Performance evaluation of typical approximation algorithms for nonconvex ℓ_p-minimization in diffuse optical tomography

Calvin B. Shaw and Phaneendra K. Yalavarthy*

Supercomputer Education and Research Centre, Indian Institute of Science, Bangalore 560 012, India
**Corresponding author: phani@serc.iisc.ernet.in*

Received December 5, 2013; revised February 19, 2014; accepted February 19, 2014;
posted February 20, 2014 (Doc. ID 202160); published March 28, 2014

The sparse estimation methods that utilize the ℓ_p-norm, with p being between 0 and 1, have shown better utility in providing optimal solutions to the inverse problem in diffuse optical tomography. These ℓ_p-norm-based regularizations make the optimization function nonconvex, and algorithms that implement ℓ_p-norm minimization utilize approximations to the original ℓ_p-norm function. In this work, three such typical methods for implementing the ℓ_p-norm were considered, namely, iteratively reweighted ℓ_1-minimization (IRL1), iteratively reweighted least squares (IRLS), and the iteratively thresholding method (ITM). These methods were deployed for performing diffuse optical tomographic image reconstruction, and a systematic comparison with the help of three numerical and gelatin phantom cases was executed. The results indicate that these three methods in the implementation of ℓ_p-minimization yields similar results, with IRL1 fairing marginally in cases considered here in terms of shape recovery and quantitative accuracy of the reconstructed diffuse optical tomographic images. © 2014 Optical Society of America

OCIS codes: (170.0110) Imaging systems; (170.3010) Image reconstruction techniques; (170.3660) Light propagation in tissues; (170.3880) Medical and biological imaging.
http://dx.doi.org/10.1364/JOSAA.31.000852

1. INTRODUCTION

Near-infrared (NIR) diffuse optical tomography is one of the emerging imaging modalities, whose major applications include breast cancer imaging [1,2] and brain function assay [1,3]. The interrogating medium in the diffuse optical tomography is NIR light in the spectral range of 600–1000 nm. A finite set of boundary measurements are acquired in NIR tomography, which in turn gets used to reconstruct the internal distribution of optical properties [4]. The NIR light is delivered through optical fibers, and the transmitted light is collected typically through the same fibers, which are in contact with the surface (boundary) of the tissue. The distribution of optical properties of the tissue are reconstructed using this measured boundary data with the help of model-based iterative algorithms [4].

The inverse problem encountered in diffuse optical tomography is underdetermined and ill-posed due to the limited availability of boundary data and diffusive nature of NIR light propagation [1,4]. To overcome this problem, the estimation of optical properties involves a regularization scheme to obtain a unique solution. Several penalty methods [5–7] have been used to alleviate the effect of ill-posedness in the image-reconstruction procedure. The Tikhonov-type (quadratic norm; ℓ_2-norm-based) regularization is widely used, as it promotes smooth solutions and thereby suppresses the high-frequency components resulting in blobby reconstructed images [6,8].

Apart from the ℓ_2-norm-based regularization schemes, the ℓ_1-norm-based (absolute value) regularization schemes were also extensively studied [8–14], as they impose a sparsity constraint on the solution, which enables the recovery of sharp features embedded in the image. Both ℓ_2 and ℓ_1-norm-based regularization functional in the minimization scheme results in convex optimization, where direct inversion methods are available to perform the diffuse optical image reconstruction [8–10,14]. Recently, nonconvex-based regularization schemes [8,15] (ℓ_p-norm with $0 < p < 1$) have also been used in the image-reconstruction framework of diffuse optical tomography to further enhance the sparsity by reweighting procedures, iterative in nature. The ℓ_p-norm-based regularization scheme, introduced in the time-domain diffuse optical tomography [15], was shown to be more effective than the conventional ℓ_1-norm regularization. The nonconvex minimization problems such as those with the ℓ_p $(0 < p < 1)$ penalties have been receiving lot of attention in the recent past, mainly because of the need for much weaker incoherence conditions among the measurements and its robustness to data noise [16,17] along with promotion of sparser solutions compared to ℓ_1-norm regularization. As a result, the ℓ_p-regularization serves as a better choice compared to the widely used ℓ_1-regularization to obtain sparse solutions [18].

In the literature [15], application of ℓ_p-norm-based regularization is achieved using various methods and a close comprehensive comparison of these implementations is nonexistent for diffuse optical tomography. Such a study is essential to know the effective utility of the existing methods as well as for better implementation of ℓ_p-norm based regularization schemes in diffuse optical tomography. The existing ℓ_p-regularization algorithms can be divided mainly into three types [18]: iteratively reweighted least squares (IRLS) minimization [17], iteratively reweighted ℓ_1 (IRL1) minimization [19], and the iteratively thresholding method (ITM) [18]. As the name suggests, the first two methods essentially solve a series of weighted ℓ_2 and ℓ_1-optimization problems, and the ITM makes use of a threshold function to optimize the ℓ_p-regularization problem. Most methods proposed in the literature for diffuse optical tomographic imaging focused on comparing reconstruction results obtained using ℓ_p-norm-based regularization (when $0 \leq p \leq 1$) with $\ell_{2/1}$-norm-based regularization, specifically comparing nonconvex regularization (nontraditional) schemes versus convex regularization (traditional) schemes [8,15].

Inspired by the recent progress in ℓ_p-regularization schemes [17,18], this paper aims at performing a systematic comparison of these ℓ_p $(0 < p < 1)$ minimization algorithms—IRLS, IRL1, and ITM—for diffuse optical tomographic imaging. This comparison included numerical and experimental phantom cases, where the test cases involving resolving power, ability to recover sharp edges, robustness against noise levels, and finally a multilayered breast mimicking experimental-gelatin phantom case were considered. Further, a model function-based choice was adopted in all three ℓ_p-regularization algorithms to determine the regularization parameter for an unbiased comparison of the algorithms [20]. The model function-based approach was proposed for a general ℓ_p-minimization framework for bioluminescence tomography (BLT) [20] and was shown to be computationally efficient and superior compared to the traditional methods. The reconstruction results using different ℓ_p-algorithms were compared quantitatively via use of both mean value of the region of interest (ROI) and Pearson correlation (PC) as the figures of merit. The discussion in this study was limited to a 2D continuous wave (CW) case, where only the amplitude of the light is collected at the boundary, and, in turn, the unknown optical absorption coefficient distribution is recovered.

2. NIR DIFFUSE OPTICAL TOMOGRAPHY: FORWARD PROBLEM

CW NIR light propagation in thick biological tissues such as the breast and brain can be modeled using diffusion equation (DE) [4], which can be written as

$$-\nabla \cdot [D(r)\nabla\Phi(r)] + \mu_a(r)\Phi(r) = Q_o(r), \qquad (1)$$

where the optical diffusion and absorption coefficients are given by $D(r)$ and $\mu_a(r)$, respectively. The CW light source, represented by $Q_o(r)$, is modeled as isotropic. The photon density at a given position r is represented by $\Phi(r)$. The diffusion coefficient is defined as $D(r) = (1/3[\mu_a(r) + \mu_s'(r)])$, where $\mu_s'(r)$ was the reduced optical scattering coefficient. In the present work, μ_s' is assumed to be known and remains constant throughout the domain. The finite element method (FEM) is used to solve Eq. (1) to generate the modeled data for a given distribution of the absorption coefficient $\mu_a(r)$ [21]. Under the Rytov approximation, the modeled data becomes the natural logarithm of the intensity (A), $\ln(A)$. In this work, the Robin (or Type III) boundary condition is incorporated to account for the refractive-index mismatch at the boundary [4]. This forward model is used repeatedly in an iterative framework to estimate the optical properties of the tissue under investigation [4,21].

3. NIR DIFFUSE OPTICAL TOMOGRAPHY: INVERSE PROBLEM

The objective of the inverse problem is to recover the optical absorption coefficient (μ_a) image (distribution) from the CW boundary measurements using a model-based approach. This is accomplished by matching the experimental measurements (y) with the model-based ones $(G(\mu_a))$ iteratively in the least squares sense over the range of μ_a. The objective function, in this case, becomes

$$\Omega(\mu_a) = \|y - G(\mu_a)\|_2^2. \qquad (2)$$

The function $G(\mu_a)$ can be expanded using a Taylor series expansion around μ_{a0} (initial guess) as

$$G(\mu_a) = G(\mu_{a0}) + \mathbf{J}\Delta\mu_a + (\Delta\mu_a)^T\mathbf{H}\Delta\mu_a + \cdots \qquad (3)$$

with $\mathbf{J} = (\partial G(\mu_a)/\partial\mu_a)$ as the Jacobian [dimension: $M \times N$, M represents the number of measurements and N the number of FEM nodes (unknowns)], and $\mathbf{H}$ represents the Hessian (second-order derivative). The $\Delta\mu_a$ is the update given as $\mu_a - \mu_{a0}$. Linearizing the above equation leads to a cost function [5]:

$$\Omega(\Delta\mu_a) = \|\delta - \mathbf{J}\Delta\mu_a\|_2^2, \qquad (4)$$

with $\delta = y - G(\mu_a)$ being the data-model misfit. In order to obtain a stable solution while minimizing Eq. (4), a regularization term (also known as penalty term) involving the unknown $\Delta\mu_a$ is added to the cost function. The generalized ℓ_p-regularization framework for Eq. (4) can be written as

$$\Omega'(\Delta\mu_a) = \|\delta - \mathbf{J}\Delta\mu_a\|_2^2 + \lambda\|\Delta\mu_a\|_p^p. \qquad (5)$$

In this work, only cases $0 < p < 1$ are considered. Note that Jacobian $(\mathbf{J})$ is typically computed using the adjoint method [22], which is known to be a computationally efficient technique. The updated equation provides a direct estimation of $\Delta\mu_a$; after computation, it is added to the current μ_a, resulting in a new estimate of μ_a. This process is repeated until the difference in $\|\delta\|_2^2$ in successive iterations becomes less than 2%. The subsequent subsections introduce the mathematical/computational framework for the ℓ_p-regularization algorithms with $0 < p < 1$ that were considered in this work.

A. Iteratively Reweighted ℓ_1 Minimization

This involves solving a series of ℓ_1-minimization problems to approximate the optimal solution of the ℓ_p-minimization problem, popularly known as the reweighted scheme. The ℓ_p-minimization problem defined in Eq. (5) with $0 < p < 1$ is solved using a reweighted ℓ_1-minimization algorithm, and the equivalent update equation can be written as [18,19]

$$\Delta\mu_a^{(k+1)} = \arg\min_{\Delta\mu_a}\left\{\|\delta - \mathbf{J}\Delta\mu_a\|_2^2 + \sum_{i=1}^{N} w(i)^{(k+1)}|\Delta\mu_a(i)|\right\}, \qquad (6)$$

where $w^{(k+1)}$ is the weight assigned to the unknown parameter $(\Delta\mu_a)$ at the $(k+1)$th iteration given as

$$w(i)^{(k+1)} = \frac{\lambda * p}{(|\Delta\mu_a(i)^{(k)}| + \epsilon_k)^{1-p}}, \tag{7}$$

in here λ represents the regularization parameter with $0 < p < 1$, and $0 < \epsilon_k \ll 1$ is a small perturbation parameter to prevent division by zeros, and ϵ_k is usually a sequence of positive real numbers that decay to zero with iterations to improve the accuracy of the solution [19]. Note that setting a value of $p = 1$ results in a conventional ℓ_1-norm minimization with uniform weights dictated by λ. Since Eq. (6) is a weighted ℓ_1-minimization problem, there exists many convex minimization schemes such as the iterative shrinkage thresholding (IST) [23], gradient projection (GP) [24], split augmented Lagrangian shrinkage algorithm (SALSA) [25,26], and many more to solve this weighted ℓ_1-minimization problem. As ℓ_1-minimization function does not have a closed-form solution, an iterative technique is adopted.

In this work, Eq. (6) is solved using SALSA, as it is known to have better convergence speed compared to all other ℓ_1-norm-based algorithms [25]. The minimization scheme deployed in SALSA makes use of a variable splitting, which is achieved using an alternating direction method of multipliers (ADMM). The detailed description of the algorithm is given in [25,26]. The original SALSA was proposed for solving only the ℓ_1-minimization problems, since Eq. (6) involves solving a weighted ℓ_1-minimization, we modified the existing SALSA to solve a nonconvex ℓ_p-minimization problem. The main change in the algorithm involves calculating the weight updates [Eq. (7)] from the previous iteration. The modified weighted SALSA used to solve the IRL1 minimization scheme is given in Algorithm 1.

Algorithm 1. Iteratively Reweighted ℓ_1 Minimization Algorithm Using Weighted SALSA

```
INPUT: $\mathbf{J}$, $\delta$, $p$, $\lambda$, $\alpha$, $Tol$, $N_{it}$ and sequence of positive numbers $\epsilon_k$ such that $lim_{k\to\infty}\epsilon_k = 0$.
OUTPUT: $\Delta\mu_a$
Initialize: $d = 0$ and $k = 1$
  1. $\Delta\mu_a = \mathbf{J}^T\delta$, $r_1 = (\mathbf{J}\Delta\mu_a - \delta)$ and $\mathbf{F} = \mathbf{J}^\mathbf{T}\mathbf{J} + \alpha\mathbf{I}$
  2. $w = ((\lambda * p)/((|\Delta\mu_a| + \epsilon_k)^{1-p}))$ [Using Eq. (7)]
  3. $v = \mathrm{soft}(\Delta\mu_a + d, (0.5 * w)/\alpha) - d$
  4. $\Delta\mu_a = \mathbf{F}^{-1}(\mathbf{J}^\mathbf{T}\delta + \alpha v)$
  5. $d = \Delta\mu_a - v$
  6. STOP if $k = N_{it}$ or $\|r_{(k+1)} - r_{(k)}\|^2 \le Tol$; otherwise set $k = k + 1$, Go TO Step 2
```

The inputs to Algorithm 1 are Jacobian $(\mathbf{J})$ matrix, δ, $0 < p < 1$, regularization parameter (λ) for ℓ_1-update, regularization parameter (α) for ℓ_2-update, Tol, indicates the absolute tolerance between consecutive residuals (r) of the solution $(\Delta\mu_a)$, maximum number of iterations (N_{it}), and ϵ_k represents the sequence of decaying positive numbers. Note that the value of α dictates the ℓ_2-weight update and $0 < \alpha < \max_{\mathrm{eig}}(\mathbf{J}^\mathbf{T}\mathbf{J})$, where $\max_{\mathrm{eig}}$ represents the maximum Eigen value of $\mathbf{J}^\mathbf{T}\mathbf{J}$. This value was set to 0.1 in all cases. The ADMM parameter (d) has a similar size of $\Delta\mu_a$, which is initialized to the zero vector, and the initial guess for the algorithm is obtained using Step 1 in the algorithm. To overcome the nondifferentiability nature of the ℓ_1-norm, a soft-threshold operation is used in Step 2. The soft-threshold represents the maximum absolute value amongst $\Delta\mu_a + d$ and $(0.5 * w)/\alpha$. Step 3 involves obtaining the update by solving normal equations, and finally Step 4 updates the ADMM parameter (d).

It is evident from Algorithm 1 that the crucial parameters that decide the sparsity level in the solution $(\Delta\mu_a)$ are the regularization parameter (λ) for ℓ_1-update and the value of p in the ℓ_p-norm. These parameters will be optimally chosen to reconstruct $\Delta\mu_a$ using a model function-based approach to determine the regularization parameter λ and the data-model misfit function [Eq. (2)] to decide the optimal p-value (discussed subsequently).

B. Iteratively Reweighted Least Squares Minimization

This scheme is similar to IRL1, except the approximation uses ℓ_2-minimization as opposed to ℓ_1-minimization. Specifically, it involves deriving the iterative weights and thereafter using the weights to solve a sequence of weighted ℓ_2-minimization problems. The equivalent unconstrained objective function can be approximated as

$$\Omega_{\mathrm{IRLS}}(\Delta\mu_a) = \arg\min_{\Delta\mu_a}\left\{\lambda\sum_{i=1}^{N}(\Delta\mu_a(i)^2 + \epsilon)^{\frac{p}{2}} + \frac{1}{2}\|\delta - \mathbf{J}\Delta\mu_a\|_2^2\right\}. \tag{8}$$

The first-order necessary condition is used to find the minimizer of the above function, given as

$$\sum_{i=1}^{N}\left(\frac{\lambda * p * \Delta\mu_a(i)}{(\epsilon + \Delta\mu_a(i)^2)^{1-\frac{p}{2}}}\right) + \mathbf{J}^\mathbf{T}(\mathbf{J}\Delta\mu_a - \delta) = 0. \tag{9}$$

The corresponding iterative scheme of the above equation as suggested in [27] is given as

$$\sum_{i=1}^{N}\left(\frac{\lambda * p * \Delta\mu_a(i)^{(k+1)}}{(\epsilon + (\Delta\mu_a(i)^{(k)})^2)^{1-\frac{p}{2}}}\right) + \mathbf{J}^\mathbf{T}(\mathbf{J}\Delta\mu_a^{(k+1)} - \delta) = 0, \tag{10}$$

or equivalently the normal equations can be written as,

$$\left\{\mathbf{J}^\mathbf{T}\mathbf{J} + \mathrm{diag}\left(\frac{\lambda * p}{(\epsilon + (\Delta\mu_a(i)^{(k)})^2)^{1-\frac{p}{2}}}\right)\right\}\Delta\mu_a^{(k+1)} = \mathbf{J}^\mathbf{T}\delta. \tag{11}$$

The above normal equation represents the weighted ℓ_2-minimization scheme, it is worth noting that when $p = 2$ and $\epsilon = 0$, the above expression reduces to the traditional ℓ_2 minimization scheme without any weights. Hence the nonconvex ℓ_p-penalty $(0 < p < 1)$ function is approximated by a quadratic penalty term in Eq. (8). Similar to IRL1, IRLS also chooses a sequence of positive numbers ϵ_k such that $\lim_{k\to\infty}\epsilon_k = 0$ to improve the numerical accuracy. The corresponding IRLS algorithm is summarized in Algorithm 2 [18].

Algorithm 2. Iteratively Reweighted Least Squares Minimization Algorithm

```
INPUT: $\mathbf{J}$, $\delta$, $p$, $\lambda$, $Tol$, $N_{it}$ and sequence of positive numbers $\epsilon_k$ such that $lim_{k\to\infty}\epsilon_k = 0$.
OUTPUT: $\Delta\mu_a$
Initialize: $k = 1$, $\Delta\mu_a^{(1)} = \mathbf{J}^T\delta$ and $r_1 = (\mathbf{J}\Delta\mu_a^{(1)} - \delta)$
  1. Solve the following weighted-normal equation:
       $\left\{\mathbf{J}^{\mathbf{T}}\mathbf{J} + diag\left(\frac{\lambda * p}{(\epsilon_k + (\Delta\mu_a(i)^{(k)})^2)^{1-\frac{p}{2}}}\right)\right\}\Delta\mu_a = \mathbf{J}^{\mathbf{T}}\delta$
  2. Update $\Delta\mu_a^{(k)}$ with $\Delta\mu_a$
  3. STOP if $k = N_{it}$ or $\|r_{(k+1)} - r_{(k)}\|^2 \leq Tol$; otherwise set $k = k + 1$, Go TO Step 1
```

One of the main advantages of IRLS lies with the existence of a closed-form solution to solve the minimization function defined by the normal equations [Eq. (11)]. On the other hand, since the ℓ_p-penalty ($0 < p < 1$) function is approximated by the ℓ_2-norm, the output of IRLS can contain few entries with large magnitudes and many entries with very small magnitudes. In other words, the solution given by IRLS could be considered as approximation to the solution obtained by IRL1. Note that the crucial parameters that dictate the sparsity level in the IRLS minimization algorithm are the regularization parameter (λ) and the value of p in the ℓ_p-norm.

C. Iteratively Thresholding Method

This method relies on minimizing the given objective function that could be approximated by a thresholding function [18,28]. The general penalized regularization problem can be defined as

$$\Omega_{\mathrm{ITM}}(\Delta\mu_a) = \arg\min_{\Delta\mu_a}\left\{\frac{1}{2}\|\delta - \mathbf{J}\Delta\mu_a\|_2^2 + P(\Delta\mu_a;\lambda)\right\}, \tag{12}$$

where $P(\Delta\mu_a;\lambda)$ is a penalty function. In order to minimize Eq. (12), an auxiliary variable z is introduced, making it equivalent to

$$\Omega'_{\mathrm{ITM}}(\Delta\mu_a, z) = \frac{1}{2}\|\delta - \mathbf{J}\Delta\mu_a\|_2^2 + P(\Delta\mu_a;\lambda) + \frac{1}{2}(\Delta\mu_a - z)^T(\mathbf{I} - \mathbf{J}^{\mathbf{T}}\mathbf{J})(\Delta\mu_a - z). \tag{13}$$

It was shown in [28] that if $\mathbf{J}$ has been scaled properly such that $\|\mathbf{J}\|_2 < 1$, where $\|\mathbf{J}\|_2$ represents the largest singular value of $\mathbf{J}$, then $\mathbf{I} - \mathbf{J}^{\mathbf{T}}\mathbf{J}$ is a positive definite matrix. In this work, this condition was ensured by normalizing the columns of the $\mathbf{J}$ matrix with respect to its Euclidean length (ℓ_2)-norm. As a result, minimizing $\Omega'_{\mathrm{ITM}}(\Delta\mu_a, z)$ over $(\Delta\mu_a, z)$ is equivalent to minimizing $\Omega_{\mathrm{ITM}}(\Delta\mu_a)$ over $\Delta\mu_a$. Now the iterative update scheme for $\Delta\mu_a$ is given as

$$\Delta\mu_a^{(k+1)} = \arg\min_{\Delta\mu_a}\left\{\frac{1}{2}\|\Delta\mu_a - ((\mathbf{I} - \mathbf{J}^{\mathbf{T}}\mathbf{J})z^{(k)} + \mathbf{J}^T\delta)\|_2^2 + P(\Delta\mu_a;\lambda)\right\}. \tag{14}$$

Now with the knowledge of $\Delta\mu_a^{(k+1)}$ minimizing over $\Omega'_{\mathrm{ITM}}(\Delta\mu_a^{(k+1)}, z)$ results in $z^{(k+1)} = \Delta\mu_a^{(k+1)}$. Finally the iterative update equation in terms of $\Delta\mu_a^{(k+1)}$ and $\Delta\mu_a^{(k)}$ is expressed in a compact form [18] as follows:

$$\Delta\mu_a^{(k+1)} = \mathbf{\Gamma}((\mathbf{I} - \mathbf{J}^{\mathbf{T}}\mathbf{J})\Delta\mu_a^{(k)} + \mathbf{J}^{\mathbf{T}}\delta;\lambda), \tag{15}$$

where $\mathbf{\Gamma}$ represents the thresholding function corresponding to a penalty function ($P(\Delta\mu_a;\lambda)$). For the penalty function, $P(\Delta\mu_a;\lambda) = \lambda\|\Delta\mu_a\|_p^p$, the corresponding threshold function can be written as (see [28] for more details)

$$\Gamma_{\ell_{\mathbf{p}}}(\Delta\mu_{\mathbf{a}};\lambda) = \begin{cases} 0, & \text{for } \Delta\mu_a \leq \tau(\lambda) \\ \mathrm{sgn}(\Delta\mu_a)\max(\theta: g(\theta) = |\Delta\mu_a|), & \text{for } \Delta\mu_a > \tau(\lambda) \end{cases}, \tag{16}$$

where $\tau(\lambda) = \lambda^{(1/(2-p))}(2 - p)[p/(1 - p)^{1-p}]^{(1/(2-p))}$ is called the threshold parameter, which is a function of regularization parameter (λ) and the value of p. Moreover, $\tau(\lambda)$ is the function that constrains the level of sparsity in the solution and has to be appropriately chosen. Note $g(\theta;\lambda) = \theta + \lambda p\theta^{(p-1)}$, and it attains its optimal (minimum) $\tau(\lambda)$ at $\theta_0 = \lambda^{(1/(2-p))}[p(1 - p)]^{(1/(2-p))}$ [28]. As a result, given any $\Delta\mu_a > \tau(\lambda)$, the expression $g(\theta) = t$ has one and only one root bounded in $[\theta_0, +\infty)$, since $g(\theta)$ is a strictly increasing function [28], and the root is found using Newton's method.

In order to improve the convergence property of the ITM algorithm, the following updating scheme was proposed:

$$\Delta\mu_a^{(k+1)} = \Gamma_{\ell_{\mathbf{p}}}((\mathbf{I} - \|\mathbf{J}\|_2^{-2}\mathbf{J}^{\mathbf{T}}\mathbf{J})\Delta\mu_a^{(k)} + \|\mathbf{J}\|_2^{-2}\mathbf{J}^{\mathbf{T}}\delta;\lambda\|\mathbf{J}\|_2^{-2}). \tag{17}$$

The corresponding ITM is summarized in Algorithm 3.

Algorithm 3. Iteratively Thresholding Method

```
INPUT: $\mathbf{J}$, $\delta$, $p$, $\lambda$, $Tol$, $N_{it}$.
OUTPUT: $\Delta\mu_a$
Initialize: $k = 1$, $\Delta\mu_a^{(1)} = \mathbf{J}^T\delta$ and $r_1 = (\mathbf{J}\Delta\mu_a^{(1)} - \delta)$
  1. Solve the following update equation:
       $\Delta\mu_a = \Gamma_{\ell_{\mathbf{p}}}((\mathbf{I} - \|\mathbf{J}\|_2^{-2}\mathbf{J}^{\mathbf{T}}\mathbf{J})\Delta\mu_a^{(k)} + \|\mathbf{J}\|_2^{-2}\mathbf{J}^{\mathbf{T}}\delta;\lambda\|\mathbf{J}\|_2^{-2})$
  2. Update $\Delta\mu_a^{(k)}$ with $\Delta\mu_a$
  3. STOP if $k = N_{it}$ or $\|r_{(k+1)} - r_{(k)}\|^2 \leq Tol$; otherwise set $k = k + 1$, Go TO Step 1
```

The IRL1 and IRLS have close similarities because they approximate the nonconvex ℓ_p-norm ($0 < p < 1$) with the ℓ_1 and ℓ_2-norm, which involves a series of reweighted ℓ_1 and ℓ_2-minimization problems, respectively. On the other hand, ITM does not derive iterative weights; instead, it solves the minimization problem using a threshold function.

D. Optimal Selection of Regularization Parameter (λ)

The outcome (solution) of IRL1, IRLS, and ITM algorithms depends essentially on the choice of regularization parameter (λ). In order to have an unbiased comparison of these three algorithms, we adopted the well-known Morozov discrepancy principle, which uses the model function approach [20]. This approach was well studied in the image-reconstruction framework for a generalized ℓ_p-regularization problem in multispectral BLT [20]; hence it is only briefly reviewed here.

The generalized objective function for the ℓ_p-regularization framework as a function of λ is defined as

$$F(\lambda) = \|\delta - \mathbf{J}\Delta\mu_{a(\lambda)}\|_2^2 + \lambda\|\Delta\mu_{a(\lambda)}\|_p^p. \tag{18}$$

The iterative update scheme for determining the λ, based on the model function [20,29], is given as

$$\lambda_{k+1} = \frac{C_k}{\sigma\tilde{m} - b} - T_k, \tag{19}$$

where k denotes the iteration number, $\sigma > 1$, $\tilde{m} = F(\lambda_k) - \lambda_k F'(\lambda_k)$ is the tangent at the point $(\lambda_k,\ F(\lambda_k))$, $b = \|\delta\|_2^2$, and $F'(\lambda_k) = \|\Delta\mu_{a(\lambda_k)}\|_p^p$. The relation among the model function parameters is given by [29]

$$C_k = -\frac{(b - F(\lambda_k))^2}{F'(\lambda_k)}, \qquad T_k = \frac{b - F(\lambda_k)}{F'(\lambda_k) - \lambda_k}. \tag{20}$$

For all three algorithms, we set the initial guess for regularization parameter $\lambda_0 = 0.5|\mathbf{J}^T\delta|_\infty$ and $\sigma = 1.01$. The steps involved in this adaptive choice of regularization parameter (λ) are given in Algorithm 4.

Algorithm 4. Adaptive Estimation of Regularization Parameter (λ)

```
INPUT: $Tol$, $\sigma = 1.01$ and $N_{it}$.
OUTPUT: $\Delta\mu_a$
Initialize: $k = 0$ and $\lambda_0 = 0.5|\mathbf{J}^T\delta|_\infty$
  1. Solve Eq. (18) for
       $\Delta\mu_{a(\lambda_k)}$ using Algorithm 1 for IRL1 (or)
       $\Delta\mu_{a(\lambda_k)}$ using Algorithm 2 for IRLS (or)
       $\Delta\mu_{a(\lambda_k)}$ using Algorithm 3 for ITM
  2. Evaluate $F(\lambda_k)$ and $F'(\lambda_k)$
  3. Update $C_k$, $T_k$ and $\tilde{m}$ using Eq. (20)
  4. Update $\lambda_{k+1}$ using Eq. (19)
  5. STOP if $k = N_{it}$ or $\|\lambda_{k+1} - \lambda_k\| \le Tol$; otherwise set $k = k + 1$,
Go TO Step 1
```

As mentioned earlier, the p-value in the ℓ_p penalty function also contributes in deciding the sparsity level apart from the regularization parameter (λ). As a result, the value of p was increased in the steps of 0.05 $(0 < p \le 1)$, and then a plot of p versus $\|y - G(\mu_{a(\mathrm{rec})})\|_2$ was considered to choose the p that resulted in a minimum value of $\|y - G(\mu_{a(\mathrm{rec})})\|_2$. Note that $G(\mu_{a(\mathrm{rec})})$ represents the model data computed with the reconstructed absorption coefficient distribution that uses optimal λ (achieved through Algorithm 4).

All methods were implemented with the help of MATLAB-based open-source NIR light propagation computational framework (NIRFAST) [21]. These algorithms are also available as open-source for the interested users [30]. The reconstructions were performed on a Linux workstation that had 2.4 GHz Intel Quadcore processor along with 8 GB RAM. The details of quantitative analysis of simulation and experimental studies performed as part of this work is presented in the following sections.

E. Quantitative Analysis

The PC was used as a figure of merit to evaluate the reconstructions obtained from different algorithms that were presented. The PC essentially measures the degree of correlation between the target $(\mu_a^{\mathrm{target}})$ and reconstructed (μ_a^{recon}) absorption coefficient distribution having a range of values from −1 to 1 and is widely used to evaluate reconstructed images in the emission tomography [31]. This measure essentially describes the detectability of the structures in the target image and is defined as [31]

$$\mathrm{PC}(\mu_a^{\mathrm{target}}, \mu_a^{\mathrm{recon}}) = \frac{\mathrm{COV}(\mu_a^{\mathrm{target}}, \mu_a^{\mathrm{recon}})}{\sigma(\mu_a^{\mathrm{target}})\sigma(\mu_a^{\mathrm{recon}})}, \tag{21}$$

where COV and σ represent the covariance and standard deviation, respectively.

The other metric used to quantify the accuracy of the contrast recovered in the ROI is the mean value in the ROI. Closer the mean value in the ROI to the expected value indicates the better quantitative accuracy (contrast recovery) of the reconstructed algorithms.

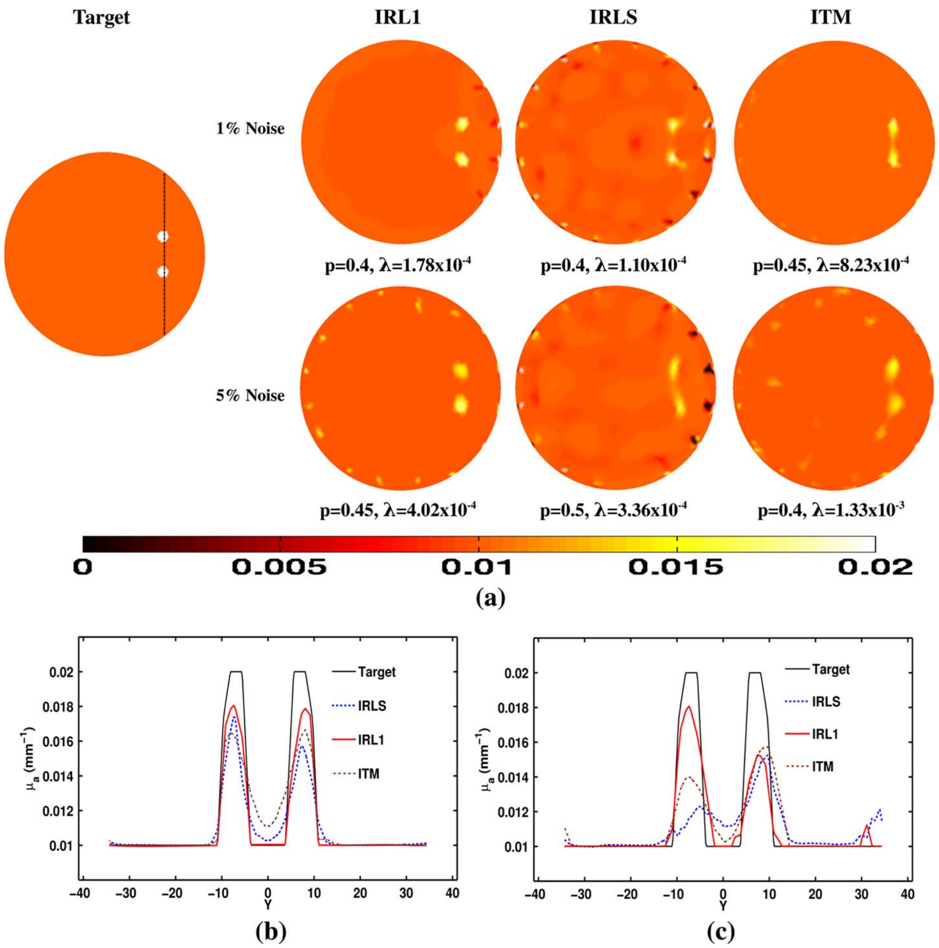
4. SIMULATION AND EXPERIMENTAL EVALUATION

A. Numerical Experimental Data

In order to evaluate different characteristics of the reconstructed images using the three nonconvex algorithms presented, numerical experiments involving three different targets were conducted. In these numerical experiments, the imaging domain was considered to be in a circular shape having a radius of 43 mm. The source-detector fibers were placed on the boundary. The data-collection setup had 16 fibers arranged in an equispaced fashion along the boundary of the imaging domain, when one fiber acted as a source, and the rest acted as detectors, resulting in 240 (16×15) measurement points [32]. All sources were positioned at one mean-free transport length inside the boundary and were modeled as having a Gaussian profile with an FWHM of 3 mm to mimic the experimental conditions [32]. The background optical properties, absorption coefficient (μ_a), and reduced scattering coefficient (μ_s') of the imaging domain were set to 0.01 mm^{-1} and 1.00 mm^{-1}, respectively. The refractive index of the imaging domain is set to be 1.33. The experimental data $[y = \ln(I)]$ was generated using a fine FEM mesh having 10,249 nodes (corresponding to 20,160 linear triangular elements), whereas the reconstruction was performed on a coarse FEM mesh with 1933 nodes (corresponding to 3726 triangular elements).

Initial numerical experiment consisting of two circular absorbing targets mimicking the tumor region was considered. The targets were circular in shape with a radius of 2.5 mm, had a contrast of 2:1 compared to the background, and were separated by a distance of 10 mm. These two targets were centered around (25, 7.5) and (25, −7.5). The tumor optical properties were set to $\mu_a = 0.02\ \mathrm{mm}^{-1}$ and $\mu_s' = 1\ \mathrm{mm}^{-1}$, whose μ_a distribution is given in Fig. 1 (first column). The numerically generated data was added with 1% (typical experimental case) and 5% (for noise-tolerance evaluation) normally distributed Gaussian noise.

Next, similar to an earlier case (Fig. 1), two rectangular targets were placed close to the center of the imaging domain



(a)

(b)

(c)

Fig. 1. (a) Reconstructed μ_a images using IRL1, IRLS, and ITM algorithms in the numerical experiment with 1% and 5% noisy data collected using the target distribution (given in the first column). The first and second rows show the reconstructed images obtained for the 1% and 5% noisy data, respectively. The reconstruction methods used are given at the top of each reconstructed image. The optimal value of p and optimal regularization parameter (λ) are also indicated at the bottom of each reconstructed image. The 1D cross-sectional profile along the dashed line of the target distribution for reconstructed μ_a images for 1% and 5% noisy cases are given in (b) and (c), respectively.

in order to test the ability to recover sharp features. The rectangles were of size 9×4 mm placed at (0, 10) and (0, −10); these targets had similar optical properties as in the earlier case. The numerically generated data had 1% normally distributed Gaussian noise, and the expected target distribution is as given in Fig. 2 (first column).

The third numerical test problem had a rectangular absorber placed horizontally of size 6×15 mm. The absorber had similar optical properties as mentioned earlier. The numerically generated data had 1% normally distributed Gaussian noise, and the expected target distribution is as given in Fig. 3 (first column). In all these cases, the numerical data was calibrated to remove biases introduced due to irregularities in modeling source-detectors [32].

B. Gelatin Phantom Data

In order to further effectively assess the reconstruction performance of different nonconvex algorithms, an experimental gelatin phantom data set was considered. In this case, a multilayered gelatin phantom of diameter 86 mm and height 25 mm was fabricated with heated mixtures of water (80%) and gelatin (20%). India ink was used for absorption and for scattering TiO_2 (titanium oxide powder) to obtain different optical properties [33]. Three distinct layers of gelatin were constructed by repeatedly hardening gel solutions to contain different amounts of ink and TiO_2 for varying optical absorption and scattering, respectively. To mimic a tumor, a cylindrical hole of diameter 16 mm and height 24 mm was filled with intralipid liquid [6]. The phantom's optical properties were estimated for 785 nm wavelength. The expected 2D cross section of the phantom is shown in the top left corner of Fig. 4(a). The region labeled as "0" represents the fatty layer with $\mu_a = 0.0065\ \mathrm{mm}^{-1}$ and $\mu_s' = 0.65\ \mathrm{mm}^{-1}$. The region labeled as "1" represents the fibro-glandular layer with $\mu_a = 0.01\ \mathrm{mm}^{-1}$ and $\mu_s' = 1.0\ \mathrm{mm}^{-1}$. Finally, the region labeled as "2" represents the tumor region with $\mu_a = 0.02\ \mathrm{mm}^{-1}$ and $\mu_s' = 1.2\ \mathrm{mm}^{-1}$ as the optical properties. The data collection setup had 16 fibers arranged in a circular fashion along the boundary of the imaging domain, where, when one fiber acted as a source, and the rest acted as detectors, resulting in 240 (16×15) measurement points [6]. This data was calibrated using a reference homogenous phantom to obtain the initial guess for optical properties (here, μ_a) [32]. For the reconstruction, an FEM mesh of 1785 nodes corresponding to 3418 linear triangular elements was used.

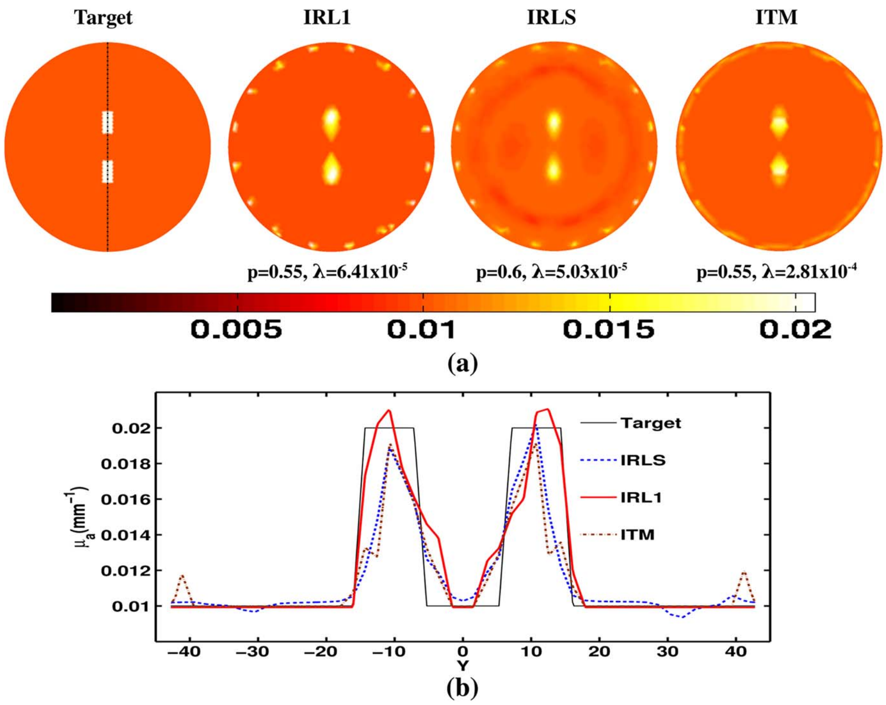


Fig. 2. (a) Reconstructed μ_a images using IRL1, IRLS, and ITM algorithms in the numerical experiment with 1% noisy data collected using the target distribution, which is rectangular in shape and placed close to the center of the imaging domain (given in the first column). The second, third, and fourth columns show the reconstructed images obtained for the 1% noisy data. The reconstruction methods used are given at the top of each reconstructed image. The optimal value of p and optimal regularization parameter (λ) are also indicated at the bottom of each reconstructed image. The 1D cross-sectional profile along the dashed line of the target distribution for reconstructed μ_a images is given in (b).

5. RESULTS

The results pertaining to the evaluation of resolving power of the three algorithms with two circular absorbing targets that are placed close to each other are given in Fig. 1(a) for the data noise level of 1% and 5% in the first and second rows, respectively. The target distribution is given in the first column. The algorithm used for performing the reconstruction was indicated on top of each reconstructed image in Fig. 1(a). The optimal p-value of the ℓ_p-norm and the regularization parameter (λ) were also indicated at the bottom of each reconstructed image in Fig. 1(a). Figures 1(b) and 1(c) are the 1D cross-sectional plots of the reconstructed images corresponding to 1% and 5% noisy data cases. It is evident from these results that visually the IRL1 reconstruction result is more appealing compared to others, and the ITM is effective in reducing the boundary artifacts. The optimal p in the ℓ_p-norm that was utilized was similar for all three cases (range 0.4–0.5). The quantitative comparison of these reconstructed results presented in Fig. 1(a) for the data noise level of 1% and 5% averaged over 10 trials, and the corresponding mean and standard deviation is given in Tables 1 (PC) and 2 (mean μ_a value in the ROI).

Similar to the earlier case, two rectangular structures were placed close to each other around the center of the imaging domain (where the sensitivity is poor); the reconstruction results are given in Fig. 2(a). The 1D cross-sectional plots along the dotted line of the target image are given in Fig. 2(b). In terms of shape recovery, the IRL1 performance is marginally superior compared with the others. The quantitative comparison of these reconstructed results presented in Fig. 2 are given in Tables 1 and 2.

The reconstruction results pertaining to rectangular structure placed horizontally close to the boundary of the imaging domain were presented in Fig. 3(a) for varying values of p. The 1D cross-sectional plots of the reconstructed results

Table 1. Pearson Correlation of the Reconstructed Optical Properties for the Results Presented in This Work[a]

Figure	IRL1	IRLS	ITM
Fig. 1 (1%)	0.788 ± 0.0127	0.344 ± 0.0152	0.759 ± 0.0111
Fig. 1 (5%)	0.247 ± 0.0115	0.085 ± 0.0094	0.223 ± 0.0132
Fig. 3	0.650	0.569	0.644
Fig. 4	0.645	0.396	0.6182
Fig. 5	0.479	0.3201	0.488

[a] The standard deviations observed among the reconstructed μ_a distributions as in Fig. 1 using 10 realizations of the noisy data (with fixed noise levels of 1% and 5%) are given in the first and second rows.

Table 2. Mean Reconstructed μ_a in the Region of Interest [Target(s)] for the Results Presented in This Work[a]

Figure	IRL1	IRLS	ITM
Fig. 1 (1%)	0.0153 ± 0.00092	0.0142 ± 0.00051	0.0148 ± 0.00072
Fig. 1 (5%)	0.0148 ± 0.00095	0.0128 ± 0.00065	0.0135 ± 0.00069
Fig. 3	0.0170	0.0151	0.0160
Fig. 4	0.0173	0.0162	0.0168
Fig. 5	0.0196	0.0168	0.0174

[a] The standard deviations observed among the reconstructed μ_a distributions as in Fig. 1 using 10 realizations of the noisy data (with fixed noise levels of 1% and 5%) are given in the first and second rows.

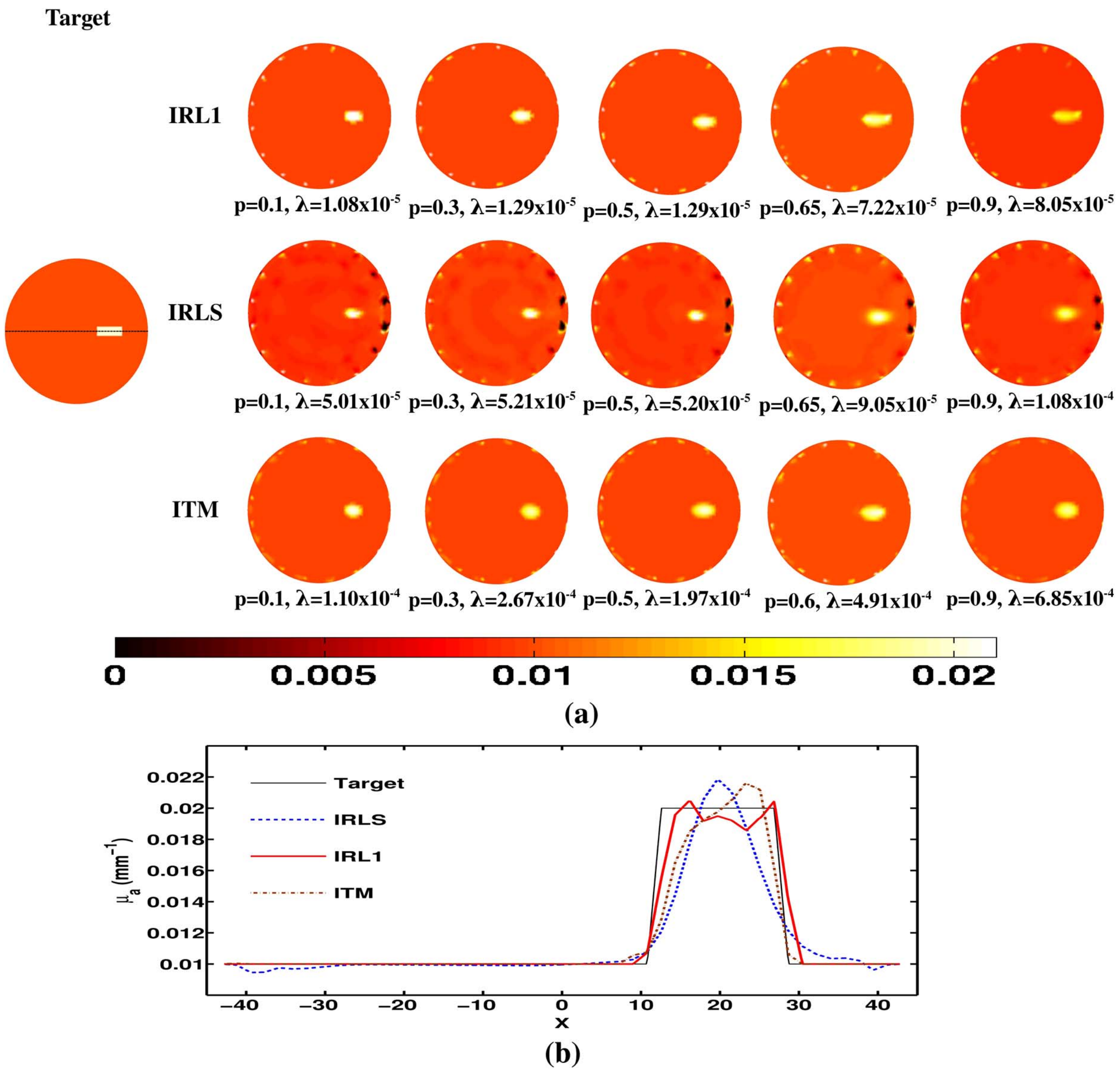


Fig. 3. (a) Reconstructed μ_a images using IRL1, IRLS, and ITM algorithms in the numerical experiment with 1% noisy data collected using the target distribution, which is rectangular in shape, placed horizontally close to the boundary of the imaging domain (given in the first column) for varying values of p (given below each reconstructed image). The first, second, and third rows show the reconstructed images obtained for the 1% noisy data using IRL1, IRLS, and ITM, respectively. The 1D cross-sectional profile along the dashed line of the target distribution for reconstructed μ_a images for the optimized p-value (last but one column) is given in (b).

along the dotted line given in the target image are presented in Fig. 3(b). The corresponding quantitative metrics are also reported in Tables 1 and 2.The different p-values considered for reconstruction are shown in the first, second, and third rows of Fig. 3(a). Note that the last but one column in Fig. 3 indicates the reconstruction obtained for optimized p-value. The reconstruction obtained from IRLS reveals that the target is shrunk compared to the results obtained from IRL1 and ITM, with IRL1 fairing marginally compared to ITM.

Finally, the reconstruction results obtained for the case of experimental gelatin phantom data are given in Fig. 4. The corresponding quantitative metrics are given in Tables 1 and 2. These results indicate that the IRL1 is marginally superior in terms of tumor-shaped recovery visually compared to other methods, while ITM tries to suppress the background artifacts considerably compared to IRL1 and IRLS. Even though the PC values are slightly higher for the ITM, the quantitative accuracy (indicated by the mean value of the target) is higher for the IRL1. It is important to note that the experimental phantom considered here represents a layered tissue model with varying optical properties mimicking a typical breast, which indicates a less sparse solution, where one can expect the ITM to perform well.

To drive the point of existence of optimal p for the ℓ_p-norm deployment, the p $(0 < p \leq 1)$ was varied in steps of 0.05, and the corresponding data-model misfit function $(\|y - G(\mu_a)\|_2^2)$ was plotted for the IRLS, IRL1, and ITM for the target case of Fig. 1 with 5% noise in the data. The same is given in Fig. 5(a), which clearly shows that there exists a minima for such functional, and an optimal p can be easily determined. Note that this approach was adapted for all cases that were presented in this work. The $G(\mu_a)$ here represents the reconstructed μ_a obtained with corresponding p coupled with the optimal regularization parameter (λ) obtained using Algorithm 4.

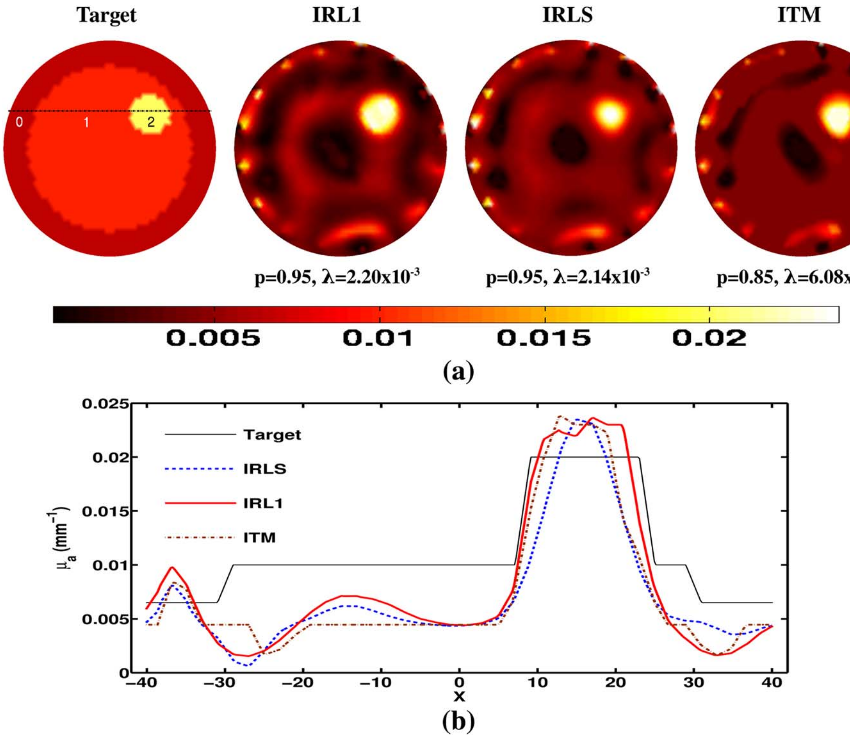


Fig. 4. (a) Comparison of reconstructed μ_a images obtained for the case of experimental gelatin phantom, mimicking a typical breast, using IRL1, IRLS, and ITM algorithms. (b) Similar to Fig. 2, the 1D cross-section of the reconstructed results along the dotted line of the target image.

For completeness, the corresponding mean squared error values for the reconstructed μ_a are plotted in Fig. 5(b).

6. DISCUSSION

The usage of ℓ_1-norm-based image reconstruction has been well studied in diffuse optical tomographic imaging [8–14,34], which is the tightest convex envelope for the NP hard ℓ_0-norm (ideal norm for recovering the sparsest solution). It has been shown in various contexts [16,17], that for this convex surrogate for a Gaussian sensing matrices, the minimum number of measurements (m) required is [16]

$$m > Ck\log^{\beta}(n), \tag{22}$$

where the value of β depends on the sensing matrix, C represents the coherence property of the sensing matrix, k is the sparsity of the solution, and n is the size of the unknown signal. The inverse problem encountered in diffuse optical tomography is typically underdetermined; as a result, one expects to have an image-reconstruction framework that can perform well even in the case of less data. To address this need, a class of nonconvex algorithms based on ℓ_p-minimization $(0 < p < 1)$ was proposed. The minimum number of measurements (m) required to solve the inverse problem with a nonconvex penalty is [16]

$$m \geq C_1 k + pk\,\log(n), \tag{23}$$

where C_1 is a constant that depends on the sensing matrix (Jacobian). Note that, as the value of p $(0 < p < 1)$ decreases, the second term in the above equation almost disappears, and the required number of measurements (m) only grows linearly with the sparsity level (k, number of nonzeros) in the solution. Hence nonconvex algorithms require lesser measurements

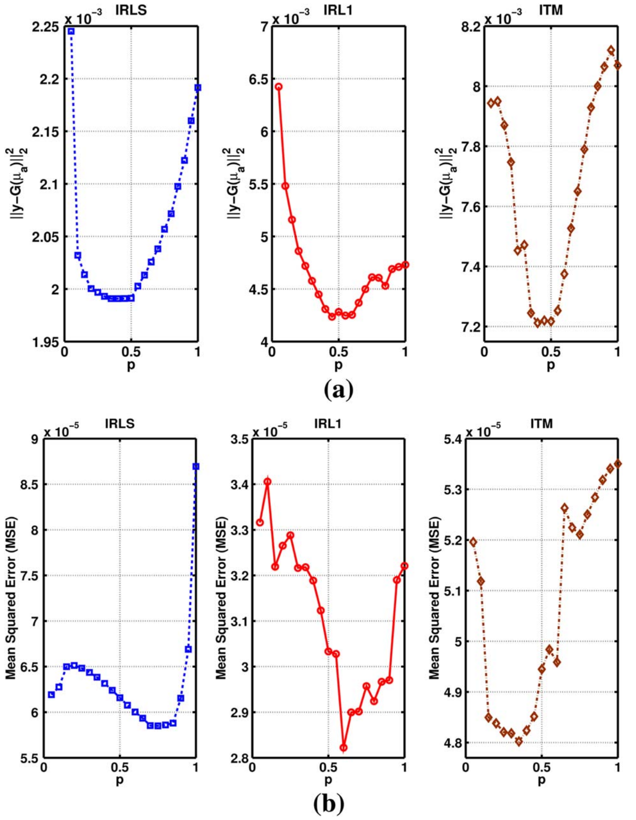


Fig. 5. (a) Plot showing the variation of p in the ℓ_p-norm $(0 < p \leq 1)$ in steps of 0.05 versus the data model misfit function, $\|y - G(\mu_a)\|_2^2$ for IRLS, IRL1, and ITM algorithms corresponding to the case of Fig. 1(a) (5% noise). The optimal value of p is the one that results in minimum value of the data model misfit (y axis). (b) Similar to (a), except the y axis gives the mean squared error, where the error is defined as reconstructed μ_a minus expected μ_a.

compared to the typical ℓ_1-norm [16]. As the inverse problem encountered in diffuse optical tomography is already severely underdetermined, we can expect the ℓ_p-norm to work well in these scenarios; the same has been shown in [8].

To effectively utilize the ℓ_p-norm, the determination of optimal regularization parameter is essential, and we have provided a scheme for determining the same. More importantly, we have developed a computationally efficient scheme for implementing the scheme, via modified SALSA (weighted SALSA). Note that Step 3 of Algorithm 1 (*soft* thresholding) is equivalent of the gradient of the ℓ_1-norm, and this approach is known as *analysis prior* (where the w vector appears only in the gradient term). On the other hand, if one has to multiply the weight matrix in Steps 1, 3, and 4 of Algorithm 1, the resulting problem is known as *synthesis prior*. It was shown that the *analysis prior* performs better than *synthesis prior* [16]; as a result in this work, only *analysis prior* is utilized in the IRL1 method (Algorithm 1).

Theoretically, a reweighted ℓ_2 (IRLS) minimization problem does not guarantee sparse solutions. The solution may not have zero entries at all, irrespective of the number of times the solution is reweighted. This effect is evident from the reconstruction results presented in this work (Figs. 1–4), where there are few background updates apart from the updates in the target. On the other hand, iterative reweighted ℓ_1 (IRL1) and iterative thresholding methods (ITM) guarantee sparse solutions at each iteration.

The IRL1 and IRLS algorithms require a sequence of positive numbers (ϵ_k) for all cases considered here. Our initialization for this vector was to make all entries equal to 0.1, and, after every iteration, the vector was multiplied by 0.5 to achieve a sequence that decays to zero. It was observed that choosing a value very close to zero leads to lot of boundary artifacts in the image. The stopping criterion (*Tol*) for IRL1, IRLS, and ITM (Algorithms 1–3) were set to 10^{-6} (single precision limit), and the adaptive regularization parameter estimation algorithm (Algorithm 4) to 10^{-5}.

The IRL1, IRLS, and ITM require $O(it \times N^3)$, $O(it \times N^3)$, and $O(it \times N^2)$ number of operations, respectively, for a fixed p and λ; here it represents the number of iterations. The ITM also requires computation of the largest singular value of Jacobian [Eq. (17)] for better implementation in real-time, making it also computationally on par with other methods. The typical computational time required for the IRL1, IRLS, and ITM algorithms for the results presented in Fig. 1(a) are 15.77, 12.03, and 13.81 s, respectively, for a fixed regularization parameter (λ) and value of p in the ℓ_p-norm.

The superiority of iteratively reweighted ℓ_1 has been demonstrated in numerous other cases [16,19,35,36] in terms of reconstructing a sparse solution compared to other existing methods; the same was also shown in this work. In terms of figures of merit, Tables 1 and 2, the IRL1 performance is comparatively superior to other methods that were discussed in this work. In terms of noise tolerance, the IRL1 also provides better robustness compared to other methods (Fig. 1, Tables 1 and 2, second row).

Moreover, this is the first comparison of ℓ_p-norm implementation schemes for biomedical optical image reconstruction, which includes modalities, namely, BLT, diffuse fluorescence optical tomography (DFT), and diffuse optical tomography. Even though the test cases presented here are limited in nature, the trends observed here should hold good in general.

7. CONCLUSIONS

The sparse recovery methods that utilize ℓ_p-norm $(0 < p < 1)$ in the regularization function have been shown to have a better utility compared to a strict convex regularization functional (ℓ_2 and ℓ_1-norm) in solving inverse problem in diffuse optical tomography. There are many ways of implementing the minimization of ℓ_p-norm cost function, and an effective comparison of these leading methods was performed in this work to know the utility of these implementations. The systematic comparison, which included three numerical and gelatin phantom cases, showed that these three typical approximation methods utilized in implementing ℓ_p-norm yield similar performance in the diffuse optical tomographic image reconstruction, with the iteratively reweighted ℓ_1-norm method performing better than others to some extent in the limited cases considered here. The algorithms used in this work are provided as open-source for enthusiastic readers [30].

ACKNOWLEDGMENTS

This work was supported by the Department of Atomic Energy, Government of India, through the DAE Young Scientist Research Award (sanction no. 2010/20/34/6/BRNS). This work in part was also supported by the Department of Biotechnology (DBT) Rapid Grant for Young Investigator (RGYI) (no. BT/PR6494/GDB/27/415/2012).

REFERENCES

1. D. A. Boas, D. H. Brooks, E. L. Miller, C. A. DiMarzio, M. Kilmer, R. J. Gaudette, and Q. Zhang, "Imaging the body with diffuse optical tomography," IEEE Signal Process. Mag. **18**(6), 57–75 (2001).
2. S. Srinivasan, B. W. Pogue, S. Jiang, H. Dehghani, C. Kogel, S. Soho, J. J. Gibson, T. D. Tosteson, S. P. Poplack, and K. D. Paulsen, "Interpreting hemoglobin and water concentration, oxygen saturation and scattering measured in vivo by near-infrared breast tomography," Proc. Natl. Acad. Sci. USA **100**, 12349–12354 (2003).
3. J. C. Hebden, A. Gibson, R. M. Yusof, N. Everdell, E. M. C. Hillman, D. T. Delpy, S. R. Arridge, T. Austin, J. H. Meek, and J. S. Wyatt, "Three-dimensional optical tomography of the premature infant brain," Phys. Med. Biol. **47**, 4155–4166 (2002).
4. S. R. Arridge and J. C. Schotland, "Optical tomography: forward and inverse problems," Inverse Probl. **25**, 123010 (2009).
5. A. H. Hielscher, A. D. Klose, and K. M. Hanson, "Gradient-based iterative image reconstruction scheme for time-resolved optical tomography," IEEE Trans. Med. Imaging **18**, 262–271 (1999).
6. P. K. Yalavarthy, B. W. Pogue, H. Dehghani, C. M. Carpenter, S. Jiang, and K. D. Paulsen, "Structural information within regularization matrices improves near infrared diffuse optical tomography," Opt. Express **15**, 8043–8058 (2007).
7. R. P. K. Jagannath and P. K. Yalavarthy, "Non-quadratic penalization improves near infrared diffuse optical tomography," J. Opt. Soc. Am. A **30**, 1516–1523 (2013).
8. J. Prakash, C. B. Shaw, R. Manjappa, R. Kanhirodan, and P. K. Yalavarthy, "Sparse recovery methods hold promise for diffuse optical tomographic image reconstruction," IEEE J. Sel. Top. Quantum Electron. **20**, 6800609 (2014), special issue on Biophotonics.
9. C. B. Shaw and P. K. Yalavarthy, "Effective contrast recovery in rapid dynamic near-infrared diffuse optical tomography using ℓ_1-norm-based linear image reconstruction method," J. Biomed. Opt. **17**, 086009 (2012).

10. N. Cao, A. Nehorai, and M. Jacobs, "Image reconstruction for diffuse optical tomography using sparsity regularization and expectation-maximization algorithm," Opt. Express **15**, 13695–13708 (2007).
11. M. Suzen, A. Giannoula, and T. Durduran, "Compressed sensing in diffuse optical tomography," Opt. Express **18**, 23676–23690 (2010).
12. J. C. Baritaux, K. Hassler, M. Bucher, S. Sanyal, and M. Unser, "Sparsity-driven reconstruction for FDOT with anatomical priors," IEEE Trans. Med. Imaging **30**, 1143–1153 (2011).
13. O. Lee, J. M. Kim, Y. Bresler, and J. C. Ye, "Compressive diffuse optical tomography: noniterative exact reconstruction using joint sparsity," IEEE Trans. Med. Imaging **30**, 1129–1142 (2011).
14. V. C. Kavuri, Z. Lin, F. Tian, and H. Liu, "Sparsity enhanced spatial resolution and depth localization in diffuse optical tomography," Biomed. Opt. Express **3**, 943–957 (2012).
15. S. Okawa, Y. Hoshi, and Y. Yamada, "Improvement of image quality of time-domain diffuse optical tomography with ℓ_p sparsity regularization," Biomed. Opt. Express **2**, 3334–3348 (2011).
16. A. Majumdar and R. K. Ward, "On the choice of compressed sensing priors and sparsifying transforms for MR image reconstruction: an experimental study," Signal Process. Image Commun. **27**, 1035–1048 (2012).
17. R. Chartrand, "Exact reconstruction of sparse signals via nonconvex minimization," IEEE Signal Process. Lett. **14**, 707–710 (2007).
18. Q. Lyu, Z. Lin, Y. She, and C. Zhang, "A comparison of typical ℓ_p minimization algorithms," J. Neurocomput. **119**, 413–424 (2013).
19. E. J. Candes, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," J. Fourier Anal. Appl. **14**, 877–905 (2008).
20. J. Feng, C. Qin, K. Jia, D. Han, K. Liu, S. Zhu, X. Yang, and J. Tian, "An adaptive regularization parameter choice strategy for multispectral bioluminescence tomography," Med. Phys. **38**, 5933–5944 (2011).
21. H. Dehghani, M. E. Eames, P. K. Yalavarthy, S. C. Davis, S. Srinivasan, C. M. Carpenter, B. W. Pogue, and K. D. Paulsen, "Near infrared optical tomography using NIRFAST: algorithms for numerical model and image reconstruction algorithms," Commun. Numer. Methods Eng. **25**, 711–732 (2009).
22. S. R. Arridge and M. Schweiger, "Photon-measurement density functions. Part 2: finite-element-method calculations," Appl. Opt. **34**, 8026–8037 (1995).
23. P. L. Combettes and V. R. Wajs, "Signal recovery by proximal forward-backward splitting," Multi-scale Model. Simul. **4**, 1168–1200 (2005).
24. M. A. Figueiredo, R. D. Nowak, and S. J. Wright, "Gradient projection for sparse reconstruction: application to compressed sensing and other inverse problems," IEEE J. Sel. Top. Signal Process. **1**, 586–597 (2007).
25. M. Figueiredo, J. Bioucas-Dias, and M. Afonso, "Fast frame-based image deconvolution using variable splitting and constrained optimization," in *IEEE Worskhop on Statistical Signal Processing*, Cardiff, Wales (2009).
26. I. Selesnick, "Introduction to sparsity in signal processing [Connexions Web site]," available at http://cnx.org/content/m43545/1.3/. (2012).
27. M. Lai and J. Wang, "An unconstrained ℓ_q minimization with $0 < q < 1$ for sparse solution of under-determined linear systems," SIAM J. Optim. **21**, 82–101 (2011).
28. Y. She, "An iterative algorithm for fitting nonconvex penalized generalized linear models with grouped predictors," Comput. Stat. Data Anal. **56**, 2976–2990 (2012).
29. Z. Wang and J. Liu, "New model function methods for determining regularization parameters in linear inverse problems," Appl. Numer. Math. **59**, 2489–2506 (2009).
30. https://sites.google.com/site/sercmig/home/complpnorm.
31. J. Kuntz, B. Flach, R. Kueres, W. Semmler, M. Kachelrie, and S. Bartling, "Constrained reconstructions for 4D intervention guidance," Phys. Med. Biol. **58**, 3283–3300 (2013).
32. T. O. McBride, B. W. Pogue, S. Jiang, U. L. Osterberg, and K. D. Paulsen, "Development and calibration of a parallel modulated near- infrared tomography system for hemoglobin imaging in vivo," Rev. Sci. Instrum. **72**, 1817–1824 (2001).
33. B. W. Pogue and M. S. Patterson, "Review of tissue simulating phantoms for optical spectroscopy, imaging and dosimetry," J. Biomed. Opt. **11**, 041102 (2006).
34. C. B. Shaw and P. K. Yalavarthy, "Prior image-constrained ℓ_1-norm-based reconstruction method for effective usage of structural information in diffuse optical tomography," Opt. Lett. **37**, 4353–4355 (2012).
35. J. Meng, L. V. Wang, L. Ying, D. Liang, and L. Song, "Compressed-sensing photoacoustic computed tomography in vivo with partially known support," Opt. Express **20**, 16510–16523 (2012).
36. N. Vaswani and W. Lu, "Modified-CS: modifying compressive sensing for problems with partially known support," IEEE Trans. Signal Process. **58**, 4595–4607 (2010).