computational cost was greatly reduced as well. The ordered subset (OS) technique provided additional 5 times and 3 times speed enhancements for simulation and phantom experiments, respectively, without degrading image qualities. When compared with the popular L^1 algorithms such as iterative soft-thresholding algorithm (ISTA) and Fast iterative soft-thresholding algorithm (FISTA) algorithms, NUMOS also outperforms them by obtaining a better image in much shorter period of time.

References and links

1. S. R. Cherry, "In vivo molecular and genomic imaging: new challenges for imaging physics," Phys. Med. Bio. **49**, 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," Nat. Biotechnol. **23**, 313–320 (2005).
3. S. R. Arridge and M. Schweiger, "Inverse methods for optical tomography," in *Information Processing in Medical Imaging* (Springer, 1993), pp. 259–277.
4. J. C. Baritaux, K. Hassler, and M. Unser, "An efficient numerical method for general l_p regularization in fluorescence molecular tomography," IEEE Trans. Med. Imaging **29**, 1075–1087 (2010).

5. J. Dutta, S. Ahn, C. Li, S. R. Cherry, and R. M. Leahy, "Joint L^1 and total variation regularization for fluorescence molecular tomography," Phys. Med. Biol. **57**, 1459 (2012).

6. D. Zhu and C. Li, "Nonconvex regularizations in fluorescence molecular tomography for sparsity enhancement," Phys. Med. Biol. **59**, 2901–2912 (2014).

7. D. Zhu, Y. Zhao, R. Baikejiang, Z. Yuan, and C. Li, "Comparison of regularization methods in fluorescence molecular tomography," Photonics **1**, 95–109 (2014).

8. A. Jin, B. Yazici, A. Ale, and V. Ntziachristos, "Preconditioning of the fluorescence diffuse optical tomography sensing matrix based on compressive sensing," Opt. Lett. **37**, 4326–4328 (2012).

9. K. Lange, "The MM algorithm," in *Optimization* (Springer, 2013), pp. 185–219.

10. J. A. Fessler, E. P. Ficaro, N. H. Clinthorne, and K. Lange, "Grouped-coordinate ascent algorithms for penalized-likelihood transmission image reconstruction," IEEE Trans. Med. Imaging **16**, 166–175 (1997).

11. H. Erdogan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography," Phys. Med. Bio. **44**, 2835 (1999).

12. 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," IEEE Trans. Image Process. **20**, 161–175 (2011).

13. D. Kim, D. Pal, J.-B. Thibault, and J. Fessler, "Accelerating ordered subsets image reconstruction for x-ray ct using spatially nonuniform optimization transfer," IEEE Trans. Med. Imaging **32**, 1965–1978 (2013).

14. M. E. Daube-Witherspoon and G. Muehllehner, "An iterative image space reconstruction algorthm suitable for volume ect," IEEE Trans. Med. Imaging **5**, 61–66 (1986).

15. A. R. De Pierro, "On the relation between the ISRA and the EM algorithm for positron emission tomography," IEEE Trans. Med. Imaging **12**, 328–333 (1993).

16. F. Fedele, J. Laible, and M. Eppstein, "Coupled complex adjoint sensitivities for frequency-domain fluorescence tomography: theory and vectorized implementation," J. Comput. Phys. **187**, 597–619 (2003).

17. C. Li, G. S. Mitchell, J. Dutta, S. Ahn,, R. Leahy, and S. R. Cherry,"A three-dimensional multispectral fluorescence optical tomography imaging system for small animals based on a conical mirror design," Opt. Express **17**, 7571–7585 (2009).

18. R. Weissleder, C.-H. Tung, U. Mahmood, and A. Bogdanov, "In vivo imaging of tumors with protease-activated near-infrared fluorescent probes," Nat. Biotechnol. **17**, 375–378 (1999).

19. H. M. Hudson and R. S. Larkin, "Accelerated image reconstruction using ordered subsets of projection data," IEEE Trans. Med. Imaging **13**, 601–609 (1994).

20. L. Zhao, H. Yang, W. Cong, G. Wang, and X. Intes, "Lp regularization for early gate fluorescence molecular tomography," Opt. Lett. **39**, 4156–4159 (2014).

21. F. Tian, G. Alexandrakis, and H. Liu, "Optimization of probe geometry for diffuse optical brain imaging based on measurement density and distribution," Appl. Opt. **48**, 2496–2504 (2009).

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

23. 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," Appl. Opt. **43**, 1053–1062 (2004).

24. C. Li, Y. Yang, G. S. Mitchell, and S. R. Cherry, "Simultaneous pet and multispectral 3-dimensional fluorescence optical tomography imaging system," J. Nucl. Med. **52**, 1268–1275 (2011).

25. E. van den Berg and M. P. Friedlander, "Probing the Pareto frontier for basis pursuit solutions," SIAM J. Sci. Comput. **31**(2), 890–912 (2008).

26. I. Daubechies, M. Defrise, and C. De Mol, "An iterative thresholding algorithm for linear inverse problems with a sparsity constraint," Comm. Pure Appl. Math., **57**, 1413–1457 (2004).

27. A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," SIAM J. Imaging Sci., **2**(1), 183–202 (2009).

28. 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," Opt. Express **18**, 8630–8646 (2010).

29. 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," pp. 79651C (2011).

30. Xu, Fang and Xu, Wei and Jones, Mel and Keszthelyi, Bettina and Sedat, John and Agard, David and Mueller, Klaus, "On the efficiency of iterative ordered subset reconstruction algorithms for acceleration on GPUs," Comput. Meth. Programs Biomed. **98**(3), 261–270 (2010).

31. C. Chen, F. Tian, H. Liu, and J. Huang, "Diffuse optical tomography enhanced by clustered sparsity for functional brain imaging," IEEE Trans. Med. Imaging. doi: 10.1109/TMI.2014.2338214

32. D. D. Lee and H. S. Seung, "Algorithms for non-negative matrix factorization," in *Advances in Neural Information Processing Systems* (2000), pp. 556–562.

33. F. Woolfe, M. Gerdes, M. Bello, X. Tao, and A. Can, "Autofluorescence removal by non-negative matrix factorization," IEEE Trans. Image Process. **20**, 1085–1093 (2011).

34. A. S. Montcuquet, L. Hervé, F. Navarro, J. M. Dinten, and J. I. Mars, "In vivo fluorescence spectra unmixing and autofluorescence removal by sparse nonnegative matrix factorization," IEEE Trans. Bio-Med. Eng. **58**, 2554–2565 (2011).

1. Introduction

In the past two decades, fluorescence molecular tomography (FMT) has been playing an important role in a number of preclinical research fields [1, 2]. In FMT, fluorescent agents (e.g. fluorophores) will be injected into the object such as mice and then near-infrared (NIR) laser beams will activate those fluorophores. The fluorophores will then emit fluorescence photons propagating out of the mouse surface, which can be captured by detectors such as CCD cameras. Meanwhile, the geometry of the mouse will be extracted so that along with the optical properties of the mouse, the system matrix can be set up. Then all the captured images can be combined tomographically to recover the distribution of the fluorophores inside the mouse. Due to the high scattering effects of photons in tissues, the system matrix of FMT is usually ill-conditioned and the reconstruction problem is ill-posed. Regularization methods have been employed to obtain reasonable results [3–7]. One may also construct an overall preconditioning matrix first before proceeding with any regularization models [8]. But ensuring the non-negativity constraint in FMT may be time consuming, for instance, an expensive line search is needed when the popular preconditioned conjugate gradient algorithm is used [5].

Also in the past two decades, the majorization-minimization (MM) algorithm has attracted considerable attention in medical imaging due to its advantages such as separating the high dimensional variable for an easier iterative update in a parallel way [9] and applying the non-negative constraints straightforwardly. Yet the MM algorithm has a critical issue of choosing the appropriate surrogate functions [9]. In medical imaging arguably the most well known application of MM framework was originated in Fessler et al. [10] and Erdogan & Fessler [11] for transmission tomography, where separable quadratic surrogate (SQS) functions were introduced based on a uniform additive type of weight functions. The uniform weighting has a clear advantage in the sense that it can be pre-computed and needs no iterative updates [10]. In FMT, Dutta et al. [5] followed [10, 11] and employed the uniform additive weight function to study the effects of the joint L^1 and total variation regularization method. Our previous work [6, 7] also followed this additive type of weighting when we investigated a family of nonconvex regularization methods. Nevertheless, uniform weighting tend to ignore the different updating needs at different nodes, which could hinder the convergence of iterations to the true solution.

Recently a nonuniform weighting strategy has been reported to improve image qualities in computed tomography [12, 13]. In this paper, we follow the spirit of [12, 13] and propose a *n*on-*u*niform *m*ultiplicative weighting with *o*rdered *s*ubsets (NUMOS) technique, which is a generalization of the image space reconstruction algorithm (ISRA) [14, 15], for the MM algorithm in FMT in small animal imaging, and validate its advantages over its uniform counterpart using both simulated data and phantom experiment.

2. Method

2.1. Forward modeling

In the continuous wave domain, photon transfer is modeled by the following coupled diffusion equation, along with Robin type (mixed) boundary conditions:

$$\begin{cases} -\nabla \cdot (D_{ex}(\mathbf{r})\nabla\Phi_{ex}(\mathbf{r})) + \mu_{a,ex}(\mathbf{r})\Phi_{ex}(\mathbf{r}) = S_{ex}(\mathbf{r}) \\ \mathbf{n} \cdot (D_{ex}(\mathbf{r})\nabla\Phi_{ex}(\mathbf{r})) + \alpha_{ex}\Phi_{ex}(\mathbf{r}) = 0 \\ -\nabla \cdot (D_{em}(\mathbf{r})\nabla\Phi_{em}(\mathbf{r})) + \mu_{a,em}(\mathbf{r})\Phi_{em}(\mathbf{r}) = \Phi_{ex}(\mathbf{r})S_{em}(\mathbf{r}) \\ \mathbf{n} \cdot (D_{em}(\mathbf{r})\nabla\Phi_{em}(\mathbf{r})) + \alpha_{em}\Phi_{em}(\mathbf{r}) = 0 \end{cases} \tag{1}$$

where ∇ denotes the gradient operator, $\mathbf{r}$ the location vector, $D_{ex}(\mathbf{r}) = [3(\mu'_{s,ex}(\mathbf{r}) + \mu_{a,ex}(\mathbf{r}))]^{-1}$ 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 coeffi-

cients and $\mu'_{s,ex}(\mathbf{r})$, $\mu'_{s,em}(\mathbf{r})$ being the reduced scattering coefficients at excitation and emission wavelengths, $\Phi_{ex}(\mathbf{r})$, $\Phi_{em}(\mathbf{r})$ are the photon fluxes, $S_{ex}(\mathbf{r})$ is the excitation source term, $S_{em}(\mathbf{r})$ is the fluorescence yield to be reconstructed, and $\mathbf{n}$ is the outward unit normal vector of the boundary, and α_{ex}, α_{em} are the Robin boundary coefficients.

The above equations can be solved using the finite element method (FEM), leading to a simple linear system of equations [16, 17]:

$$\mathbf{Ax} = \mathbf{b} \qquad (2)$$

where $A = (a_{ij}) \in \mathbf{R}^{N_m \times N_n}$, $a_{ij} > 0$, is the system matrix, $\mathbf{x} = (x_j) \in \mathbf{R}^{N_n \times 1}$ is a short way to write the fluorescence yield $S_{em}(\mathbf{r})$, $\mathbf{b} = (b_i) \in \mathbf{R}^{N_m \times 1}$ the measurements, and N_m, N_n the total number of measurements and FEM mesh nodes, respectively. Note that $N_m = N_d * N_s$ with N_d, N_s being the number of detectors and excitation sources.

2.2. Regularized least squares

A typical solution of (2) is obtained by minimizing the following regularized squared data-measurement misfit under the non-negativity constraint:

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}, \mathbf{x} \geq \mathbf{0}} \Phi(\mathbf{x}) := \frac{1}{2} \|A\mathbf{x} - \mathbf{b}\|_2^2 + \lambda R(\mathbf{x}), \qquad (3)$$

where λ is the regularization parameter and $R(x)$ is the regularization term, such as the L^p (semi-)norm: $R(\mathbf{x}) = \|\mathbf{x}\|_p^p$, $p \geq 0$.

Recall in the MM algorithm, the definition of a surrogate function $\Phi^{sur}(\mathbf{x})$ in the minimization problem, the following three conditions should hold:

$$\Phi^{sur}(\mathbf{x}) \geq \Phi(\mathbf{x}), \text{ for all } \mathbf{x}; \quad \Phi^{sur}(\mathbf{x}^k) = \Phi(\mathbf{x}^k), \text{ at point } \mathbf{x}^k; \quad \nabla\Phi^{sur}(\mathbf{x}^k) = \nabla\Phi(\mathbf{x}^k), \text{ at } \mathbf{x}^k. \quad (4)$$

We follow the SQS routine [11] to construct Φ^{sur} (the right-hand side of the inequality below) for the data-fitting term in (3) as follows:

$$\frac{1}{2}\|\mathbf{b} - \mathbf{Ax}\|_2^2 = \frac{1}{2} \sum_{i=1}^{N_m} (b_i - (A\mathbf{x})_i)^2 \leq \frac{1}{2} \sum_{i=1}^{N_m} \sum_{j=1}^{N_n} \beta_{ij} \{ b_i - (A\mathbf{x}^k)_i - \frac{a_{ij}}{\beta_{ij}}(x_j - x_j^k) \}^2$$

$$= \sum_{j=1}^{N_n} \{ \frac{(x_j - x_j^k)^2}{2} \sum_{i=1}^{N_m} \frac{a_{ij}^2}{\beta_{ij}} - x_j \sum_{i=1}^{N_m} a_{ij}(b_i - (A\mathbf{x}^k)_i) + constant \} \qquad (5)$$

where *constant* is independent of x_j, $\beta_{ij} > 0$ and $\sum_{j=1}^{N_n} \beta_{ij} = 1$. Note that superscript k refers to values of variables at k^{th} iteration, which are known.

The regularized non-negative minimization problem (3) may then be easily solved, especially when we employ the L^1 regularization. Denoting $\sum_{i=1}^{N_m} \frac{a_{ij}^2}{\beta_{ij}}$ by κ_j and the regularization parameter by λ_1, each x_j has the following analytical iterative updating formula:

$$x_{j,L^1}^{k+1} = \left(x_j^k + \frac{\sum_{i=1}^{N_m} a_{ij}(b_i - (A\mathbf{x}^k)_i) - \lambda_1}{\kappa_j} \right)_+ , \qquad (6)$$

where $u_+ = \max(0, u)$, representing the positive part of any function u. For the details of the above derivations, we refer readers to [6], where analytical formulas for nonconvex L^p, $0 < p < 1$, and log regularizations were also presented.

2.3. Nonuniform update

As we mentioned in the introduction section, we clearly see in (5) and (6) that choosing a suitable weight β_{ij} is critical for the success of the MM algorithm, since it determines κ_j, which dictates the convergence speed of the algorithm. Contrary to the popular uniform additive weighting [5, 6, 10, 11]: $\beta_{ij}^A = \frac{a_{ij}}{\sum_{l=1}^{N_n} a_{il}}$, we propose to adopt in this paper the following nonuniform multiplicative weighting [11, 15]:

$$\beta_{ij}^M = \frac{a_{ij}x_j^k}{\sum_{l=1}^{N_n} a_{il}x_l^k}. \tag{7}$$

We briefly analyze why (7) meets our needs better in the following. Substituting (7) into the κ_j in (6), we have $\kappa_j^M = (A^t A \mathbf{x}^k)_j / x_j^k$ and (6) has the following much simplified form, noting that the positive function $(\cdot)_+$ is not needed if $\lambda_1 = 0$:

$$x_{j,L^1}^{M,k+1} = x_j^k \cdot \frac{((A^t \mathbf{b})_j - \lambda_1)_+}{(A^t A \mathbf{x}^k)_j}. \tag{8}$$

In comparison with its counterpart obtained by using the additive form, where $\kappa_j^A = (A^t A \mathbf{1}_{N_n})_j$ with $\mathbf{1}_{N_n}$ being the N_n-dimensional vector with all entries equal 1 :

$$x_{j,L^1}^{A,k+1} = \left(x_j^k + \frac{(A^t \mathbf{b})_j - (A^t A \mathbf{x}^k)_j - \lambda_1}{(A^t A \mathbf{1}_{N_n})_j} \right)_+, \tag{9}$$

we observed two potential advantages of (8): firstly, the nonuniform update (8) clearly uses only part of the calculations in (9) so it is less expensive; and secondly, (8) implies that during the iterative process, once $x_j^k = 0$, then $x_j^{\tilde{k}}$ remains 0 for all $\tilde{k} \geq k$, as long as not all components are 0, so it naturally promotes sparsity, which is desirable in most FMT problems [18].

We then utilized a random version of the ordered subsets (OS) technique [5, 11, 19] to enhance the convergence speed of (8). In case an even division of subsets is not possible, we simply ignore those "extra" measurements. We then have the following NUMOS algorithm:

Initialization: $\mathbf{x}^0 = x_0 * \mathbf{1}_n$, $B = (A^t b - \frac{\lambda}{nOS} * \mathbf{1}_{N_n})_+$;
for $k = 1$ *to* $\frac{N_{max}}{nOS}$ **do**
  Compute A_i based on a random partition of the N_d detectors, and select the corresponding B_i ;
  for $i = 1$ *to* nOS **do**
    $\mathbf{x}^{k+1} = B_i .* \mathbf{x}^k ./ (A_i^t A_i \mathbf{x}^k)$;
  end
  If $\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2 / \|\mathbf{x}^k\|^2 < \delta_{stop} * nOS$, break;
end
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, and δ_{stop} the empirically chosen stopping criterion.

Algorithm 1: NUMOS

2.4. Remarks

We remark here that the better performing nonconvex ($L^p, 0 < p < 1$, and log) regularizations as identified by our previous work [6, 7] and a recent work from another group [20] are essentially doing nonuniform updates, the effects of which will be undesirably canceled in part, if

not all, by the effects of the nonuniform updating of κ_j. So in this paper we only include L^1 regularization to retain the nonuniform updating of κ_j in hope of further gain in image qualities.

We also remark that the recent NU-SQS multiplicative variant [13]: $\beta_{ij}^{nM} = \frac{a_{ij}\Delta x_j^k}{\sum_{l=1}^{N_n} a_{il}\Delta x_l^k}$, where $\Delta x_j^k = |x_j^k - x_j^{k-1}|$, does not lead to a simple update as (8), hence it does not carry any of the aforementioned benefits when compared with the uniform additive weighting. It may however significantly enhance the convergence if the initialization is close to the truth as shown in [13]. But since we don't have a good initialization as the filtered-back-projection result used in [13] for computed tomography, we will investigate this variant in the future.

2.5. Image quality metrics

In the analysis of the qualities of the reconstructed images and the selection of the best regularization parameter, we used on the same four metrics as proposed in our previous work [6]: the volume ratio (VR) [21] that is the ratio of the reconstructed volume (typically defined as the voxels with intensities greater than half of the maximum) to the true volume, the dice similarity coefficient (Dice) [22] that is the ratio of the intersection to the average of two volumes, and the well known mean squared error (MSE) and contrast-to-noise ratio (CNR) [23]. VR is a measure of sparsity of the reconstructed target and Dice quantifies the shape and location accuracy. The closer VR and Dice are to 1, the better. MSE is only calculated if the ground-truth is known and usually the smaller it is, the better. CNR is about the reconstructed image only and higher values are preferred.

3. Numerical simulations and results

To validate the advantages of NUMOS, we first tested it on simulated data. For a fair comparison with the uniform additive weighting without OS we adopted in [6], here we used the same simulated data set and first set $nOS = 1$, i.e. no OS, then increased nOS by orders of 2 [11]: $2, 2^2, ..., 2^7$. We briefly recapitulate the simulation setup as follows. Two fluorescent capillary tube sources were inserted inside a mouse mesh, which consisted of 32,332 nodes and 161,439 tetrahedral elements. The tubes diameters were 2 mm and lengths 20 mm. The fluorophore concentration inside the tubes were 1 and outside 0. A total of 60 laser excitations source nodes were uniformly selected along the trunk of the mouse and all the 4020 surface nodes covering the trunk were set as detectors. The tissue optical properties were $\mu_a = 0.007\,\text{mm}^{-1}$, $\mu_s' = 0.72\,\text{mm}^{-1}$ at 650 nm, the excitation wavelength, and $\mu_a = 0.014\,\text{mm}^{-1}$, $\mu_s' = 0.78\,\text{mm}^{-1}$ at 700 nm, the emission wavelength. White Gaussian noise with a signal-to-noise ratio of 1 was added to the simulated measurements. All our calculations were done on an Intel i5 2400 3.1GHz PC with 16GB memory.

The simulation result is shown in Fig. 1, with detailed image quality metrics in Table 1. Note that in this and the following table, we also included the metrics for the best images reconstructed using the uniform additive weighting (Uniform) along with nonconvex regularizations, which were directly copied from our previous work [6]. We clearly see that the proposed NUMOS with $nOS = 1$ significantly outperforms "Uniform" in obtaining a more localized and less noisy image. The sparsity (VR) and location accuracy (Dice) have been significantly enhanced, CNR is more than two times higher, and MSE is about 40% lower, as indicated by Table 1. The computational cost of NUMOS, $nOS = 1$, is less than 1/4 of Uniform. This is partly because that NUMOS needed only about 1400 iterations to stop, while Uniform took 2000 iterations to reach that result. As for results from different nOS, we presented the cases when $nOS = 16, 32, 64$ and 128 in Fig. 1(c)-(f). The results shown are averaged over 6 runs. For $nOS = 16$ and 32, we obtained further speed gain, without sacrificing much of the image qualities. From Fig. 1(d) and Table 1, we can see that $nOS = 32$ is a reasonably good choice
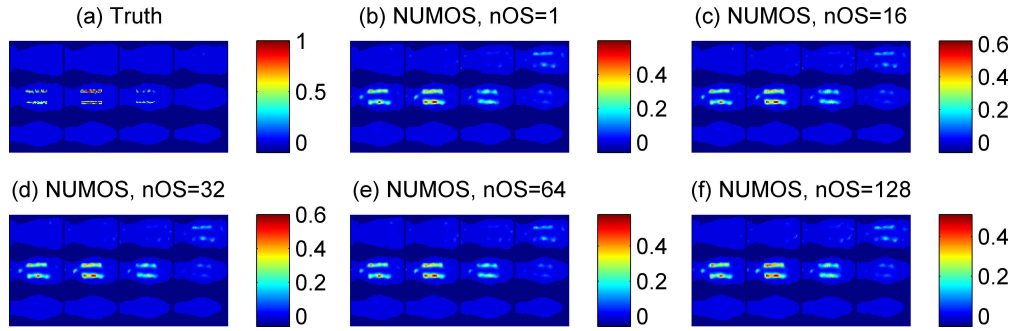
Fig. 1. The coronary sections from bottom to top of the simulated mouse: (a) Truth and reconstructions with our proposed NUMOS algorithm using $\lambda_1 = 1.0E\text{-}03$ and (b) $nOS = 1$, (c) $nOS = 16$, (d) $nOS = 32$, (e) $nOS = 64$, and (f) $nOS = 128$, respectively.

since we have about 5.4X speed gain against the $nOS = 1$ case. For $nOS = 64$ and 128, image qualities are still acceptable although they start to deteriorate. We may still opt for these cases if we want further reduction of computation time by another 24% and 40%, respectively.

Table 1. Metrics for simulated mouse: reconstructed using NUMOS with $nOS = 1, 16, 32, 64$, and 128, respectively, and $\lambda_1 = 1.0E\text{-}03$ v.s. uniform weighting with non-convex $L^{1/2}$ regularization, $\lambda_{1/2} = 4.0E\text{-}05$, the best result among different regularization methods [6]. The overall best performance is highlighted in bold.

Algorithm	VR	Dice	CNR	MSE	Time (seconds)
NUMOS, nOS=1	1.01	0.61	9.10	2.18E-03	740.4
NUMOS, nOS=16	1.04	0.61	9.01	2.21E-03	221.8
NUMOS, nOS=32	**1.08**	**0.59**	**8.94**	**2.23E-03**	**139.2**
NUMOS, nOS=64	1.18	0.57	8.81	2.27E-03	104.8
NUMOS, nOS=128	1.39	0.55	8.40	2.38E-03	81.9
Uniform	3.74	0.26	4.31	3.64E-03	3201.3

4. Phantom experiments and results

We further validated NUMOS with phantom experiment and also tested it with $nOS = 1$ and with nOS ranging from 2^1 to 2^7. The cubic phantom was again the same as we used in [6] with dimension 32 mm by 32 mm by 29 mm and was composed of 1% intralipid, 2% agar, and water in the background. In the two capillary target tubes, both 6.5 μm DiD fluorescence dye solution and uniformly distributed ^{18}[F]-fluoro-2-deoxy-D-glucose (FDG) at activity level of 100 μCi were injected for simultaneous FMT and positron emission tomography (PET) scans. The PET result will be used to validate our FMT. The excitation laser at wavelength of 650 nm scanned the front surface of the phantom at 20 illumination nodes. The emission wave length was 700 nm and there were 1057 detectors for measurement collection. 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. The FEM mesh has 8690 nodes and 47,581 tetrahedral elements. The PET images were thresholded 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 [24].

The phantom experiment result is shown in Fig. 2, with detailed image quality metrics in Table 2. The reconstructed images are normalized so that they can be compared with the truth
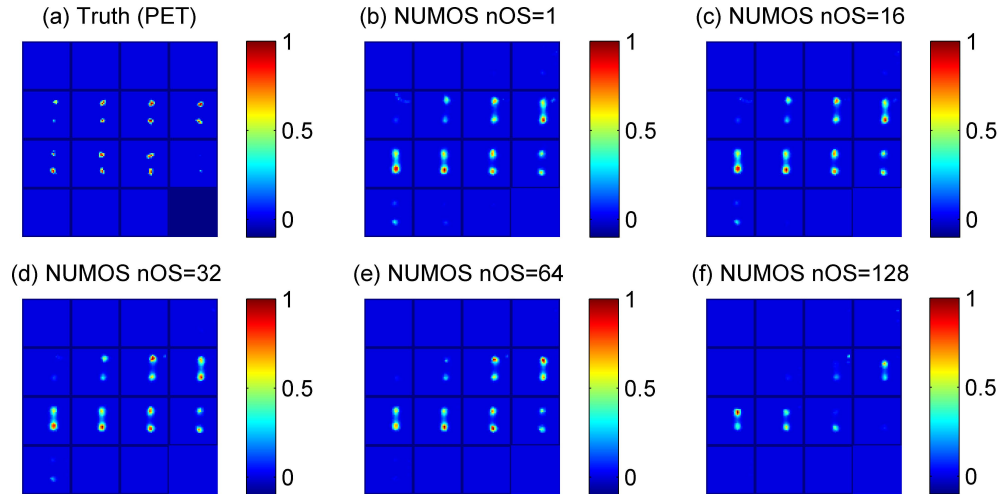
Fig. 2. The coronary sections from bottom to top of the cubic phantom: (a) the truth (PET) and reconstructions with our proposed NUMOS algorithm using $\lambda_1 = 5.0\text{E}+03$ and (b) $nOS = 1$, (c) $nOS = 16$, (d) $nOS = 32$, (e) $nOS = 64$, and (f) $nOS = 128$, respectively.

we obtained from PET, where intensity information is not comparable. The findings here are similar to the simulation case. We see that in comparison with "Uniform", NUMOS, nOS=1 takes 12x less of time to obtain a results with improved metrics: VR much closer to 1, CNR much higher, and Dice coefficient also slightly higher. The results for different nOS's are averaged from 6 runs. The results from cases of $nOS = 16$ and 32 are again very comparable to those of the case of $nOS = 1$. And $nOS = 32$ provides an additional about 3x speed gain. When nOS is further increased, the image qualities start to drop although we again see some speed gain.

Table 2. Metrics for cubic phantom, reconstructed using NUMOS with $\lambda_1 = 5.0\text{E}+03$, and $nOS = 1, 16, 32, 64,$ and 128, respectively, v.s. uniform weighting with $L^{1/2}, \lambda_{1/2} = 5.0\text{E}+07$, the best result among different regularization methods [6]. The overall best performance is highlighted in bold.

Algorithm	VR	Dice	CNR	Time (seconds)
NUMOS, nOS=1	1.13	0.41	7.90	14.7
NUMOS, nOS=16	1.23	0.39	8.00	8.3
NUMOS, nOS=32	**1.22**	**0.41**	**8.20**	**5.3**
NUMOS, nOS=64	0.77	0.40	9.11	3.5
NUMOS, nOS=128	0.26	0.14	6.38	3.1
Uniform	2.56	0.39	5.11	180.5

5. Discussion and conclusion

In this paper, we proposed NUMOS, a robust nonuniform multiplicative weighting strategy for the MM framework in FMT with L^1 regularization and ordered subsets acceleration. Using both simulation and phantom data, we compared NUMOS with the other popular uniform additive weighting strategy in combination with best possible regularizations. We started the iteration from an arbitrary uniform initialization, ran a maximum of 5000 steps and stopped

when $\|\mathbf{x}^{k+1} - \mathbf{x}^k\|^2/\|\mathbf{x}^k\|^2 < \delta_{stop} * nOS$, with δ_{stop} empirically set as 4E-04 for simulations and 9E-04 for phantom. NUMOS was robust for noise levels as high as having SNR of 1. NU-MOS performed better in the sense that the reconstructed targets were more localized (smaller VR, higher Dice, higher CNR, and smaller MSE (only for simulation case) than those from its uniform counterpart, and also in terms of its significantly faster speed of convergence. On an Intel i5 2400 3.1GHz PC with 16GB memory, each iteration took about 0.6 seconds (v.s 1.6 seconds for the uniform weighting in [6]) for the simulated data with SNR of 1, and about 0.025 (vs. 0.09) seconds for the cubic phantom. NUMOS ($nOS = 1$) also needed much less iterations to converge, bringing the total speed to 4x faster for the simulation and 12x faster for the phantom than its uniform counterparts. By setting the ordered subset number to be 32, we were able to get an additional 5.5x and 3x speed enhancements for the numerical simulation and cubic phantom respectively.

For a better understanding of the uniform additive and nonuniform multiplicative types of β_{ij} in section 2.3, we examined a small variation of the multiplicative form (7), $\beta_{ij}^{vM} = \frac{a_{ij}(x_j^k + z_j)}{\sum_{l=1}^{N_n} a_{il}(x_l^k + z_l)}$, where z_j is a small positive number [10]. It is essentially using the maximum of β_{ij}^M and β_{ij}^A at each step, which performed almost the same as β_{ij}^M, but with a slightly slower convergence speed as expected. We also examined a more general additive form $\beta_{ij}^{gA} = \frac{a_{ij}^q}{\sum_{l=1}^{N_n} a_{il}^q}, q > 0$ [9]. But not much improvement in the reconstructed images was observed when we varied the p to different values other than 1, with the exception of $0.9 \leq p < 1$, where we were able to see some slight improvement in the convergence rate.

Although the L^1 regularization is not our emphasis as we already remarked in the method section, we understand that there are a lot of efficient L^1 algorithms available. A good summary of five types of closely related L^1 optimization models can be found in [25], where more than ten specific algorithms are also discussed. However, as the authors there stated, all those algorithms are not equivalent to the regularized model we adopted (QR_λ in [25]). There are a lot of subtleties to be taken care of before they can be adopted into FMT. Here we briefly compared NUMOS with the popular iterative soft-thresholding algorithm (ISTA) [26] and fast iterative soft-thresholding algorithm (FISTA) [27] algorithms, which have been successfully implemented in FMT [28, 29] recently. We first tested the algorithms on the phantom experiment data since its size is much smaller and the calculation of its largest eigenvalue is easy. The results are shown in Fig. 3 (a) and (b) and Table 3. We have not been able to directly carry out the calculation on the digimouse due to the fact that ISTA and FISTA need the estimation of the eigenvalues of the system matrix, which is too large (requires $\sim$ 62GB memory for storage alone). We noticed there was a lot of background noise in the reconstruction result of FISTA as shown in Fig. 3(b). Further investigation revealed that FISTA did not monotonically decrease the objective function when nonnegative solutions are sought. We applied the back-tracking technique to guarantee the strict decrease of the objective function hoping to get rid of the noise. The result is shown in Fig. 3(c). To the best of our knowledge, FISTA with back-tracking line search has not been applied to FMT before. The image quality has been greatly improved and is comparable to the result from our proposed NUMOS with $nOS = 1$ as can be seen from Fig. 3(d). Yet NUMOS is already about 10 time faster. It will be a few extra times faster if OS feature is turned on. In short, our proposed NUMOS outperforms both ISTA and FISTA in yielding an accurate image in much short period of time.

For regularization problems, selection of the optimal regularization parameters has always been a issue. Although we mentioned in section 2.5 that we followed our previous work [6] where a detailed comparison of metrics VR, Dice, CNR and MSE were used, it would still be interesting if we can find an automatic way for quick selection. This is also related to how the
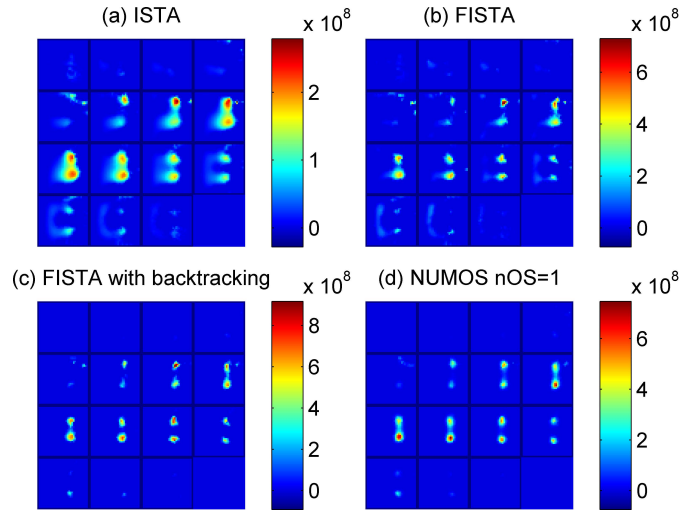
Fig. 3. The coronary sections from bottom to top of the cubic phantom: L^1 regularized reconstructions with parameter $\lambda_1 = 5.0E+03$ using (a) ISTA [26, 28], (b) FISTA [27, 29], (c) FISTA with backtracking line search [27] and (d) NUMOS algorithm with $nOS = 1$, respectively.

Table 3. Metrics for cubic phantom, reconstructed by L^1 regularization with parameter $\lambda_1 = 5.0E+03$, using (a) ISTA, (b) FISTA, (c) FISTA with backtracking line search, and (d) NUMOS with $nOS = 1$, respectively.

Algorithms	VR	Dice	CNR	Time (seconds)
ISTA	4.38	0.25	4.66	189.0 s
FISTA	1.34	0.41	7.12	24.7 s
FISTA with backtracking	1.28	0.41	8.46	138.0 s
NUMOS, nOS=1	1.13	0.41	7.90	14.7 s

reconstructed targets are defined and how the stopping criteria are chosen. We plan to do more involved comparisons and explorations in this direction in the future.

It was brought to our attention that a clustered/group sparsity promoting algorithm has been introduced into diffuse optical imaging recently [31]. We will consider similar techniques in FMT as well in the future.

For the OS technique, the choice of nOS depends on the size of our measurements, the stopping criterion, and how much we want to sacrifice the image qualities for a faster convergence speed [11, 30]. For our simulation data, we found that for $nOS = 16, 32$, the result is not too different from the case for $nOS = 1$ and for $nOS = 64, 128$, the results are still quite good. For phantom experiments, $nOS = 16, 32, 64$ are all good but $nOS = 128$ is not working well. A possible explanation is that much less measurements in each subset may cause the algorithm unstable. One important requirement for the OS technique is to choose the subsets in a balanced way [19]. That's why we did the randomization trick, which has unfortunately increased quite some computational cost and prevented us from obtaining a theoretical nOS times of speed increase [11]. We plan to further study the distribution of our detector positions so that a deterministic way of choosing ordered subsets for the FMT system may become possible, which may make it possible for NUMOS to take even less time.

Lastly, as we mentioned in the introduction, when $\lambda_1 = 0$ and $nOS = 1$, the NUMOS updating

formula (8) was also known as the ISRA algorithm [14, 15], and the "non-negative matrix factorization" (NNMF) algorithm [32]. It is interesting to note that this special case of NUMOS, under the NNMF framework, was used to remove background autofluorescence effectively [33, 34]. In our future in vivo study, we may follow the NNMF framework and preprocess the data straightforwardly to guarantee sparsity.

In conclusion, our proposed NUMOS algorithm is superior to its uniform counterpart that was popular for the MM framework in FMT and it is also significantly faster than some state-of-the-art algorithms such as FISTA, even when we disable the OS feature for speed enhancement. When a reasonably large number of ordered subsets are used, our NUMOS can be $3 \sim 5$ extra times faster.

Acknowledgments