An Efficient Method for Model Reduction in Diffuse Optical Tomography

A.-R. Zirak*, **, M. Khademi* and M.-S. Mahloji**

Abstract: We present an efficient method for the reduction of model equations in the linearized diffuse optical tomography (DOT) problem. We first implement the maximum a posteriori (MAP) estimator and Tikhonov regularization, which are based on applying preconditioners to linear perturbation equations. For model reduction, the precondition is split into two parts: the principal components are considered as reduced size preconditioners applied to linear perturbation equations while the less important components are marginalized as noise. Simulation results illustrate that the new proposed method improves the image reconstruction performance and localizes the abnormal section well with a better computational efficiency.

Keywords: DOT, Bayesian Methods, Model Reduction, Preconditioner.

1 Introduction

Diffuse optical tomography (DOT) involves the technique of using Near Infrared light for imaging specific parts of the body by diffusive nature of photons in turbid media. It has the advantages of good temporal resolution an order of magnitude faster than functional MRI, which is also portable, low-cost, non-invasive and non-ionizing. Furthermore, it offers unique physiological information about the metabolic status that is unavailable from other imaging methods [1,2]. It is also useful for obtaining information about tissue abnormalities, such as breast and brain tumors. For these reasons, DOT is becoming a useful complement to the current tomographic modalities.

Despite the unique feature it possesses, the extracting of optical properties or image reconstruction in DOT is a nonlinear, ill-posed problem which usually suffers from a low spatial resolution. To remedy these, various approaches including regularization and Bayesian methods with spatial prior have drawn a significant attention. These approaches have led to the development of many enhanced diffuse optical imaging systems such as MR-guided optical breast imaging [2] and the X-ray-guided DOT systems [3]. Owing to sparse and correlated measurements data, Bayesian methods are suitable. In such problems, which are solved by iterative methods, for stabilization and better convergence, the solution space must be small. These constraints lead to extensive and over determined system of equations so that model reduction to efficiently minimize computation and model retrieving criteria to refine model error are often required.

In the literature, some methods for model reduction are proposed. Popular methods are bound to tessellate the domain for the forward problem into sparse or coarse meshes or truncate the model, which has some disadvantages such as reduction of image quality. In [4] a model reduction method with better performance is proposed. It represents computational model inaccuracy as a random variable or noise. But in this method, the non-linearity nature of the optimization problem has some disadvantages such as restriction of applying some criteria or constraints, e.g. regularized total least squares (RTLS) and principal component analysis (PCA), both of which are used in this paper. Also, lack of prior information in model reduction has some disadvantage such as blind and case sensitive model reduction.

An efficient framework for model reduction in DOT is presented in this paper. For this reason, the principal components of covariance matrices are considered as the preconditioner matrices. However, the less important ones are marginalized as noise. This type of preconditioners, have properties such as maximum a posteriori (MAP) estimator and Tikhonov regularization, which simplify applying proposed methods without changing the linearity of the problem. In this paper, model reduction and refinement is performed using PCA method and RTLS criteria respectively, for treating the linearized preconditioned DOT problem. All of the proposed methods in this

Iranian Journal of Electrical & Electronic Engineering, 2008.
Paper first received 14th May 2007 and in revised form 15th March 2008.
* A.-R. Zirak and M. Khademi are with the Department of Electrical Engineering, Ferdowsi University of Mashhad, Mashhad, Iran.
E-mail: zirak@um.ac.ir.
** A.-R. Zirak and M.-S. Mahloji are with the Laser & Optics Research School, Nuclear Science & Technology Research Institute (NSTRI), P.O.Box 14155-1339, Tehran, Iran.

paper are concentrated in one algorithm which is named PCAME (Principal Component analysis with marginalization errors). In this paper, we have compared proposed PCAME algorithm with traditional model reduction methods.

This paper is organized as follows: In Section 2, some useful theories in DOT are discussed. In Section 3, numerical methods for solving linear equations which we are concerned with are recalled. Section 4 introduces a method for reduction of the model size which is named PCAME algorithm. Simulation results illustrating the performance of the proposed methods in DOT problems are presented in Section 5. Section 6 contains the conclusions.

2 Diffuse Optical Tomography

2.1 Diffusion Approximation

Light photons undergo absorption and scattering processes when passed through domains such as tissue and the diffusion equation approximates the bulk light propagation under the assumption that the diffuse fluency behaves as though the scattering is uniformly isotropic with a reduced scattering coefficient, μ_s', when measured over long distances [5,6]. This condition exists under the assumption that scattering dominates over absorption which is true in the case of several tissue types, including the human breast, in the wavelength region of 650-1350 nm [2]. Let $\Omega \subset R^P (p = 2,3)$ denote the domain under test. In the frequency-domain, this differential equation can be described accurately by the Diffusion Approximation (DA) which is expressed by [2,7]:

$$-\nabla.\left(\kappa(r)\nabla\Phi(r,\omega)\right)+\left(\mu_a(r)+\frac{i\omega}{c}\right)\Phi(r,\omega) = q_0(r,\omega), \qquad r \in \Omega \tag{1}$$

where $\Phi(r,\omega)$ is the photon density at position *r* and modulation frequency ω (in this work, ω = 100MHz), $\kappa(r)=1/3(\mu_a+\mu_s')$ denotes the spatially varying diffusion coefficient with absorption coefficient μ_a and reduced scattering coefficient μ_s', c is the speed of light in the medium, p=2,3 is the dimension of the domain, $q_0(r,\omega)$ is the isotropic source term and $i=\sqrt{-1}$. In the DA, it is assumed that the angular distribution of the radiance is almost uniform. This assumption can be achieved within highly scattering medium ($\mu_s' >> \mu_a$) relatively far (a few $1/\mu_s'$) from the sources [2,8].

DA can be solved using standard numerical techniques, and here a finite element model (FEM) for this equation, has been applied. After solving this equation, one can obtain the fluency for a given distribution of optical properties by applying suitable boundary conditions. In this paper, a Robin (*Type III*) boundary condition is used, which is given as follows [2,9]:

$$\Phi(r,\omega)+\frac{1}{2\gamma}\kappa(r)\frac{\partial\Phi(r,\omega)}{\partial\nu} = g(r,\omega) \qquad r \in \partial\Omega, \tag{2}$$

where $g(r,\omega)$ models the boundary sources, γ is a dimension-dependent constant ($\gamma_2 = 1/\pi, \gamma_3 = 1/4$) and ν is the outer normal at boundary domain, $\partial\Omega$.

2.2 Image Reconstruction

Image reconstruction methods in DOT differ in the type of data being considered, the type of solutions being sought, the physical model assumed for light propagation, and in the algorithmic details [10,11], which is always done by sequential solving of two problems, forward and inverse problems. Roughly speaking, the forward problem in DOT describes the photon propagation in tissue and is usually formulated by the DA equation. The inverse problem consists in retrieving the spatially varying image of the object by comparing predicted and actual measurements. In this paper, the forward problem is solved with Finite Element based methods (FEM). The key principle in the FEM is the reduction of the general, continuous problem to one of matrix algebra of finite size [12]. Each node in the FEM mesh is labeled according to the region obtained from simulated MRI images [13]. In the inverse problem, the actual and predicted measurements are often used by optimization based procedure to get the retrieved image.

The inverse problem solving methods are divided into two classes, which are the linear methods based on perturbation theory and the non-linear methods based on optimization approach [10]. Two main distinctions exist between linear and non-linear methods. In the former category, a perturbation model is postulated which corresponds to the first term in a Born or Rytov expansion of the Lippman–Schwinger equation [13]. We have employed the perturbation approach with a first-order Rytov approximation in the frequency domain [7,11]. The second approach considers the model in terms of explicit parameters and adjusts these parameters in order to optimize an objective function combining a data fitting and regularization term. The nonlinear minimization problem is usually solved iteratively using gradient methods such as the nonlinear conjugate gradient [14,15] or Newton based methods [11] to minimize the objective function over the search space of optical parameters. Newton based methods converges faster owing to the contribution of higher order information. However, the price paid is the reduction in robustness, i.e. it is more sensitive to poor initial estimates than the conjugate gradient (CG) method. To provide a significant computation time and error saving of Hessian matrix (second derivative of forward model) which is used in Newton based

methods, a Gauss–Newton approach [11] to the inverse solver in optical tomography is used.

It is shown that [10], the linear methods has a good performance such as robustness for poor initial estimates in the first stage of the iterative process. But in the final stages of the iterations, where the estimate of optical property is closed to the actual ones, Gauss–Newton method is very good because of its quadratic convergence and low oscillation during convergence [11]. For this reasons, we have used the linear perturbation method at the first and the non-linear Gauss–Newton method at the final stage. In order to focus on numerical aspects of perturbation and Gauss-Newton equations, they can be represented in the matrix form by concerning with the solution of large linear equation systems contaminated by additive Gaussian noise:

$$b = Ax + e, \quad A \in R^{m \times n}, x \in R^{n}, b, e \in R^{m}, \tag{3}$$

where e is the additive noise vector, x is the image perturbation and b and A are named the measurement vector and model matrix respectively. The model matrix A includes *Jacobian* structural elements [10]. The measurements in equation (3) would be complex, and inversion would lead to a complex parameter update. For this reason we split the measurement vector into real and imaginary parts with a commensurate splitting of the linearized derivative operators. In addition, when considering log of the data the splitting associates the real part with logarithmic amplitude, and the imaginary part with phase [11]. In DOT the actual measurement of each photon is most often corrupted by shot noise statistics, which originates from Poisson statistics. However, with a sufficiently large number of detected photons, as the intensity tends to infinity, the probability distribution of the shot noise is normally considered in the Gaussian limit [7]. So, e in (3) is assumed to have Gaussian distribution, where its variance is expected to be proportional to the number of photons at the detector and spatially uncorrelated for simplicity.

In particular, we are concerned about the solution of linear discrete ill-posed problems for which the corresponding model matrices A are of ill-determined rank, contaminated by noise, and as large as to make its factorization impossible and not explicitly available. Besides, the model matrix A is very extensive and large so that model reduction to efficiently minimize computation is often required.

3 Numerical Methods

3.1 Regularization

When the linear equation (3) comes from the discretization of a diffusion approximation, many of the singular values of the coefficient matrix *A* are very close to the origin, so the ill-conditioning of the coefficient matrix for these linear systems is typically very large [17]. Quite often, one of the following results can occur in the evaluation of the goal of the inverse problem: 1) the solution does not exist, 2) the solution is not unique, 3) the solution is not converge with acceptable computational time or 4) solving the solution is not stable, i.e. a tiny perturbation (error) in b (measurements) will be amplified so that it results in a large perturbation in x (image). If any of the above results occurs, the inverse problem is said to be ill-posed (in the Hadamard sense) [18], sometimes so much as to make the computed image useless. An ill-posed problem must be converted into a well-posed version in order to be solved. The technique for this conversion is called regularization.

We extend the discussion into three useful and popular regularization methods for the conversion of linear systems to have a better behavior and solutions.

The first method for regularizing is preconditioning. In traditional iterative reconstruction methods, preconditioning is a technique which improves the convergence rate by transforming the matrix A into a new matrix with more desirable spectral characteristics [16]. With this perspective, the equation

$$\left(L_1 A L_2^{-1}\right)\left(L_2 x\right) = L_1 b \tag{4}$$

is named the preconditioned system of equations (3), with the same solution. L_1 and L_2^{-1} are symmetric positive definite matrices which are named left and right preconditioners respectively, such that the condition number of the resulting matrix $L_1 A L_2^{-1}$, in general, will be smaller. Furthermore, the side of the preconditioner is very important. In the iterative solution of linear discrete ill-posed problems, a right preconditioner is closely related to available or inferred information about the solution, while a left preconditioner conveys information about the noise in the data or model matrix whose statistical properties may be known. A survey of popular preconditioning strategies can be found, e.g., in [16].

The second but most popular method for regularization in DOT is Tikhonov regularization which replaces the linear system (3) with the minimization problem

$$x_{\text{tikhonov}} = \arg\min_{x} \left\{ \|Ax - b\|^2 + \lambda \|L(x - x_0)\|^2 \right\} \tag{5}$$

where the second term in the functional to be minimized penalizes unexpected growth of the solution. The Matrix L decides how the computed solution should be allowed to grow and the regularization parameter λ balances the effect of the prior with predicted-actual measurements mismatch. Details of how to choose the optimal regularization parameters L and λ can be found in [11,20,21].

The third method for regularizing linear ill-posed problems with the application of DOT is TLS (total least squares) based methods. It can guarantee the existence of the solution by adapting the measurement space with the model space. The TLS method can produce a robust solution only in linear systems such as equation (3) when A and b are both contaminated with noise whereas other methods only consider the noise in b. The TLS estimation is obtained by [22]:

$$\min \left\| (E, f) \right\|_F \text{ subject to } (A + E)x = b + f \tag{6}$$

where both E and f must be minimized with the Frobenius norm $\| \ \|_F$. For more detail see [23]. Also, in practical situations, the linear system is often ill-conditioned. In these cases the TLS solution can be physically meaningless and thus regularization is essential for stabilizing the solution. For this reason, RTLS was addressed by several approaches such as truncation methods and Tikhonov regularization [18,22,23]. In this paper, RTLS is applied to all of the preconditioned linear equations as the minimization problem of [23]:

$$x_{RTLS} = \arg \min_{x \in R^n} \left\{ \frac{\|Ax - b\|^2}{\|x\|^2 + 1} : \|L_s x\|^2 \le \rho \right\}, \tag{7}$$

where $\rho > 0$ is a regularization parameter and $L_s \in R^{k \times n} (k \le n)$ is a regularization matrix that defines a (semi)norm on the solution which is frequently chosen to approximate the first or second derivative operator.

3.2 Dimension Reduction of Equations

Dimension reduction is often required in several applications, typically due to limited available time, computer memory or other restrictions. In problems that are related to partial differential equations, this often means that we are bound to use sparse or coarse meshes in the model for the forward problem, to reduce the dimension of equation (3). Typically, the distributed parameter x is approximated by a representation in lower dimension basis such that the linear equation (3) is replaced by an reduced size equation,

$$\begin{aligned} b &= A_h x_h + e, \\ A_h &\in R^{m \times n_1}, \ x \in R^{n_1}, \ b, e \in R^m, \end{aligned} \tag{8}$$

Where n_1 is the degrees of the approximation of x ($n_1 \le n$), and $h>0$ is a parameter controlling the level of discretization. Conversely, if we are given more and more accurate measurements, we have to employ increasingly accurate forward problem solvers in order to exploit the information in the measurements. In DOT, the typical required accuracy for the forward problem solver leads to computational times that may be unacceptable both in biomedical and industrial end applications. On the other hand, if in (8) the forward model is inaccurate; the discretization error may become significant compared to the measurement error. Together with the fact that the inverse problem is ill-posed, the approximation error may destroy the quality of the estimate of x. This dichotomy is one of the bottlenecks of diffuse tomographic methods, where the computational complexity is an issue even when relatively coarse meshes are used.

To overcome this dichotomy, in [4], statistical inversion theory is employed such that represents computational model inaccuracy as a random variable and thus treat it as noise. Hence, instead of the model (8), an *accurate* model can be written as,

$$\begin{aligned} b &= A_h x_h + \left[Ax - A_h x_h \right] + e \\ &= A_h x_h + \varepsilon(x) + e, \end{aligned} \tag{9}$$

where the term $\varepsilon(x)$ is the modeling error. But in this method, the non-linearity nature of objective function has some disadvantages such as computational complexity and restriction of applying some criteria or constraints, e.g. regularized total least squares (RTLS). Another drawback of this method is blind reduction, for example, restriction of using principal components of priors to reduce the model size. This will fade small but important parts of the image, such as anomalies. Also lack of prior information in model reduction has some disadvantage such as space invariant model reduction.

4 Implementation

4.1 Statistics and Regularization as Preconditioners

It is well-known that traditional DOT reconstruction algorithms do not produce satisfactory reconstructions when applied to sparse projection data or the poor model matrix. Besides, the optical property distributions within the studied domains usually contain distinct jumps at the boundaries of different organs. Despite this fact, the regularization schemes that are most commonly utilized in DOT carry the implicit assumptions that the optical property distribution is smooth (in the sense of Sobolev norms) and as a consequence the reconstructed images are often too smooth and the organ boundaries and anomalies are blurred. In recent years, many papers present and review results suggesting that statistical inversion methods can be successfully used for reconstruction in DOT. The statistical inversion approach has the following benefits:

- Any collection or less informative projection data can be used for tomographic reconstruction. In particular, scattering geometry and truncated

projections are not more complicated to work with than non-scattering geometry and full projections.

- Application-dependent a priori information on the target can be used in a natural and systematic way to recast the classically ill-posed problem in a well-posed stochastic form. With a well constructed prior model one can obtain better improved image quality than in traditional methods.

Although many statistical methods have been proposed for the restoration of tomographic images, their use in medical environments, especially in functional imaging, has been limited due to two important factors. These factors are the need for greater computational time than deterministic methods and the selection of the hyperparameters in the image models. In this section, a statistically based preconditioning framework which is basically used in [19] for numerical procedure, is introduced and improved in DOT based systems. From the point of view of regularization, the traditional preconditioners which only improve the speed of convergence seem to be of little use in DOT. Indeed, accelerated convergence by preconditioning may lead to an iterative method where the noise takes over immediately and the regularization property is lost. In this work, preconditioning is investigated from the point of view of Bayesian statistics. In this case, we consider the compatibility of prior assumptions of regularization scheme and the actual prior information on optical property distribution from a Bayesian perspective. In this paper, random variables are denoted by capital letters and their realizations are denoted by lowercase letters. In the Bayesian approach, the optical property distribution X and the measurements Y are assumed to be multivariate random variables with some joint probability density $p(x,y)$. Hence, instead of the deterministic equation (3), we consider its stochastic extension,

$$\begin{aligned} &AX = B + E, \\ &A \in R^{m\times n},\ X \in R^{n},\ B,E \in R^{m}, \end{aligned} \tag{10}$$

where X, B and E are random variables instead of x, b and e in deterministic model respectively, and A is the deterministic model matrix. Most papers use the additive noise model E, but the ideas of this paper are more generally applicable, so that the model error in A can be added to E with good approximation. Let $p(x)$ denotes the prior probability density of image X, which expresses the degree of information about the values of X prior to measuring B. The likelihood density, denoted by $p(b\,|\,x)$, is the probability density of B given the realization X=x. The probability density of X given B=b is called the posterior density and is denoted by $p(x\,|\,b)$,

$$p(x\,|\,b) = \frac{p(x)p(b\,|\,x)}{p(b)}. \tag{11}$$

This is the solution of equation (10) in the Bayesian frame of mind. Based on the posterior density, we may define various estimates of the image x. The most commonly used statistical estimates are the *conditional mean* (CM) and *maximum a posteriori* (MAP) estimates,

$$\begin{aligned} x_{MAP} &= \arg\max_{x} p(x\,|\,y), \\ x_{CM} &= \int x\, p(x\,|\,y)dx, \end{aligned} \tag{12}$$

provided that such estimates exist.

Consider the linear additive noise model (10) under the assumption that X and E are mutually independent Gaussian random variables with $X \sim N(\mu_X, \Gamma_X)$ and $E \sim N(0,\Gamma_e)$, that is, the random variable X has mean $\mu_x \in R^n$ and its covariance matrix $\Gamma_x \in R^{n\times n}$ is symmetric, positive definite, and E is zero mean Gaussian noise with covariance $\Gamma_e \in R^m$. Then, Bayes' formula implies that the posterior density is

$$\begin{aligned} p(x\,|\,b) &\propto \exp\left(-\frac{1}{2}\left[(x-\mu_x)^T\Gamma_x^{-1}(x-\mu_x) + (b-Ax)^T\Gamma_e^{-1}(b-Ax)\right]\right) \\ &= \exp\left(-\frac{1}{2}\left[\,\|L_x(x-\mu_x)\|^2 + \|L_e(b-Ax)\|^2\,\right]\right) = \exp\left(-\Psi(x\,|\,b)\right), \end{aligned} \tag{13}$$

where the Cholesky factors of Γ_x^{-1} and Γ_e^{-1} are denoted by L_x and L_e, respectively, i.e., $L_x^T L_x = \Gamma_x^{-1}$, with L_x upper triangular. The value of x which maximizes (13) is the MAP estimate. Under these assumptions, x_{MAP} coincides with the conditional mean estimate x_{CM}, which is the centre point of the posterior density given above, and they are the solutions to the minimization problem,

$$x_{MAP} = x_{CM} = \arg\min(\Psi(x\,|\,b)). \tag{14}$$

In order to have a linear equation which its solution is equivalent to the non-linear MAP estimator, a new random variable is introduced:

$$W = L_x(X - \mu_x). \tag{15}$$

It is shown that W is a Gaussian white noise with $W \sim N(0, I)$, since

$$\begin{aligned} \mu_W &= E\{W\} = 0, \\ \Gamma_W &= E\{WW^T\} = L_x \Gamma_x L_x^T = I. \end{aligned} \tag{16}$$

where I is the identity matrix. Therefore, the upper triangular matrix L_x is a *whitening matrix* for X and owing to the above properties of W, the resulting equation has a better convergence property. Then, for a given realization of the image x, if we define $b_0 = b - A\mu_x$, we can write

$$\begin{aligned} \Psi(x \mid b) &= \|w\|^2 + \left\|L_e(b_0 - AL_x^{-1}w)\right\|^2 \\ &= \left\| \begin{bmatrix} L_e A L_x^{-1} \\ I \end{bmatrix} w - \begin{bmatrix} L_e b_0 \\ 0 \end{bmatrix} \right\|^2. \end{aligned} \tag{17}$$

Also, it is shown that the MAP estimator is the solution of this linear system in the least square sense, in the context of iterative solvers with appropriate regularization:

$$L_e A L_x^{-1} w = L_e b_0, \qquad w = L_x(x - \mu_x). \tag{18}$$

In this equation, L_x^{-1} and L_e are named right and left preconditioners respectively, or as is inferred from their nature, are named priorconditioners [19]. A central challenge in statistical modeling of inverse problems is constructing informative and reliable prior densities. For further details on constructing prior, see [4]. In this paper, we discuss sample-based estimation of the prior. Sample-based priors have been discussed in [19,25]. Another equally important question concerning priors is how to avoid typical priors that are biasing towards a reasonable, but incorrect, solution. In particular, in medical imaging, the prior should favor the typical or normal solutions that we expect to see, but at the same time it should allow the appearance of abnormalities or anomalies that are often of central interest. This issue has been previously addressed in [26]. Our approach to all of these issues is based on random sampling. In this method, it is assumed that a sample of realizations of the random variable X is accessible, as well as of the corresponding set of data. This sample of pairs are named a *training set* and denoted by

$$\begin{aligned} \vartheta_0 = \{(x_1, b_1), (x_2, b_2), ..., (x_N, b_N)\}, \\ x_j \in R^n, \quad b_j \in R^m, \end{aligned} \tag{19}$$

with $N \geq n$. If the prior is Gaussian or another standard parametric distribution, efficient random vector generators can be used. More generally, the sample can be generated by using Markov chain Monte Carlo (MCMC) techniques [25]. In DOT, the training set could consist, e.g., of previous measurements b_j combined with information obtained by surgical or experimental interventions, or it could have been generated, e.g., by using a computational anatomical or physiological model such as previous MRI or CT image. This set is used to set up a prior model that is adjusted to the estimation method of choice. Here, we assume that the vectors x_j are represented as discretized approximations. If N is large and the sample is representative, it is possible to estimate the probability density of the underlying variable X using this sample. For constructing priorconditioners, we must seek a Gaussian approximation of the prior density. The Gaussian distributions are completely characterized by the second-order statistics. Based on the available sample, the sample mean ($\tilde{\mu}_x$) and the sample covariance ($\tilde{\Gamma}_x$) which are estimation of real mean and covariance obtained as

$$\begin{aligned} \tilde{\mu}_x &= \frac{1}{N}\sum_{j=1}^{N} x_j \approx E\{X\} = \mu_x, \\ \tilde{\Gamma}_x &= \frac{1}{N}\sum_{j=1}^{N} x_j x_j^T - \tilde{\mu}_x \tilde{\mu}_x^T \approx E\{XX^T\} - \mu_x \mu_x^T = \Gamma_x \end{aligned} \tag{20}$$

4.2 Model Reduction by Error Marginalization

In medical applications, the vectors x_j represent typical features of the random variable X. For this reason, the vectors can not be very dissimilar. Consequently, the space spanned by the realizations may be a proper subspace even if $N \geq n$, and hence Γ_x is rank deficient or of ill-determined rank with the satisfactory approximation of its real value. Without loss of generality, also in linearized perturbation DOT problems, we may assume that the mean of X vanishes. Considering the singular value decomposition of the matrix Γ_x,

$$\begin{aligned} \Gamma_x &= VDV^T, \\ V &= [v_1, v_2, ..., v_n],\ D = \mathrm{diag}[d_1, d_2, ..., d_n], \end{aligned} \tag{21}$$

where the orthonormal singular vectors v_j correspond to the singular values d_j so that we have;

$$d_1 \geq d_2 \geq \ldots \geq d_r \geq \varepsilon > d_{r+1} \geq \ldots \geq d_n \geq 0, \tag{22}$$

where ε is a threshold value. Further, we split X in parts as

$$\begin{aligned} X = V_0(V_0^T X) + V_1(V_1^T X) = V_0 X_0 + V_1 X_1, \\ X_0 \in R^r,\ X_1 \in R^{n-r}. \end{aligned} \tag{23}$$

In the PCA method, after zero approximation of small singular values which provided that deleting the second term in equation (23), the original problem is transformed into a simplified linear equation that is well posed. This approximation which is particularly attractive for computations, leads to some disadvantages. Specially, if the forward model is inaccurate, the model reduction error may become significant compared to the measurement error. Together with the fact that the inverse problem is ill posed, the approximation error may destroy the quality of the image.

For treating errors in the forward model while reducing its size, we employ the Bayesian statistical inversion theory. The key idea in this paper is to represent not just the measurement noise, but also less important parts of model components as a random variable. The aim is to obtain an analogous formulation for the compensation by marginalization model. Hence, we don't ignore small singular values. For this reason, the second term in the right hand side of equation (23) must be considered as noise, so that:

$$\begin{aligned} B &= AX + E \\ &= A_0 X_0 + A_1 X_1 + E = A_0 X_0 + U, \end{aligned} \tag{24}$$

where:

$$A_0 = AV_0,\quad A_1 = AV_1,\quad U = A_1 X_1 + E. \tag{25}$$

We consider now the general case when X_0 and X_1 are not independent. We write equation (24) as

$$B = \begin{bmatrix} A_0 & I \end{bmatrix} \begin{bmatrix} X_0 \\ U \end{bmatrix}. \tag{26}$$

If we define

$$\Gamma_{ij} = E\{X_i X_j^T\},\qquad i,j = 0,1 \tag{27}$$

the joint covariance matrix becomes

$$\begin{aligned} \operatorname{cov}\left(\begin{bmatrix} X_0 \\ U \end{bmatrix}\right) &= E\left\{\begin{bmatrix} X_0 \\ U \end{bmatrix}\begin{bmatrix} X_0^T & U^T \end{bmatrix}\right\} \\ &= R = \begin{bmatrix} \Gamma_{00} & \Gamma_{01}A_1^T \\ A_1\Gamma_{10} & A_1\Gamma_{11}A_1^T + \Gamma_e \end{bmatrix}. \end{aligned} \tag{28}$$

A decomposition for R^{-1} can be written as

$$R^{-1} = L^T L,\qquad L = \begin{bmatrix} L_{11} & 0 \\ L_{21} & L_{22} \end{bmatrix}. \tag{29}$$

Now the multivariate Gaussian white noise can be defined as

$$\begin{bmatrix} W \\ V \end{bmatrix} = L\begin{bmatrix} X_0 \\ U \end{bmatrix}. \tag{30}$$

Writing the inverse of L in the form

$$L^{-1} = \begin{bmatrix} L_{11}^{-1} & 0 \\ -L_{22}^{-1}L_{21}L_{11}^{-1} & L_{22}^{-1} \end{bmatrix}, \tag{31}$$

and expressing (24) in terms of the newly defined white noise variables, we obtain

$$B = \begin{bmatrix} A_0 & I \end{bmatrix} L^{-1}\begin{bmatrix} W \\ V \end{bmatrix} = \left(A_0 - L_{22}^{-1}L_{21}\right)L_{11}^{-1}W + L_{22}^{-1}V, \tag{32}$$

which, by multiplication with L_{22} from the left, yields the whitened equation

$$\begin{aligned} L_{22}B = L_{22}\left(A_0 - L_{22}^{-1}L_{21}\right)L_{11}^{-1}W + V, \\ W = L_{11}X_0. \end{aligned} \tag{33}$$

Note that if X_0 and X_1 are mutually independent, the matrices R and L are block diagonal, and therefore (33) reduces to (18).

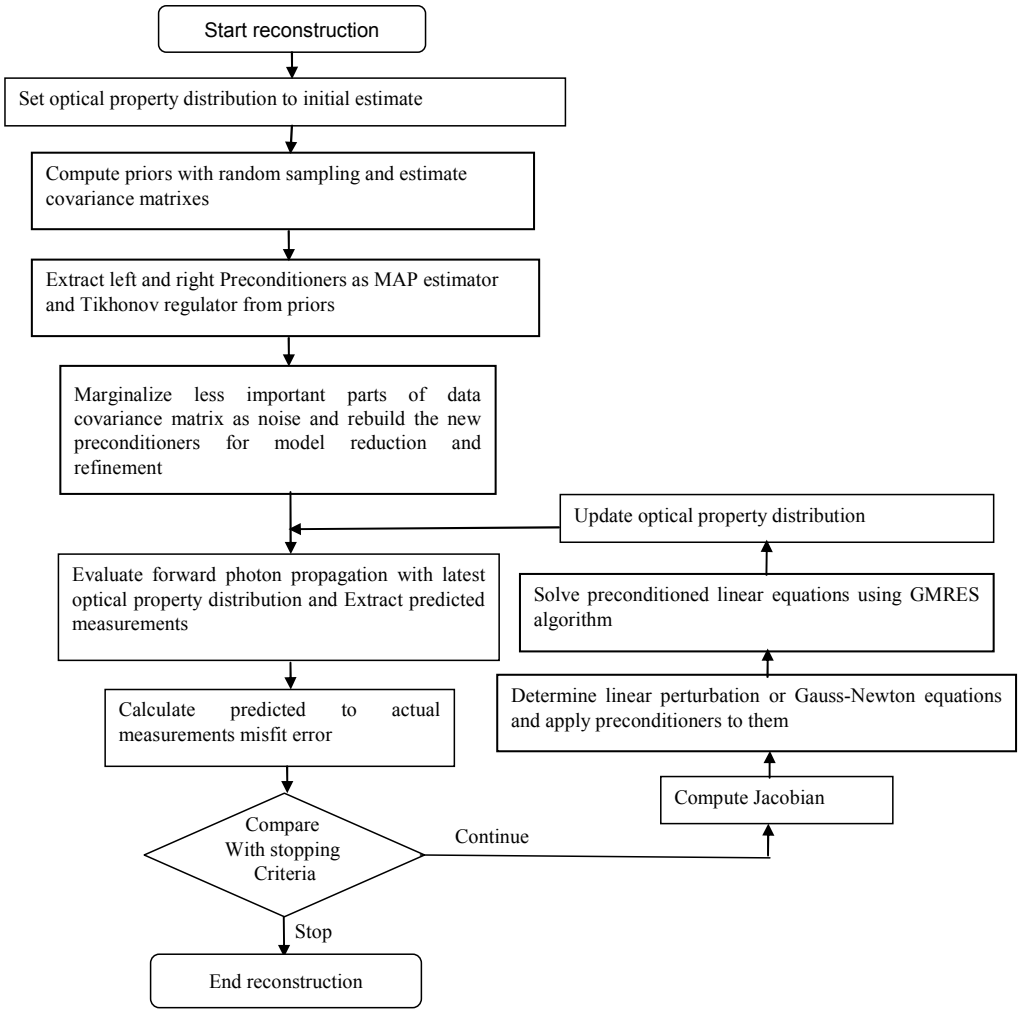


Fig. 1 Flow chart of the proposed model reduction method.

4.3 Flow Chart of PCAME Algorithm

Based on the proposed methods in Section 4, a flow chart which is named PCAME algorithm is drawn in figure (1) to illustrate the detailed computational steps for a complete reconstruction. From the flow chart, five proposed key steps can be identified: (a) computation of priors with random sampling, (b) extraction of preconditioners from priors or data covariance matrix, (c) Marginalize less important parts of priors as noise and rebuild the new preconditioners for model reduction and refinement, (d) updating the reconstruction equations by applying new preconditioners to them and (e) solving preconditioned linear equations using CGLS iterative method.

5 Numerical Results

5.1 Simulated Test Phantom

A multi-layered phantom is simulated in a circular object of diameter 100 mm and of infinite height. A two-dimensional cross-section of this phantom is shown in figure 2(a). In this figure, three regions are shown, Region 0 ($\mu_a = 0.015\,\mathrm{mm}^{-1}$ and $\mu_s' = 1.5\,\mathrm{mm}^{-1}$) and Region 1 ($\mu_a = 0.15\mathrm{mm}^{-1}$ and $\mu_s' = 15\,\mathrm{mm}^{-1}$) with typical prior information and finally Region 2 ($\mu_a = 0.01$ and $\mu_s' = 1\mathrm{mm}^{-1}$), which is simulated as a anomaly with few prior information.

The 16 optical channels are simulated, as is the standard practice in human imaging studies. They are used to collect data using 16 sources and 16 detectors (giving 240 measurements) in a single plane geometry so that there are one source and one detector together for each 16 optical channels. Sources are considered to be intensity modulated with a frequency of 100 MHz, and measurements consisting of the logarithmic modulation amplitude b^A and phase shift b^φ, so that the target data $b = \left(b^A, b^\varphi\right)$ is calculated by the FEM diffusion forward model. With this model, the object is discretized into 2880 non-overlapping triangular elements connected with 1501 nodes, defining a piecewise quadratic unstructured basis expansion. Both b^A and b^φ are then contaminated with 2% zero mean additive Gaussian noise, which is always considered in practical simulations. For *a priori* information, pixels of similar intensity as segmented MRI are assumed to represent the same material or tissue such as [12].

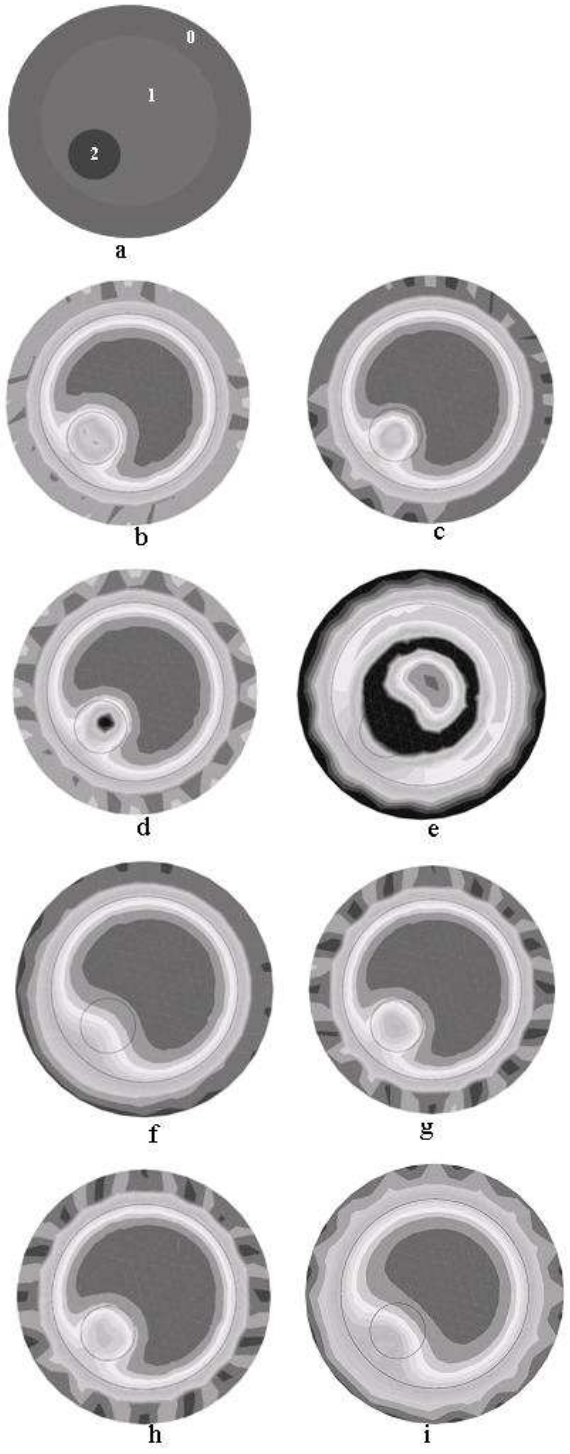


Fig. 2 (a) The simulated test phantom, (b) the reconstructed image with traditional MAP estimator, (c) the reconstructed image with preconditioning, (d) preconditioning and RTLS criteria, (e) preconditioning and traditional TSVD for dimension reduction, (f) preconditioning and dimension reduction with PCA, (g) preconditioning, dimension reduction with PCA and then RTLS criteria, (h) Proposed PCAME algorithm, (i) Compensation by marginalization with MAP estimator (with coarse meshes).

5.2 Phantom Imaging: Image Quality

Figure 2(b) gives the reconstructed image of μ_a when using the traditional MAP estimator with sample based image and noise covariance matrixes as in Section 4.1. By comparison, figure 2(c) gives the reconstructed image with the preconditioners which are introduced in equation (18). The preconditioners are constructed by assumption covariance matrices of noise, typical data and anomaly. These kinds of preconditioners with inherent regulator and MAP estimator properties lead to flexible and less computational operations with traditional iterative algorithms. The details of how RTLS criteria can improve the quality of image is shown in figure 2(d). Furthermore, the better results when using RTLS appears where applied to the reduced model by PCA method which approximates the subspace associated with the small singular values of preconditioners with zero which provided that deleting the second term in equation (23). It is shown by comparison of next three images. Figure 2(e) gives the reconstructed image when traditional TSVD (truncated singular value decomposition) model reduction is applied to preconditioned equation (18), without using RTLS [17]. In this method, the subspace spanned by small singular values is truncated such that the model dimension is reduced with well posed behavior. In this image, the regions are faded. Also 'bumps', 'valleys' and pseudo objects are observed. Figure 2(f) gives the reconstructed image by proposed PCA method for dimension reduction with zero approximation of small singular values of data covariance matrix (without error marginalization which is yielded with zero approximation of the second term in equation (23)). Comparison of figures 2(e) and 2(f) would demonstrate that dimension reduction with PCA method has the superior quality in anomaly detection than TSVD. The better results are available when RTLS is used to the previous PCA method, shown in figure 2(g). But when using the proposed PCAME algorithm which is based on the Compensation by marginalization method that is presented in (4.2), the best results especially for computational efficiency are observed. Figure 2(h) shows that the reconstructed image with PCAME algorithm is almost identical to figure 2(g). This is because equation (33) acts as a simple model refinement while model reduction with preconditioning perspective, which has a better computational efficiency. Also the image quality and computational efficiency of the proposed method is observed as compared with the method that is stated in [4]. In the latter, which its related image is shown in figure 2(i), some deterministic information is eliminated to reduce the model size, for example replacing fine meshes with coarse meshes, and the approximation error is compensated by marginalizing it as noise. This method for model reduction is somewhat computational expensive and faded, due to the nonlinear nature of Bayesian formula

and leaving out some important parts of data respectively.

5.3 Phantom Imaging: Computational Efficiency

Table 1 represents the reconstruction time and iteration numbers required for the iterative algorithm. Each rows of the table corresponds to one of the images from figure 2(b) to 2(i), respectively. In this paper, we have used CGLS iterative method on a 2.8 GHz Pentium 4 processor with 1 GB RAM. The performance of proposed model reduction method, which is marked with asterisk, can be inferred from this table.

Table 1 The convergence time and number of iterations to reconstruct the images shown in figure 2.

Method	Related figure	Converge time (sec)	Iteration number
Traditional MAP estimator	2(b)	6270	18
Preconditioning	2(c)	4976	19
Preconditioning with RTLS	2(d)	3012	14
Preconditioning with TSVD	2(e)	1647	12
Preconditioning with PCA	2(f)	1470	11
Preconditioning with PCA and RTLS	2(g)	1562	9
Proposed PCAME algorithm*	2(h)	931	10
Compensation by marginalization with coarse meshes	2(i)	1543	12

6 Conclusion

We used a preconditioning scheme for the linearized DOT problem which is constructed by covariance matrices. This help us to use the proposed model reduction method with advantages as constructing only important parts of the image or estimating only the required portion of the image in the field of view as the new preconditioners. Moreover, we treat the contribution coming from other parts, with less important properties, as noise. In this method, the approximated error in the model or covariance matrices is marginalized and evaluated as noise. For these reasons, the proposed method is named compensation by marginalization. We found that the preconditioning and model reduction by error marginalization is fairly robust to small absorption and scattering perturbation levels. Quantitatively, model reduction based on proposed PCAME algorithm is close to the expected results with respect to other modalities.

References

[1] Guy C. and Fytche D., *An Introduction to the Principles of Medical Imaging*, London, Imperial College Press, 2000.

[2] Arridge S., "Optical tomography in medical imaging," *Inverse Problems*, Vol. 15, pp. 41-93, Apr. 1999.

[3] Zirak A. R. and Mafinejad K., "Improving the iterative method in image reconstruction in CT-Scan," *8th ICEE conference*, Iran pp. 415-420, May 2000.

[4] Arridge S., Kaipio J. P., V. Kolehmainen, M. Schweiger, E. Somersalo,T. Tarvainen and M Vauhkonen, "Approximation errors and model reduction with an application in optical diffusion tomography," *Inverse Problems*, Vol. 22, pp. 175-195, Jan. 2006.

[5] Kolehmainen V., "Novel approaches to image reconstruction in diffusion tomography," Ph.D. thesis University of Kuopio, Kuopio, Finland, 2003.

[6] Selb J. and Joseph D., "Time-gated optical system for depth-resolved functional brain imaging," *Biomedical Optics*, Vol. 11, No. 4, Aug. 2006.

[7] Cong A. and Wang G., "A finite-element-based reconstruction method for 3D fluorescence tomography," *Optics Express*, Vol. 13, Issue 24, pp. 9847-9857, Nov. 2005.

[8] Hielscher A. H. and Yodh A. G, "Diffuse Optical Tomography of Highly Heterogeneous Media," *IEEE Transaction on Medical Imaging*, Vol. 20, No. 6, pp. 98-109, Jun. 2001.

[9] Cao N., "Tumor localization using diffuse optical tomography and linearly constrained minimum variance beamforming," *Optics Express*. Vol. 15, No. 3, pp. 896-909, Mar. 2007.

[10] Gibson A. P., Hebden J. C. and Arridge S. R., "Recent advances in diffuse optical imaging," *Phys. Med. Biol.* 50, pp. R1–R43, Feb. 2005.

[11] Schweiger M., Arridge S. R. and Nissil I., "Gauss–Newton method for image reconstruction in diffuse optical tomography," *Phys. Med. Biol*. 50, pp. 2365–2386, May. 2005.

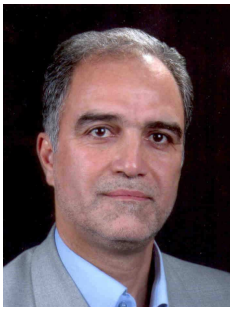
[12] Brooksby B., Dehghani H. and Pogue B.W., "Near infrared (NIR) tomography breast image reconstruction with a priori structural information from MRI: algorithm development for reconstructing heterogeneities," *IEEE J. Quantum Electron*. 9, pp. 199–209, Mar. 2003.

[13] Kilmer M., Miller E. and Barbaro A., "Three-dimensional shape-based imaging of absorption perturbation for diffuse optical tomography," *Appl. Opt*. 42, pp. 3129-3144, Jan. 2003.

[14] Arridge S. R. and Schweiger M., "A gradient-based optimisation scheme for optical tomography" *Opt. Express 2*, pp. 213-26, Mar. 1998.

[15] Hielscher A. H. and Klose A. D., "Gradient-Based Iterative Image Reconstruction Scheme for Time-Resolved Optical Tomography," *IEEE Trans. on Med. Imag*., Vol. 18, No. 3, pp. 32-46, Mar. 1999.
[16] Hanke M., Nagy J. G. and Plemmons R. J., "Preconditioned iterative regularization for ill-posed problems, in Numerical Linear Algebra and Scientific Computing," eds. L. Reichel, A. Ruttan and R.S. Varga, de Gruyter, Berlin, pp. 141-163, Jan. 1993.
[17] Hansen P. C., "Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion," *SIAM, Philadelphia*, 1998.
[18] Bellomo N., Bellouquid A., and Delitala M., "Mathematical topics on the modelling complex multicellular systems and immune competition," *Math. Models Meth. Appl. Sci*., 14, pp. 1683-1733, Nov. 2004.
[19] Calvetti D., Somersalo E., "Priorconditioners for linear systems," *Inverse Problems* 21, pp. 1379-1418, Jul. 2005.
[20] Anton H. and Rorres C., "Linear Algebra: Applications Version," 8th edition, New York: John Wiley and Sons, 2000.
[21] Brooks D. H., Ahmad G. F. and MacLeod R. S., "Inverse electrocardiography by simultaneous imposition of multiple constraints," *IEEE Trans. Biomed. Eng*. Vol. 46, pp.3-18, Jan. 1999.
[22] Renaut R. A. and Guo H., "Efficient algorithms for solution of regularized total least squares," *SIAM J. Matrix Anal. Appl*., Vol. 26(2), pp. 457-476, March 2005.
[23] Guo H. and Renaut R. A., "Iterative Method for Regularized Total Least Squares," Department of Mathematics and Statistics, Arizona State University, 2005.
[24] Calvetti D., Reichel L. and Shuibi A., "Invertible smoothing preconditioners for linear discrete ill-posed problems," *Appl. Numer. Math*., Vol. 54, pp. 135-149, Jul. 2005.
[25] Kaipio J. and Somersalo E., "Statistical and Computational Inverse Problems," *Springer-Verlag*, 2004.
[26] Huttunen T., Kaipio J., Hynynen K., "Modelling of anomalies due to hydrophones in continuous-wave ultrasound fields", *IEEE Trans UFFC* 50, pp. 1486-1500, Nov. 2003.

Ali-Reza Zirak was born in Iran in 1974. He received the B.Sc. degree in 1996, the M.sc. degree in 1999 and the Ph.D. degree in 2008 in electrical engineering, all from Ferdowsi University of Mashhad, Iran.
Since 2004, he has been with the Laser & Optics Research School, Nuclear Science & Technology Research Institute (NSTRI), P.O.Box 14155-1339, Tehran, Iran, where he is currently a Faculty Member in charge of advanced research and development activities in the fields of signal processing, electronic and communication system design and medical imaging. His research interests are in optimization, in particular, large-scale problems, statistical and digital signal and image processing, estimation theory and communication.



Morteza Khademi received the B.Sc. and the M.Sc. degrees in electrical engineering from Isfahan University of Technology in 1985 and 1987, respectively. He received the Ph.D. degree in 1995 in electrical engineering from Australia.
He is currently Associate Professor of Electrical Engineering at Ferdowsi University of Mashhad. His research interests are in video, audio and image coding, digital signal and image processing and error control coding.

Mohammad-Saeid Mahlooji was born in Iran in 1971. He received the B.Sc. degree in 1993 and the M.sc. degree in 1997 in electrical engineering, from Sharif University of Technology and Tabriz University, respectively.
Currently, he is with the Laser & Optics Research School, Nuclear Science & Technology Research Institute (NSTRI), P.O.Box 14155-1339, Tehran, Iran. He is a Faculty Member in charge of research and development activities in the fields of Laser developments and applications and electronic system design. His research interests are in power and laser based electronics.