Image reconstruction for diffuse optical tomography using sparsity regularization and expectation-maximization algorithm

Nannan Cao[1], Arye Nehorai[1]*, and Mathews Jacob[2]

[1] *Department of Electrical and Systems Engineering, Washington University in St. Louis, Bryan Hall, Room 201, Campus Box 1127, One Brookings Drive, St. Louis, MO 63130, USA.*
[2] *Department of Biomedical Engineering, University of Rochester, Rochester, NY 14642, USA.*
* Corresponding author: nehorai@ese.wustl.edu

Abstract: We present an image reconstruction method for diffuse optical tomography (DOT) by using the sparsity regularization and expectation-maximization (EM) algorithm. Typical image reconstruction approaches in DOT employ Tikhonov-type regularization, which imposes restrictions on the L_2 norm of the optical properties (absorption/scattering coefficients). It tends to cause a blurring effect in the reconstructed image and works best when the unknown parameters follow a Gaussian distribution. In reality, the abnormality is often localized in space. Therefore, the vector corresponding to the change of the optical properties compared with the background would be sparse with only a few elements being nonzero. To incorporate this information and improve the performance, we propose an image reconstruction method by regularizing the L_1 norm of the unknown parameters and solve it iteratively using the expectation-maximization algorithm. We verify our method using simulated 3D examples and compare the reconstruction performance of our approach with the level-set algorithm, Tikhonov regularization, and simultaneous iterative reconstruction technique (SIRT). Numerical results show that our method provides better resolution than the Tikhonov-type regularization and is also efficient in estimating two closely spaced abnormalities.

OCIS codes: (100.3190) Image processing: Inverse problem; (170.3010) Medical optics and biotechnology: Image reconstruction technique

---

References and links

1. D. Grosenick, T. Moesta, H. Wabnitz, J. Mucke, C. Stroszcynski, R. Macdonald, P. Schlag, and H. Rinnerberg, "Time-domain optical mammography: Initial clinial results on detection and characterization of breast tumors," Appl. Opt. **42**, 3170–3186 (2003).
2. X. Intes, J. Ripoll, Y. Chen, S. Nioka, A. Yodh, and B. Chance, "In vivo continuous-wave optical breast imaging enhanced with Indocyanine Green," Med. Phys. **30**, 1039–1047 (2003).
3. G. Strangman, D. Boas, and J. Sutton, "Non-invasive neuroimaging using near-infrared light," Biol. Psychiatry **52**, 679–693 (2002).
4. A. Villringer and B. Chance, "Non-invasive optical spectroscopy and imaging of human brain function," Trends Neurosci. **20**, 435–442 (1997).
5. A. N. Tikhonov and V. Y. Arsenin, *Solutions of ill-posed problems* (V. H. Winston Sons, Washington D. C.).
6. B. Pogue, T. McBride, J. Prewitt, U. Osterberg, and K. Paulsen, "Spatially variant regularization improves diffuse optical tomography," Appl. Opt. **38**, 2950–2961 (1999).

7. A. Li, G. Boverman, Y. Zhang, D. Brooks, E. L. Miller, M. E. Kilmer, Q. Zhang, E. M. C. Hillman, and D. Boas, "Optimal linear inverse solution with multiple priors in diffuse optical tomography," Appl. Opt. **44**, 1948–1956 (2005).

8. K. D. Paulsen and H. Jiang, "Enhanced frequency-domain optical image reconstruction in tissues through total-variation minimization," Appl. Opt. **35**, 3447–3458 (1996).

9. H. Dehghani, B. W. Pogue, S. Jiang, B. A. Brooksby, and K. D. Paulsen, "Three-dimensional optical tomography: Resolution in small-object imaging," Appl. Opt. **42**, 3117–3128 (2003).

10. A. Douiri, M. Schweiger, J. Riley, and S. R. Arridge, "Anisotropic diffusion regularization methods for diffuse optical tomography using edge prior information," Meas. Sci. Technol. **18**, 87–95 (2007).

11. M. E. Kilmer, E. L. Miller, D. Boas, and D. Brook, "A shape-based reconstruction technique for DPDW data," Opt. Express **72**, 481–491 (2000).

12. M. E. Kilmer, E. L. Miller, A. Barbaro, and D. Boas, "Three-dimensional shaped-based imaging of absorption perturbation for diffuse optical tomography," Appl. Opt. **42**, 3129–3144 (2003).

13. G. Boverman and E. Miller, "Estimation-theoretic algorithms and bounds for three-dimensional polar shape-based imaging in diffuse optical tomography," in *Proceedings of IEEE International Symposium on Biomedical Imaging*, pp. 1132–1135 (2006).

14. P. Charbonnier, L. Blanc-Feraud, G. Aubert, and M. Barlaud, "Deterministic edge-perserving regularization in computed imaging," IEEE Trans. Image Process. **6**, 298–310 (1997).

15. M. A. T. Figueiredo, "Adaptive sparseness for supervised learning," IEEE Trans. Pattern Anal. Mach. Intell. **25**, 1150–1159 (2003).

16. P. S. Bradley, O. L. Mangasarian, and W. N. Street, "Feature selection via mathematical programming," IN-FORMS J. Comput. **10**, 209–217 (1998).

17. D. Malioutov, M. Cetin, and A. S. Willsky, "A sparse signal reconstruction perspective for source localization with sensor arrays," IEEE Trans. Signal Process. **53**, 3010–3022 (2005).

18. M. Cetin and W. C. Karl, "Feature-enhanced synthetic aperture radar image formation bsed on nonquadratic regularization," IEEE Trans. Image Process. **10**, 623–631 (2001).

19. I. Gorodnitsky and B. D. Rao, "Sparse signal reconstruction from limited data using FOCUSS: A re-weighted minimum norm algorithm," IEEE Trans. Signal Process. **45**, 600–616 (1997).

20. K. Matsuura and Y. Okabe, "Selective minimum-norm solution of the biomagnetic inverse problem," IEEE Trans. Biomed. Eng. **42**, 608–615 (1995).

21. K. Matsuura and Y. Okabe, "A robust reconstruction of sparse biomagnetic sources," IEEE Trans. Biomed. Eng. **44**, 720–726 (1997).

22. M. Huang, A. Dale, T. Song, E. Halgren, D. Harrington, I. Podgorny, J. Ganive, S. Lewis, and R. Lee, "Vector-based spatial-temporal minimum L1-norm solution for MEG," NeuroImage **31**, 1025–1037 (2006).

23. M. Jacob, Y. Bresler, V. Toronov, X. Zhang, and A. Webb, "Level-set algorithm for the reconstruction of functional activation in near-infrared spectroscopic imaging," J. Biomed. Opt. **11**, 064,029–1–12 (2006).

24. O. Dorn, "A shape reconstruction method for diffuse optical tomography using a transport model and level sets," in *Proceedings of IEEE International Symposium on Bioomedical Imaging*, pp. 1015–1018 (2006).

25. R. Tibshirani, "Regression shrinkage and selection via the Lasso," J. Royal Statistical Soc. (B) **58**, 267–288 (1996).

26. M. A. O'Leary, "Imaging with diffuse photon density waves," Ph.D. thesis, University of Pennsylvania (1996).

27. R. J. Gaudette, D. H. Brooks, C. A. DiMarzio, M. E. Kilmer, E. L. Miller, T. Gaudette, and D. A. Boas, "A comparison study of linear reconstruction techniques for diffuse optical tomographic imaging of absorption co-efficient," Phys. Med. Biol. **45**, 1051–1070 (2000).

28. S. R. Arridge, "Optical tomography in medical imaging," Inverse Problems **15**, R41–R93 (1999).

29. K. D. Paulsen and H. Jiang, "Spatially-varying optical property reconstruction using a finite element diffusion equation approximation," Med. Phys. **22**, 619–701 (1995).

30. B. W. Pogue, M. S. Patterson, H. Jiang, and K. D. Paulsen, "Initial assessment of a simple system for frequency domain diffuse optical tomography," Phys. Med. Biol. **40**, 1709–1729 (1995).

31. M. S. Patterson, B. Chance, and B. C. Wilson, "Time resolved reflectance and transmittance for the noninvasive measurement of tissue optical properties," J. Appl. Opt. **28**, 2331–2336 (1989).

32. R. Aronson, "Boundary conditions for diffusion of light," J. Opt. Soc. Am. A **12**, 2532–2539 (1995).

33. R. C. Haskell, L. O. Svaasand, T.-T. Tsay, T.-C. Feng, and M. S. McAdams, "Boundary conditions for the diffusion equation in radiative transfer," J. Opt. Soc. Am. **10**, 2727–2741 (1994).

34. J.-J. Fuchs, "On sparse representations in arbitrary redundant bases," IEEE Trans. Inf. Theory **50**, 1341–1344 (2004).

35. M. A. T. Figueiredo and R. D. Nowak, "An EM algorithm for wavelet-based image restoration," IEEE Transactions on Image Processing **12**, 906–916 (2003).

36. G. McLachlan and T. Krishnan, *The EM algorithm and extensions* (Wiley, New York).

37. C. Wu, "One the convergence properties of the EM algorithm," Ann. Stst. **11**, 95–103 (1983).

38. D. Donoho and I. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," Biometrika **81**, 425–455 (1994).

39. D. L. Donoho, "De-noising by soft-threshold," IEEE Trans. Inf. Theory **41**, 613–627 (1995).

40. P. Moulin and J. Liu, "Analysis of multiresolution image denoising schemes uisng generalized-Gaussion and complexity priors," IEEE Trans. Inf. Theory **45**, 909–919 (1999).
41. S.-J. Kim, K. Koh, M. Lustig, S. Boyd, and D. Gorinevsky, "An efficient method for l1-regularized least squares," IEEE Trans. Selected Topics in Signal Process. (2007).
42. M. J. Holboke, B. J. Tromberg, X. Li, N. Shah, J. Fishkin, D. Kidney, J. Butler, B. Chance, and A. Yodh, "Three-dimensional diffuse optical mammography with ultrasound localization in a human subject," J. Biomed. Opt. **5** (2000).

1. Introduction

Diffuse optical tomography (DOT) is a non-invasive functional imaging modality whose goal is to estimate the optical properties of human tissue. It provides useful physiological information about blood volume and oxygenation and has applications in optical mammography [1, 2] and functional brain imaging [3, 4].

The forward problem in DOT describes the photon propagation in tissue and the inverse problem involves estimating the absorption and scattering coefficients of tissue from light measurements on the surface. The inverse problem is ill posed and highly underdetermined, and thus regularization is typically required through which certain prior assumptions on the solution (e.g., an assumption on its smoothness or a bound on its norm) can be imposed. The most common choice is the Tikhonov-type regularization [5, 6, 7], where the least-square residual is regularized using the L_2 norm of the unknown parameters. This adjustment aims to reduce high-frequency noise in the reconstructed images; however, it tends to produce an over-smooth solution and performs best when the real solution assumes a Gaussian distribution. Different methods have been developed to overcome these drawbacks. Paulsen and Jiang proposed a non-linear regularization method that minimizes the total-variation norm [8]. The anisotropic diffusion regularization was applied in [9, 10], which provides better performance in preserving geometrical structures such as edges or corners and can readily incorporate prior structural information from other imaging modalities. Pogue *et al.* employed spatially variant regularization parameters to compensate for the spatial dependence of the contrast and resolution in the reconstruction [6], and several shape-based approaches were proposed in [11, 12] with their performances analyzed using the Cramér-Rao bound [13].

In this paper, we propose an alternative method to improve the spatial resolution of the reconstructed images in DOT. Our method is based on the observation that abnormalities in the domain (e.g., tumors in the breast or activations in the brain) are typically spatially concentrated and thus sparse. Therefore, it would be helpful to incorporate such information in solving the inverse problem. In recent years, the topic of sparse signal representation and estimation has developed in a variety of applications, including image reconstruction and restoration [14], feature selection and classification in machine learning [15, 16], radar signal processing and imaging [17, 18], and biomedical imaging [19, 20, 21, 22, 23, 24]. It was also commonly known as LASSO (Least Absolute Shrinkage and Selection Operator) in the field of regression [25]. In [15], the authors used a Laplacian prior to enforce sparsity and interpreted it using a hierarchical Bayesian method. In [19], a re-weighted minimum norm algorithm called FOCal Underdetermined System Solver (FOCUSS) was proposed, whereby a low resolution estimate of the sparse signal was first obtained and then pruned to a sparse signal representation. Different level-set algorithms were developed to incorporate the sparse nature of the activations in DOT and consider both the support and values of the activated regions [23, 24].

We propose imposing the sparseness of the optical abnormality using the L_1 norm regularization and solving it using the expectation-maximization (EM) algorithm. The EM algorithm was originally formulated to obtain the maximum likelihood estimation with missing data. In our case, we reformulate the measurement model in such a way that we first consider the real absorption perturbation as a "clean" perturbation corrupted by white Gaussian noise, regard it

as a type of missing data, and then estimate the pure perturbation iteratively using the EM algorithm. In particular, we apply a soft-thresholding approach in the maximization step to simplify the computation. We validate our method using 3D simulated examples where the scattering coefficients are assumed to be spatially constant and known and we recover only the absorption coefficient. We compare our approach with standard methods, including the Tikhonov regularization and simultaneous iterative reconstruction technique (SIRT) [26, 27], as well as a level-set algorithm [23]. The level-set method represents the support of the abnormality using a level-set scheme and makes the inverse problem well-posed by exploiting the spatially concentrated nature of the abnormality. We show that our method provides better resolution than the standard Tikhonov-type regularization and is efficient in estimating two closely-spaced abnormalities.

This paper is organized as follows: In Section 2 we briefly review the forward and measurement models in DOT. In Section 3, we propose the image reconstruction method using sparse regularization and provide the solution using the EM algorithm. Numerical examples are given in Section 4, where we present our reconstruction results and compare the performance with other methods. Finally, we offer conclusions in Section 5.

2. Forward and measurement models

Photon propagation in human tissue is mathematically described by the Boltzmann transport equation. It can be simplified under certain assumptions using the diffusion approximation (DA) equation, which can be solved using the finite element method (FEM) [28, 29], finite difference method (FDM) [30], and Born or Rytov approximations [26]. In this paper, we employ the Rytov solution to DA, where the scattered field is assumed to be slowly varying in space. Furthermore, we assume that the scattering coefficient μ_s is spatially constant and known, and we focus solely on the reconstruction of the absorption coefficient μ_a.

Denote N_s the number of light sources, N_d the number of detectors, and $N_m = N_s \times N_d$ the number of measurements. Assuming that the domain is divided into N voxels with a constant absorption coefficient μ_{ai} in the ith voxel, the Rytov solution to DA is expressed as [26]

$$\phi^c = A^c \mu, \tag{1}$$

where the superscript "c" denotes complex values, $\phi^c \in C^{N_m}$ the Rytov phase, $A^c \in C^{N_m \times N}$ the weighting matrix, and $\mu \in R^N$ the change of μ_a in each voxel compared with the background. More specifically, we can write $\mu = [\delta\mu_{a1}, \delta\mu_{a2}, \ldots, \delta\mu_{aN}]^T$, where $\delta\mu_{ai} = \mu_{ai} - \mu_{a0}$ with μ_{a0} denoting the homogeneous background absorption coefficient. Note that Eq. (1) can be used for different simple geometries such as infinite, semi-infinite, or slab, by modifying the elements in A^c using the method of image sources [31] and extrapolated zero boundary conditions [32, 33]. It can also be easily extended to consider multiple frequencies.

Separating the real and imaginary parts of ϕ^c and A^c and letting $\phi = [\mathscr{R}\phi^{cT}, \mathscr{I}\phi^{cT}]^T$ and $A = [\mathscr{R}A^{cT}, \mathscr{I}A^{cT}]^T$, the complex Eq. (1) can be rewritten as a real equation:

$$\phi = A\mu, \tag{2}$$

with $\phi \in R^{N_{tot}}$, $A \in R^{N_{tot} \times N}$, and $N_{tot} = 2 \times N_m$. Assuming additive Gaussian noise e, the measurement model is given as

$$y = A\mu + e, \tag{3}$$

where $y \in R^{N_{tot}}$ denotes the measurement vector. In this paper, we assume e to be zero-mean and spatially uncorrelated with a covariance matrix $\sigma^2 I$. Namely,

$$e \sim \mathscr{N}(0, \sigma^2 I), \tag{4}$$

where $\mathscr{N}(\nu, \Sigma)$ denotes a Gaussian distribution with mean ν and covariance matrix Σ.

3. Inverse problem and expectation-maximization algorithm

In this section, we first present the inverse problem and interpret it from a regularization perspective. We then introduce the concept of sparsity regularization and formulate the inverse problem using the L_1 norm of the unknown vector μ. We propose solving the optimization problem using the EM algorithm and give the solution in the end.

3.1. Inverse problem from a regularization perspective

The inverse problem in DOT involves estimating μ from the measurements y. The problem is ill posed and highly underdetermined since N is typically much larger than N_{tot}. To solve this problem, regularization is usually required, where additional information is introduced about the solution to μ (e.g., an assumption on its smoothness or a bound on its norm). Using regularization, the inverse solution is formulated as

$$\hat{\mu} = \arg\max_{\mu}\{-||y - A\mu||^2 - \gamma g(\mu)\}, \tag{5}$$

where the first term in brackets denotes the least-square error, $g(\mu)$ the regularization function, and γ the regularization parameter controlling the tradeoff between the noise residual and the prior. In the Tikhonov regularization,

$$g(\mu) = ||\mu||_2^2 = \sum_i \mu_i^2, \tag{6}$$

which assumes that the solution μ is smooth over the domain. The corresponding solution is given as

$$\hat{\mu} = (A^T A + \gamma I)^{\#} A^T y, \tag{7}$$

where I denotes the identity matrix and $A^{\#}$ the pseudo-inverse of A. It can be solved either directly or iteratively using, for example, the conjugate gradient (CG) method.

Equation (5) can also be interpreted from a Bayesian point of view. If we assume that the measurements y are corrupted by Gaussian noise and μ is normally distributed with a zero mean, Eq. (5) is equivalently expressed as

$$\hat{\mu} = \arg\max_{\mu}\{\ln p(y|\mu) + \gamma \ln p(\mu)\}, \tag{8}$$

where $p(y|\mu)$ denotes the likelihood function with $\ln p(y|\mu) \propto -||y - A\mu||^2$ and $p(\mu)$ the prior distribution with $\ln p(\mu) \propto -||\mu||_2^2$. Therefore, the solution using the Tikhonov regularization is equivalent to the Bayesian solution assuming a Gaussian prior on μ.

3.2. Introduction of the sparsity regularization

Note that in the image reconstruction of DOT, the distribution of μ (i.e., the change of the absorption coefficient) is not necessarily Gaussian in 3D. In particular, it appears in a sparse format where most of its elements are zeros (corresponding to the unchanged background), and only a few of them are nonzero (corresponding to the abnormalities). Based on this fact, we propose incorporating this sparseness prior into the image regularization formulation to improve the reconstructed resolution.

The ideal measure of the sparseness of a vector x is its L_0 norm: $||x||_0$, which is the number of non-zero entries in x. However, the minimization/maximization involving $||x||_0$ is NP-hard (Nondeterministic Polynomial-time hard) and can be solved only using a combinatorial optimization approach. Instead, it is usual to use the L_1 norm $||x||_1$ as an approximated measure, where

$$||x||_1 = \sum_i |x_i|. \tag{9}$$

It has been shown that if the solution μ is sparse enough, the solution to the L_1 norm minimization is equivalent to minimizing the L_0 norm [34].

Under the measurement model (3) and the sparsity assumption on μ, our inverse problem is expressed as

$$\min ||\mu||_1 \quad \text{subject to} \quad ||y - A\mu|| \leq \xi^2, \tag{10}$$

where ξ^2 is a parameter on the noise level. It can be equivalently expressed in an unconstrained form as

$$\hat{\mu} = \arg\max_{\mu}\{-||y - A\mu||_2^2 - \gamma||\mu||_1\}. \tag{11}$$

Equation (11) is a convex problem but it is non-differentiable. It can be solved using linear or quadratic programming, but in this paper we propose solving it using the expectation-maximization (EM) algorithm. By decomposing measurement model (3), the EM method takes advantage of certain structures and results in computationally simple solutions.

3.3. Equivalent model and EM algorithm

We propose solving the optimization problem in (11) using the EM algorithm. The EM algorithm is a general method to obtain the maximum penalized log-likelihood estimator (MPLE) by introducing missing data and maximizing the complete penalized log-likelihood. In our case, the MPLE is expressed as

$$\hat{\mu} \;=\; \arg\max_{\mu}\{\log p(y|\mu) - \gamma||\mu||_1\} \tag{12}$$

$$\;=\; \arg\max_{\mu}\{-\frac{||y - A\mu||_2^2}{2\sigma^2} - \gamma||\mu||_1\}, \tag{13}$$

which provides the same solution as (11). To apply EM, we first reformulate the measurement model (3) so that

$$\begin{cases} x = \mu + \alpha e_1, \\ y = Ax + e_2, \end{cases} \tag{14}$$

where α is a positive constant. This reformulation is equivalent to expressing

$$e = \alpha A e_1 + e_2, \tag{15}$$

where e_1 and e_2 are independent, such that

$$e_1 \;\sim\; \mathcal{N}(0, I), \tag{16}$$
$$e_2 \;\sim\; \mathcal{N}(0, \sigma^2 I - \alpha^2 A A^T). \tag{17}$$

Note that in order for the matrix $\sigma^2 I - \alpha^2 A A^T$ to be positive definite (so that it is a valid covariance matrix), we must have

$$\alpha^2 \leq \sigma^2 / \beta_1, \tag{18}$$

where β_1 denotes the largest eigenvalue of $A A^T$.

The decomposition in (14) introduces a hidden variable x, which is the noisy version of the true absorption perturbation μ. By regarding x as the missing data, we can estimate the real μ using the EM algorithm. The EM algorithm produce a sequence of estimates $\hat{\mu}^{(k)}, k = 1, 2, \ldots$ by alternating two steps (see below) until some stopping criterion is met.

(1) E-step: Compute the conditional expectation of the complete log-likelihood (of y and x) given the observed data y and the current estimate $\hat{\mu}^{(k)}$. Namely, compute

$$Q(\mu, \hat{\mu}^{(k)}) = \mathrm{E}\left[\log p(y, x|\mu)|y, \hat{\mu}^{(k)}\right]. \tag{19}$$

In this particular case, we can show that it is equivalent to computing

$$\hat{x}^{(k)} = \hat{\mu}^{(k)} + \frac{\alpha^2}{\sigma^2} A^T (y - A\hat{\mu}^{(k)}). \tag{20}$$

See Appendix A for a detailed derivation.

(2) M-step: Update the estimated $\hat{\mu}^{(k)}$ according to

$$\hat{\mu}^{(k+1)} = \arg\max_{\mu} \left\{ Q(\mu, \hat{\mu}^{(k)}) - g(\mu) \right\}, \tag{21}$$

which in our case can be expressed as

$$\begin{aligned}
\hat{\mu}^{(k+1)} &= \arg\max_{\mu} \left\{ -\frac{||\mu - \hat{x}^{(k)}||^2}{2\alpha^2} - \gamma||\mu||_1 \right\} \\
&= \arg\max_{\mu} \left\{ -||\mu - \hat{x}^{(k)}||^2 - 2\alpha^2\gamma||\mu||_1 \right\}. \tag{22}
\end{aligned}$$

Equation (22) can be solved separately for each element $\hat{\mu}_i^{(k+1)}$, $i = 1, \ldots, N$, as

$$\hat{\mu}_i^{(k+1)} = \arg\max_{\mu_i} \left\{ -\mu_i^2 + 2\mu_i x_i - 2\alpha^2\gamma|\mu_i| \right\}, \tag{23}$$

where x_i denotes the ith element of x. According to [15, 35], (23) can be solved using a soft-threshold method [15]:

$$\hat{\mu}_i^{(k+1)} = \mathrm{sgn}(\hat{x}_i^{(k)})(|\hat{x}_i^{(k)}| - \gamma\alpha^2)_+, \tag{24}$$

where $(\cdot)_+$ denotes the positive part operator defined as $(x)_+ = \max\{x, 0\}$, and $\mathrm{sgn}(\cdot)$ is the sign function defined as $\mathrm{sgn}(x) = 1$ if $x > 0$, and $\mathrm{sgn}(x) = -1$ if $x < 0$.

3.4. Some comments

A few comments deserve mention here regarding the aforementioned image reconstruction algorithm.

3.4.1. On the model (14)

By introducing the hidden variable x, we decompose the inverse problem (the mapping of $y \rightarrow \mu$) into two parts: The first part can be interpreted from the typical image reconstruction point of view, which removes the effects of measurement noise (i.e., the mapping of $y \rightarrow x$); the second part can be regarded as a denoising procedure (the mapping of $x \rightarrow \mu$). This formulation can help improve the resolution of the reconstructed images, as we will show in Section 4.

3.4.2. On the convergence of the EM algorithm

Based on the information matrices, it has been shown in [36] that each iteration of EM is guaranteed to increase the penalized log-likelihood function, namely,

$$-\log p(y|\hat{\mu}^{(k+1)}) - \gamma||\mu^{(k+1)}||_1 \geq -\log p(y|\hat{\mu}^{(k)}) - \gamma||\mu^{(k)}||_1. \tag{25}$$

However, it does not mean that the sequence will converge to the maximum likelihood estimator of μ. If the distribution is multimodal, Wu showed in [37] that the EM algorithm is guaranteed to converge to a stationary point (i.e., the local maximum or saddle point) provided that the Q function and the penalty term are continuous in both μ and $\hat{\mu}$. The convergence performance also depends highly on the choice of the starting value $\mu^{(1)}$. To escape from the stationary point, we can use several different random initial points and also incorporate the prior information regarding the abnormality distribution.

3.4.3. On the soft-threshold method

In the M-step, we employed the soft-threshold method to implement the maximization. This approach is commonly used in the wavelet denoising scenario for image processing [15, 38, 39, 40], where the goal is to solve the problem of $\arg\min_x ||y - Ax||_2^2 + \gamma ||x||_1$ with A being orthogonal and x sparse. Certainly, other algorithms, such as the subgradient method and the interior-point method can also be used to solve (23), but we choose the thresholding algorithm for the sake of simplicity. From (20) and (24), we can see that our method has a computational complexity of $O(N^2)$ for the update in the E-step and $O(N)$ in the M-step.

3.4.4. On the choice of the parameters

We need to consider the following issues when determining the model parameters: (1) the convergence rate of the EM algorithm and (2) the effect of the sparsity regularization. The first factor is affected by α. According to the theory of the convergence rate of EM [36], α should be made as large as possible for a faster convergence. In our case, since we need to satisfy the condition $\alpha^2 \leq \sigma^2/\beta_1$ (see Subsection 3.3) for the validity of the EM model (14), we should choose α to be as close to $\sigma/\sqrt{\beta_1}$ as possible.

The effect of the sparsity regularization is controlled by α, γ, and σ as shown in Eqs. (13) and (22). Once σ is given and α selected based on (18), it is obvious that the higher the γ, the sparser the reconstructed image. According to the first-order optimality condition in the convex optimization, Kim *et al.* showed in [41] that for the general problem of $\arg\min_x ||Ax - y||_2^2 + \gamma||x||_1$, an upper bound on the useful range of γ is given as $(\gamma)_{max} = ||2A^Ty||_\infty$, where $||x||_\infty = \max_i\{x_i\}$. For γ that is higher than $(\gamma)_{max}$, the estimated $\hat{x}$ would have all its elements be zero. After applying this result to Eqs. (13) and (22), we derive that σ, γ, and α should satisfy the following equations:

$$\alpha^2\gamma \leq (\alpha^2\gamma)_{max} = ||x||_\infty, \tag{26}$$

$$\sigma^2\gamma \leq (\sigma^2\gamma)_{max} = ||A^Ty||_\infty. \tag{27}$$

Note that Eqs. (26) and (27) provide only an upper bound on the model parameters. Their main function is to avoid selecting a γ that is too large for obtaining a meaningful result. The right-hand side of Eq. (27) can be easily computed once the model is determined. Regarding the right-hand side of Eq. (26), although x denotes the noisy version of the true perturbation μ, its maximum can be estimated given certain prior information. For example, according to [26], the perturbation in the abnormality's absorption value ranges from 0.02 cm^{-1} to 0.3 cm^{-1} for most biological tissue, depending on the tissue type. Based on this information, we would be able to obtain a reasonable estimate of the bound on $\alpha^2\gamma$. In practice, we first determine the range of α, γ, and σ based on Eqs. (18), (26), and (27) and then select their values experimentally based on the signal-to-noise ratio to obtain the optimal reconstruction results.

4. Numerical examples

In this section, we provide numerical examples to illustrate the reconstruction results using our sparsity regularization and compare it with other methods. We considered a 3D transmission geometry as shown in Fig. 1, where the origin is at the center of the bottom surface ($z = 0$ cm) and the sides are of length 8 cm, 8 cm, and 6 cm along the x, y, and z directions, respectively. We placed 25 sources on the bottom surface at 1.75 cm intervals and 25 detectors on the top surface ($z = 6$ cm) at 1.5 cm intervals. We chose a source modulation frequency of 200 MHz and wavelength $\lambda = 750$ nm.

We chose the background optical parameters as $\mu_{a0} = 0.05$ cm^{-1}, $\mu'_{s0} = 9.5$ cm^{-1}, and the speed of light $c = 22$ cm/ns [42]. We divided the domain into small voxels with size $4 \times 4 \times 5$
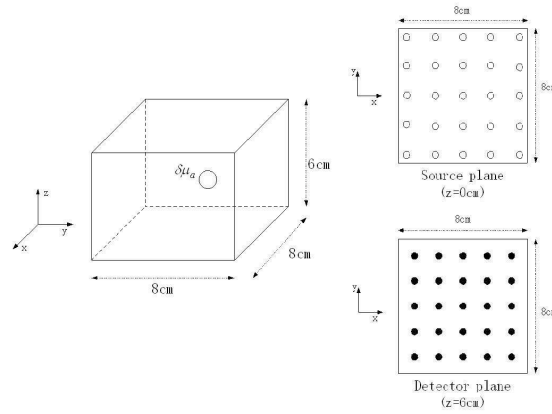
Fig. 1. Simulation setup. The domain is composed of an $8 \times 8 \times 6$ cm^3 cube, where twenty-five sources (circles) are placed on the bottom surface ($z = 0$ cm) and twenty-five detectors (black dots) on the top ($z = 6$ cm). The range along the x and y axes is $[-4, 4]$ cm.

mm^3 and set up the forward model (2) using the method of image sources with the extrapolated zero boundary condition. We added Gaussian noise with standard deviation $\sigma = 0.01$ and performed image reconstruction using four methods: (a) our L_1-EM approach, (b) the level-set algorithm proposed in [23], (c) the Tikhonov regularization, and (d) the simultaneous iterative reconstruction technique (SIRT). In the following, we first demonstrate the reconstruction results of four methods assuming one absorbing abnormality and then discuss their performances in separating two closely spaced abnormalities.

4.1. Image reconstruction results

We assumed that there is one spherical absorption abnormality centered at $[-1.5, 1.25, 2.9]$ cm with radius $R = 1$ cm and absorption perturbation $\delta \mu_a = 0.2$ cm^{-1}. The abnormality covers 54 voxels out of a total of 4,800 voxels in the domain. The original μ_a distribution is shown in Fig. 2, where each small image shows the cross-section layer at different z values at 0.5 cm intervals.

For our L_1-EM approach, we selected $\alpha = 3.5 \times 10^{-4}$, $\gamma = 10^4$, and stopped the EM iteration when $|\hat{u}^{(k+1)} - \hat{u}^{(k)}| \leq 10^{-3}$. Under the current setup, the solution converges after 400 iterations, taking about 45 seconds using Matlab 7 on a PC with Pentium 4 2.6 GHz CPU and 1 G of RAM. For the level-set algorithm, we used $\kappa = 0.4$, $\lambda = 0$, $\mu = 2 \times 10^{-4}$, and stopped the iteration when the relative change $e(k)$ is less than 10^{-8}; see [23] for the notation and implementation. We selected $\gamma = 0.05$ for the Tikhonov regularization (see Eq. (7)) and terminated the SIRT after 800 iterations.

The results are shown in Fig. 3. We can see that all of the four methods can recover the center location of the abnormality correctly. However, clear differences in performance can also be observed. More specifically,

- Due to the incorporation of the sparse nature of μ, our method and the level-set algorithm provide the best performance in terms of resolution. We observe the least amount of background noise and biggest contrast between the activation and background in Figs. 3a and b. The level-set algorithm overestimates the size a little bit and the L_1-EM approach appears to have a "spiky" effect due to the nature of the L_1 norm.
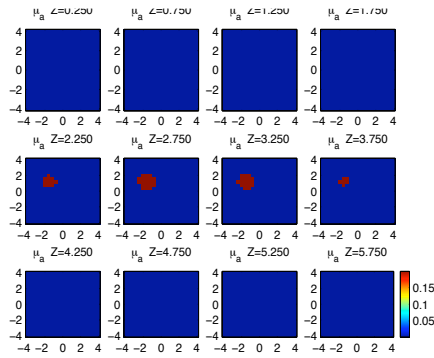
Fig. 2. Original $\delta\mu_a$ distribution in cm^{-1}, assuming only one absorbing abnormality. Small images show the cross-section layers at different z values at 0.5 cm intervals.

- The result using the Tikhonov regularization is blurred and appears as a Gaussian distribution due to the effect of the L_2 norm; see Fig. 3c. There is also more background noise.

- The SIRT introduces the biggest side-lobe in the reconstructed results. Furthermore, the reconstructed values are biased, with a zero initial vector.

From the above example, we can see that our method can provide image reconstruction result with fairly good resolution. Combined with other methods using Tikhonov regularization, it can be useful in determining the range of the activation with more accuracy.

4.2. Performance analysis

In this subsection, we study the performance of our method for separating two closely spaced absorbing abnormalities. We assumed two spherical absorptive abnormalities with radius $R = 0.75$ cm and $\delta\mu_a = 0.2$ cm^{-1}. We considered the following two cases: In the first case, the centers of these two spheres are at $[-1.5, 0.8, 2]$ cm and $[1.5, -0.8, 4]$ cm, with a distance of 4 cm between the two centers and 2.5 cm between the two closest points on the spheres. In the second case, the centers are moved to $[-1, 0.5, 2.5]$ cm and $[1, -0.5, 3.5]$ cm, and the distance becomes 2.45 cm between centers and 1 cm between the closest points. The original $\delta\mu_a$ distributions are shown in Fig. 4.

The reconstruction results are shown in Fig. 5 and Fig. 6, respectively. We can see that for the first case, all four methods can separate the two abnormalities with different levels of accuracy. In a similar fashion as the results shown in Fig. 3, the L_1-EM method and level-set algorithm reconstruct the problem with the best resolution, and the Tikhonov regularization and SIRT method show wider side lobes than the true distribution. When the two abnormalities get closer, the difference in the reconstruction performance becomes more obvious. We can tell that the SIRT method fails to differentiate the two abnormalities: There is only one big sphere showing around the midpoint along the two abnormalities. Both the L_1-EM and level-set algorithm can separate the abnormalities well. However, the level-set algorithm underestimates the size in this case. This might be due to the small size of the activations, leading to a poor approximation of curve evolution; the discretization of the problem on a finer grid may yield better estimates. The result using the Tikhonov regularization again shows bigger side lobes.
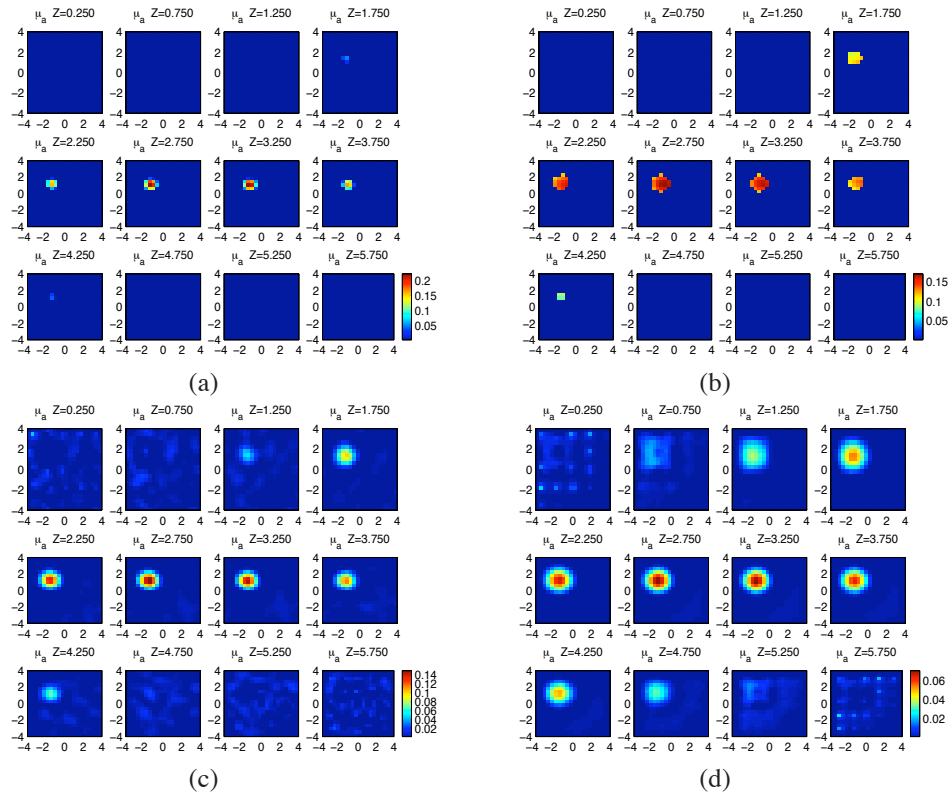
Fig. 3. Reconstructed $\delta\mu_a$ distributions in cm^{-1}, assuming only one absorbing abnormality as shown in Fig. 2. (a) Using the sparse regularization with EM algorithm; (b) using the level-set algorithm; (c) using the Tikhonov regularization; (d) using SIRT. Small images show the cross-section layers at different z values at 0.5 cm intervals.
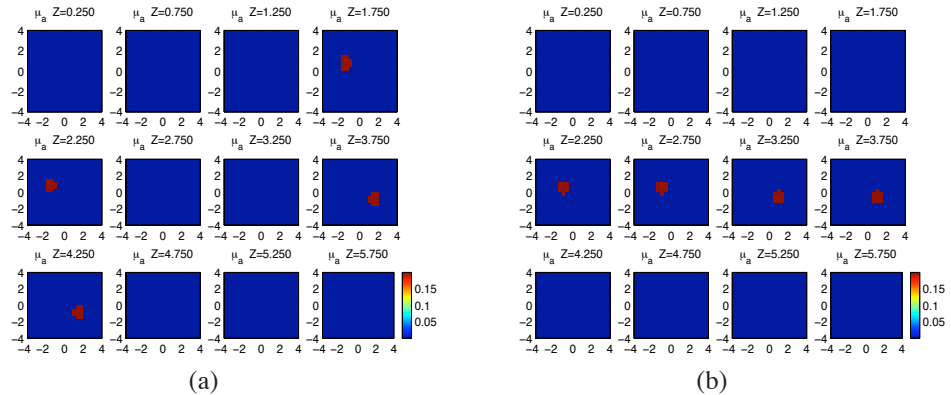

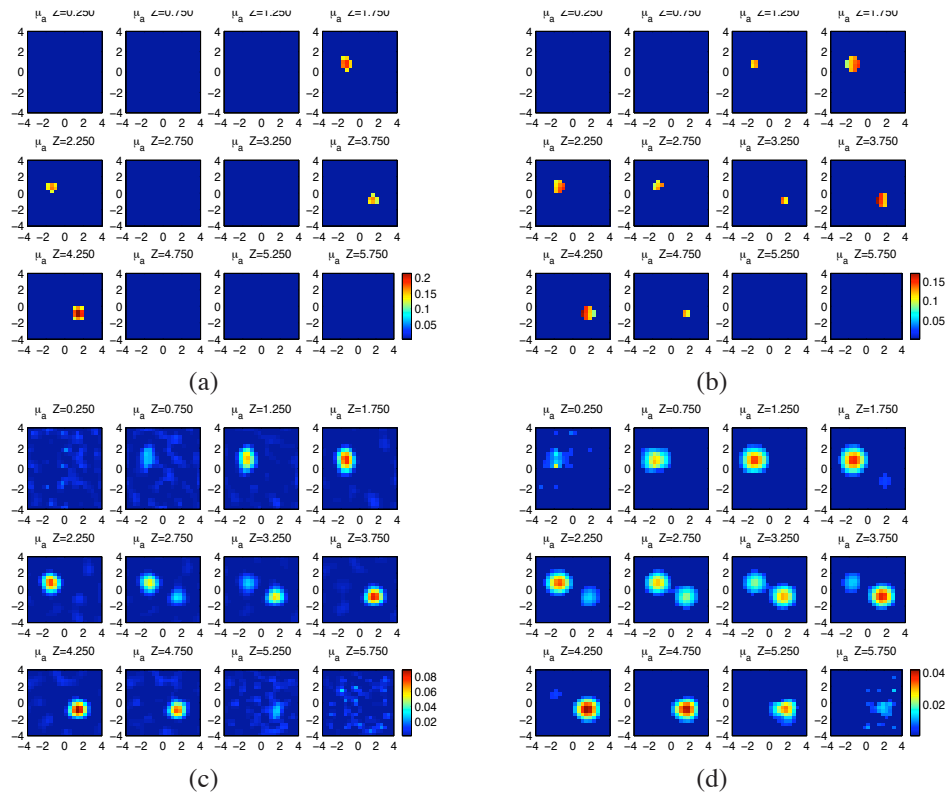Fig. 4. Original $\delta\mu_a$ distributions in cm^{-1}, assuming two spherical absorptive abnormalities.

Fig. 5. Reconstructed $\delta\mu_a$ distributions in cm^{-1}, assuming two spherical absorptive abnormalities as shown in Fig. 4(a). (a) Using the sparse regularization with EM algorithm; (b) using the level-set algorithm; (c) using the Tikhonov regularization; (d) using SIRT.

5. Conclusions

We proposed an image reconstruction method for diffuse optical tomography by introducing the sparsity regularization. We formulated the inverse problem by regularizing the L_1 norm of the unknown absorption coefficients and solved the optimization problem using the expectation-maximization (EM) method with a soft-threshold approach. We compared our method with other image reconstruction approaches and showed its efficiency in improving the reconstruction resolution.

In the current work, we applied our method to a transmission geometry, using the forward model based on the Rytov approximation to the diffuse approximation equation. This framework is useful for breast tumor detection. Our future work includes applying the proposed methods to the FEM forward model with a spherical domain shape and testing it using the real optical measurements from brain imaging. In order to apply our method to a nonlinear forward solver like FEM, a linearized model using the Taylor series needs to be obtained first for the applicability of the decomposition (14). More details on model linearization and computation of the Jacobian matrix for the FEM model can be found in [28, 29].
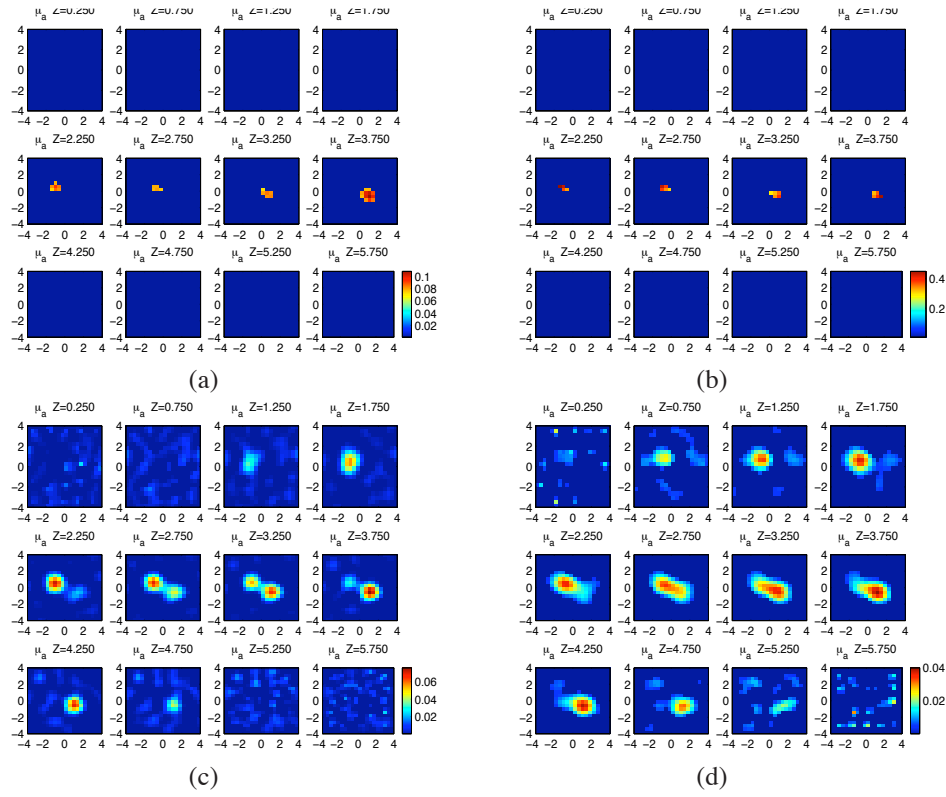
Fig. 6. Reconstructed $\delta\mu_a$ distributions in cm^{-1}, assuming two spherical absorptive abnormalities as shown in Fig. 4(b). (a) Using the sparse regularization with EM algorithm; (b) using the level-set algorithm; (c) using the Tikhonov regularization; (d) using SIRT.

Appendix A

In this appendix, we derive the E-step given in Section 3.3. The complete likelihood function is expressed as

$$p(y,x|\mu) = p(y|x,\mu)p(x|\mu) = p(y|x)p(x|\mu), \tag{28}$$

where the second equality follows because y is independent of μ when conditioned on x. Then we have

$$\log p(y,x|\mu) = -\frac{||x-\mu||^2}{2\alpha^2} + C_1 \tag{29}$$

$$= -\frac{\mu^T\mu - 2\mu^T x}{2\alpha^2} + C_2, \tag{30}$$

where C_1 and C_2 are constants that do not depend on μ. Comparing (29) with (19), we can see that the E-step is equivalent to computing the conditional expectation of x given the observed data y and current estimated parameter $\hat{\mu}^{(k)}$, i.e., computing

$$\hat{x}^{(k)} = \mathrm{E}\left[x|y,\hat{\mu}^{(k)}\right] = \int x p(x|y,\hat{\mu}^{(k)})dx. \tag{31}$$

Since

$$y|x \quad \sim \quad \mathcal{N}(Ax, \sigma^2 I - \alpha^2 AA^T), \tag{32}$$

$$x|\hat{\mu}^{(\kappa)} \quad \sim \quad \mathcal{N}(\hat{\mu}^{(\kappa)}, \alpha^2 I), \tag{33}$$

and

$$p(x|y, \hat{\mu}^{(k)}) = \frac{p(y|x)p(x|\mu^{(k)})}{p(y|\mu^{(k)})} \propto p(y|x)p(x|\mu^{(k)}), \tag{34}$$

it can be shown that $x|y, \hat{\mu}^{(k)}$ is also Gaussian with its mean given by

$$\hat{x}^{(k)} = \hat{\mu}^{(k)} + \frac{\alpha^2}{\sigma^2}A^T(y - A\hat{\mu}^{(k)}). \tag{35}$$