

A gradient-based optimisation scheme for optical tomography

Simon R. Arridge

*Dept. of Computer Science, University College London, Gower Street, London
WC1E 6B, UK*

S.Arridge@cs.ucl.ac.uk

Martin Schweiger

Department of Medical Physics, University College London, WC1E 6JA, UK

martins@medphys.ucl.ac.uk

Abstract: Optical tomography schemes using non-linear optimisation are usually based on a Newton-like method involving the construction and inversion of a large Jacobian matrix. Although such matrices can be efficiently constructed using a reciprocity principle, their inversion is still computationally difficult. In this paper we demonstrate a simple means to obtain the gradient of the objective function directly, leading to straightforward application of gradient-based optimisation methods.

©1998 Optical Society of America

OCIS codes: (100.3190) Inverse problems; (110.6960) Tomography; (170.3010) Image reconstruction techniques; (170.3660) Light propagation in tissues; (170.6920) Time-resolved imaging; (170.6960) Tomography

References

1. A. D. Edwards, J. S. Wyatt, C. E. Richardson, D. T. Delpy, M. Cope and E. O. R. Reynolds, "Cotside measurement of cerebral blood flow in ill newborn infants by near infrared spectroscopy," *Lancet* **2**, 770-771 (1988).
2. J. S. Wyatt, M. Cope, D. T. Delpy, C. E. Richardson, A. D. Edwards, S. C. Wray and E. O. R. Reynolds, "Quantitation of cerebral blood volume in newborn infants by near infrared spectroscopy," *J. Appl. Physiol.* **68**, 1086-1091 (1990).
3. M. Tamura, "Multichannel near-infrared optical imaging of human brain activity," in *OSA Trends in Optics and Photonics on Advances in Optical Imaging and Photon Migration*, R. R. Alfano and J. G. Fujimoto, eds. (Optical Society of America, Washington, DC 1996) Vol. 2, pp. 8-10.
4. J. C. Hebden, R. A. Kruger and K. S. Wong, "Time resolved imaging through a highly scattering medium," *Appl. Opt.* **30**, 788-794 (1991).
5. *Near-infrared spectroscopy and imaging of living systems*, special issue of *Philos. Trans. R. Soc. London Ser. B*, Vol. 352 (1997).
6. J. C. Hebden, S. R. Arridge and D. T. Delpy, "Optical imaging in medicine: I. Experimental techniques," *Phys. Med. Biol.* **42**, 825-840 (1997).
7. S. R. Arridge and J. C. Hebden, "Optical imaging in medicine: II. Modelling and reconstruction," *Phys. Med. Biol.* **42**, 841-853 (1997).
8. S. B. Colak, G. W. Hooft, D. G. Papaioannou and M. B. van der Mark, "3D backprojection tomography for medical optical imaging," in *OSA Trends in Optics and Photonics on Advances in Optical Imaging and Photon Migration*, R. R. Alfano and J. G. Fujimoto, eds. (Optical Society of America, Washington, DC 1996) Vol. 2, pp. 294-298.
9. S. A. Walker, S. Fantini and E. Gratton, "Back-projection reconstructions of cylindrical inhomogeneities from frequency domain optical measurements in turbid media," in *OSA Trends in Optics and Photonics on Advances in Optical Imaging and Photon Migration*, R. R. Alfano and J. G. Fujimoto, eds. (Optical Society of America, Washington, DC 1996) Vol. 2, pp. 137-141.

10. M. A. O'Leary, D. A. Boas, B. Chance and A. G. Yodh, "Experimental images of heterogeneous turbid media by frequency-domain diffusing-photon tomography," *Opt. Lett.* **20**, 426-428 (1995).
11. S. R. Arridge, M. Schweiger and D. T. Delpy, "Iterative reconstruction of near-infrared absorption images," in *Inverse Problems in Scattering and Imaging*, M. A. Fiddy, ed., *Proc. SPIE* **1767**, 372-383 (1992).
12. H. Jiang, K. D. Paulsen, U. L. Osterberg, B. W. Pogue and M. S. Patterson, "Optical image reconstruction using frequency-domain data: Simulations and experiments," *J. Opt. Soc. Am. A* **13**, 253-266 (1995).
13. B. W. Pogue, M. S. Patterson, H. Jiang and K. D. Paulsen, "Initial assessment of a simple system for frequency domain diffuse optical tomography," *Phys. Med. Biol.* **40**, 1709-1729 (1995).
14. D. T. Delpy, M. Cope, P. van der Zee, S. R. Arridge, S. Wray and J. Wyatt, "Estimation of optical pathlength through tissue from direct time of flight measurement," *Phys. Med. Biol.* **33**, 1433-1442 (1988).
15. B. Chance, M. Maris, J. Sorge and M. Z. Zhang, "A phase modulation system for dual wavelength difference spectroscopy of haemoglobin deoxygenation in tissue," in *Time-resolved Laser Spectroscopy in Biochemistry II*, J. R. Lakowicz, ed., *Proc. SPIE* **1204**, 481-491 (1990).
16. J. D. Moulton, *Diffusion modelling of picosecond laser pulse propagation in turbid media*, M. Eng. thesis, McMaster University, Hamilton, Ontario (1990).
17. S. R. Arridge, M. Schweiger, M. Hiraoka and D. T. Delpy, "A finite element approach for modelling photon transport in tissue," *Med. Phys.* **20**, 299-309 (1993).
18. M. Schweiger, S. R. Arridge, M. Hiraoka and D. T. Delpy, "The finite element model for the propagation of light in scattering media: Boundary and source conditions," *Med. Phys.* **22**, 1779-1792 (1995).
19. S. R. Arridge, "Photon measurement density functions. Part 1: Analytic forms," *Appl. Opt.* **34**, 7395-7409 (1995).
20. S. R. Arridge and M. Schweiger, "Photon measurement density functions. Part 2: Finite element calculations," *Appl. Opt.* **34**, 8026-8037 (1995).
21. M. S. Bazaraa, H. D. Sherali and C. M. Shetty, *Nonlinear Programming: Theory and Algorithms*, 2nd edition (Wiley, New York, 1993).
22. S. R. Arridge and M. Schweiger, "Direct calculation of the moments of the distribution of photon time of flight in tissues with a finite-element method," *Appl. Opt.* **34**, 2683-2687 (1995).
23. M. Schweiger and S. R. Arridge, "Direct calculation of the Laplace transform of the distribution of photon time of flight in tissue with a finite-element method," *Appl. Opt.* **36**, 9042-9049 (1997).
24. M. Schweiger, S. R. Arridge and D. T. Delpy, "Application of the finite-element method for the forward and inverse models in optical tomography," *J. Math. Imag. Vision* **3**, 263-283 (1993).
25. K. D. Paulsen and H. Jiang, "Spatially-varying optical property reconstruction using a finite element diffusion equation approximation," *Med. Phys.* **22**, 691-701 (1995).
26. S. S. Saquib, K. M. Hanson and G. S. Cunningham, "Model-based image reconstruction from time-resolved diffusion data," in *Medical Imaging: Image Processing*, K. M. Hanson, ed., *Proc SPIE* **3034**, 369-380 (1997).
27. O. Dorn, *Das inverse Transportproblem in der Lasertomographie*, Ph. D. thesis, University of Münster, 1997.
28. R. Roy, *Image reconstruction from light measurements on biological tissue*, Ph. D. thesis, University of Hertfordshire, 1997.
29. S. R. Arridge and M. Schweiger, "A general framework for iterative reconstruction algorithms in optical tomography, using a finite element method," in *Computational Radiology and Imaging: Therapy and Diagnosis* C. Borgers and F. Natterer, eds., IMA Volumes in Mathematics and its Applications (Springer 1998, in press).
30. S. R. Arridge, M. Hiraoka and M. Schweiger, "Statistical basis for the determination of optical pathlength in tissue," *Phys. Med. Biol.* **40**, 1539-1558 (1995).
31. M. Schweiger and S. R. Arridge, "Optimal data types in optical tomography," in *Information Processing in Medical Imaging*, Lect. Notes Comput. Sci., **1230**, 71-84 (1997).

1. Introduction

Optical tomography is potentially a powerful tool to noninvasively obtain spatially resolved data of the optical parameters of tissue, from which physiologically relevant information such as local oxygenation can be calculated. Primary applications of this new imaging modality are the monitoring of cerebral blood and tissue oxygenation of newborn and preterm infants [1,2] to prevent death or permanent brain damage caused by asphyxiation during birth, functional mapping of brain activation during physical or mental exercise [3], and imaging of the breast to detect tumours [4]. Recent developments can be found in several review articles in the recent Royal Society meeting [5], and other journals [6,7].

The principal problem that arises in optical tomography is the dominance of scattering, which causes light to propagate diffusely in tissue and thus prohibits the application of direct reconstruction methods using the Radon transform. Various reconstruction methods, including *ad hoc* backprojection [8,9] and semi-analytic [10] schemes have been proposed, but increasingly attention is turning to iterative, optimisation based reconstruction methods [11-13].

In this approach, the problem is regarded as the optimisation of an objective function representing the sum-squared difference of the data to the model, plus additional regularisation terms representing prior knowledge. The most prevalent approach to date has been a Newton-like method such as Levenberg-Marquardt which involves repeated construction and inversion of the problem Jacobian. This method rapidly becomes intractable as the size of the problem domain increases, for example when a fully 3D problem is considered. An alternative is to consider gradient-based algorithms, such as conjugate gradients (CG). In this paper we present a method to derive the gradient efficiently using an adjoint source scheme.

2. Problem definition

Let Ω be the domain under consideration, with surface $\partial\Omega$. We consider S source positions $\vec{\zeta}_j \in \partial\Omega$ ($j = 1 \dots S$) and M_j measurement positions $\vec{\xi}_{j,i} \in \partial\Omega$ for each source j ($i = 1 \dots M_j$), resulting in a total number of measurements $M_{\text{TOT}} = \sum_{j=1}^S M_j$, with $M_{\text{UNIQ}} \leq M_{\text{TOT}}$ distinct measurement positions. The forward problem is non-linear and is represented by:

$$\vec{\mathbf{F}} = \vec{\mathcal{P}}[\mu, \kappa] \quad (1)$$

Under the usual assumptions of a system corrupted by multivariate Gaussian noise, and a maximum-likelihood approach to the solution, we may define the inverse problem as the optimisation of an objective function

$$\Psi = \frac{1}{2} \sum_{j=1}^S \sum_{i=1}^{M_j} \left(\frac{y_{j,i} - \mathcal{P}_{j,i}[\mu, \kappa]}{\sigma_{j,i}} \right)^2 \quad (2)$$

which we write in vector form as

$$\Psi = \frac{1}{2} (\vec{\mathbf{y}} - \vec{\mathbf{F}})^T \mathbf{R}^{-2} (\vec{\mathbf{y}} - \vec{\mathbf{F}}) = \frac{1}{2} \vec{\mathbf{b}}^T \vec{\mathbf{b}} \quad (3)$$

where $y_{j,i}$ is the data for the i^{th} measurement from source j with standard deviation $\sigma_{j,i}$ and $b_{j,i} = \sigma_{j,i}^{-1} (y_{j,i} - F_{j,i})$ is the *residual data* for this measurement. We use subscripted vectors $\vec{y}_j, \vec{F}_j, \vec{\sigma}_j$ to represent all the relevant quantities for a single source j .

Thus $\mathcal{P}_j[\mu, \kappa]$ is the *projection operator* for source j , \vec{F}_j is the *projection data* obtained by sampling \mathcal{P}_j at the discrete measurement positions $\{\vec{\xi}_{j,1}, \vec{\xi}_{j,2} \dots \vec{\xi}_{j,M_j}\}$ and $\vec{b}_j = \mathbf{R}_j^{-1}(\vec{y}_j - \vec{F}_j)$. Bold vectors $\vec{y}, \vec{P}, \vec{F}, \vec{b}$ represent the combined quantities from all sources. \mathbf{R} is the data-space correlation matrix, which we here take to be

$$\mathbf{R} = \text{diag}(\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_S) = \text{diag}(\sigma_{1,1}, \sigma_{1,2}, \dots, \sigma_{j,i}, \dots, \sigma_{S,M_s}) \quad (4)$$

3. Transport model

Data acquisition for optical tomography is not limited to steady-state attenuation measurements. Time-of-flight systems using an ultra-short pulsed laser as a light source and time-resolved detectors allow the boundary measurement of the temporal intensity response function with a resolution of a few picoseconds [14]. Frequency domain systems use a radio frequency modulated light source and measure the phase shift and modulation amplitude of the transmitted light [15]. While the data acquisition equipment for the steady-state, temporal and frequency domain are quite different, the methods are related from the theoretical viewpoint: the time-of-flight method includes the steady-state case via the integral over time of the detected light, and is related to the frequency domain case via its Fourier transform.

We assume here that the propagation model is the diffusion approximation to the radiative transfer equation which in the frequency domain is:

$$-\nabla \cdot \kappa(\vec{r}) \nabla \hat{\Phi}(\vec{r}, \omega) + \mu(\vec{r}) \hat{\Phi}(\vec{r}, \omega) + \frac{i\omega}{c} \hat{\Phi}(\vec{r}, \omega) = \hat{q}_0(\vec{r}, t), \quad (5)$$

and in the time domain :

$$-\nabla \cdot \kappa(\vec{r}) \nabla \Phi(\vec{r}, t) + \mu(\vec{r}) \Phi(\vec{r}, t) + \frac{1}{c} \frac{\partial \Phi(\vec{r}, t)}{\partial t} = q_0(\vec{r}, t), \quad (6)$$

where Φ is the isotropic photon density, q_0 is an isotropic source distribution and c is the speed of light in the medium. For a detailed discussion of the physical and mathematical models employed in this field, refer to the recent review articles by Hebden and Arridge [6,7]. The model is characterised by the two spatially varying functions $\mu(\vec{r})$ (absorption)¹ and $\kappa(\vec{r})$ (diffusion), which gives rise to the dual search-space nature of the optimisation problem defined in Eq. 2.

For both the frequency-domain and time-domain cases, the boundary measurement $\Gamma(\xi)$ at $\xi \in \partial\Omega$ is related to $\Phi(\vec{r})$ by

$$\Gamma(\xi) = -c\kappa(\xi) \vec{n} \cdot \nabla \Phi(\xi), \quad (7)$$

where \vec{n} is the outer normal of $\partial\Omega$ at ξ . We use the Robin-type boundary condition

$$\Phi(\xi) + \frac{\kappa}{\alpha} \vec{n} \cdot \nabla \Phi(\xi) = 0, \quad (8)$$

where α is a term to incorporate boundary reflections as a result of a refractive index mismatch at $\partial\Omega$ [16,18]. A collimated source incident at $\zeta \in \partial\Omega$ is commonly represented by a diffuse point source $q_0(\vec{r}) = \delta(\vec{r} - \vec{r}_s)$ where \vec{r}_s is located at a depth of one scattering length below the surface.

¹In biomedical optics the absorption parameter is usually denoted μ_a , with the scattering parameter denoted μ'_s , and the diffusion parameter given by $\frac{1}{3(\mu_a + \mu'_s)}$. We will use μ to denote μ_a throughout, to avoid a proliferation of subscripts.

4. Finite element approach

Although the above approach is general, and can be applied to analytical forward models, the most successful approaches use a numerical approach to solve the Forward problem. Here we assume it is an FEM approach.

The domain Ω is divided into P elements, joined at D vertex nodes. The solution Φ is approximated by the piecewise linear function $\Phi^h(\vec{r}) = \sum_i^D \Phi_i u_i(\vec{r}) \in \mathcal{U}^h$, where \mathcal{U}^h is a finite dimensional subspace spanned by basis functions $u_i(\vec{r})$, $i = 1 \dots D$ chosen to have limited support. The problem of solving for Φ^h becomes one of sparse matrix inversion for which standard methods such as Cholesky decomposition or conjugate gradient solvers are readily available. The advantage of the FEM approach is its versatility which makes it applicable to complex geometries and highly inhomogeneous parameter distributions.

As developed in [17,18], the diffusion equation in the FEM framework is expressed, in the frequency domain as :

$$(\mathbf{K}(\kappa) + \mathbf{C}(\mu) + \alpha\mathbf{A} + i\omega\mathbf{B})\vec{\Phi}(\omega) = \vec{Q}(\omega) \quad (9)$$

and in the time-domain as :

$$(\mathbf{K}(\kappa) + \mathbf{C}(\mu) + \alpha\mathbf{A})\vec{\Phi}(t) + \mathbf{B}\frac{\partial\vec{\Phi}(t)}{\partial t} = \vec{Q}(t) \quad (10)$$

where the *system matrices* $\mathbf{K}, \mathbf{C}, \mathbf{A}$ and \mathbf{B} have entries given by :

$$K_{ij} = \int_{\Omega} \kappa(\vec{r}) \nabla u_i(\vec{r}) \cdot \nabla u_j(\vec{r}) d^n \vec{r} \quad (11)$$

$$C_{ij} = \int_{\Omega} \mu(\vec{r}) u_i(\vec{r}) u_j(\vec{r}) d^n \vec{r} \quad (12)$$

$$B_{ij} = \frac{1}{c} \int_{\Omega} u_i(\vec{r}) u_j(\vec{r}) d^n \vec{r} \quad (13)$$

$$A_{ij} = \int_{\partial\Omega} u_i(\vec{r}) u_j(\vec{r}) d(\partial\Omega) \quad (14)$$

To relate the FEM approach to the the forward model, we define :

$$\vec{F}_j = \mathcal{M}[\vec{\Phi}_j] \quad (15)$$

where $\vec{\Phi}_j$ is the solution to Eq. 9 or 10 for the j^{th} source, and $\mathcal{M} : L_2(\Omega) \rightarrow L_2(\partial\Omega)$ is a *measurement operator*.

5. Measurement types

The simplest measurement operator to consider is the *boundary operator* \mathcal{B} which implements Eq. 7. In the discrete case this is a $M_j \times D$ matrix. Since \mathcal{B} is linear it can be applied to either temporal data $\Phi(t)$ or frequency domain data $\Phi(\omega)$. We define the integrated intensity measurement \mathcal{E} to represent both $\mathcal{E}[\Phi(\omega)] = \mathcal{B}[\Phi(\omega = 0)]$ (i.e. the solution to the DC steady problem) or $\mathcal{E}[\Phi(t)] = \mathcal{B}[\int \Phi(t) dt]$ (the sum over time of all arriving photons).

In our approach we have advocated the use of measurement operators that are *self-normalised*

$$\bar{\mathcal{M}} = \frac{\mathcal{T}}{\mathcal{E}} \quad (16)$$

where \mathcal{T} is the integral of $\Gamma(t)$ weighted by a temporal filter. Examples that have been suggested are :

$$\text{time-gated intensity: } \mathcal{M}_{\bar{E}(T)}[\Phi(t)] = \frac{1}{\mathcal{E}[\Phi(t)]} \int_0^T \mathcal{B}[\Phi(t)] dt, \quad (17)$$

$$\text{n-th temporal moment: } \mathcal{M}_{\langle t^n \rangle}[\Phi(t)] = \frac{1}{\mathcal{E}[\Phi(t)]} \int_0^\infty t^n \mathcal{B}[\Phi(t)] dt, \quad (18)$$

$$\text{n-th central moment: } \mathcal{M}_{c_n}[\Phi(t)] = \frac{1}{\mathcal{E}[\Phi(t)]} \int_0^\infty (t - \langle t \rangle)^n \mathcal{B}[\Phi(t)] dt, \quad (19)$$

$$\text{normalised Laplace transform: } \mathcal{M}_{\bar{L}(s)}[\Phi(t)] = \frac{1}{\mathcal{E}[\Phi(t)]} \int_0^\infty e^{-st} \mathcal{B}[\Phi(t)] dt. \quad (20)$$

Additional features could be chosen, such as the logarithmic slope of the temporal decay of Γ , the peak intensity, etc. However, we aim to chose features that satisfy two conditions: robustness of the experimental measurements with respect to systematic errors such as fluctuations of the source power, detector sensitivity, or fibre coupling losses, and secondly the ability of the forward model to generate the corresponding data efficiently. Both conditions are satisfied for measurement types (18), (19) and (20), since these are normalised and therefore not dependent on absolute intensity measurements, and they can be calculated directly by our forward model without the need for explicitly generating the temporal profile of $\Gamma(t)$ [22,23].

6. Reconstruction methods

To solve the optimisation problem, we consider the *gradient* of the objective function. Letting x represent either μ or κ the k^{th} component of the gradient is given by :

$$\frac{\partial \Psi}{\partial x_k} = \sum_{j=1}^S \sum_{i=1}^{M_j} \left(\frac{y_{j,i} - \mathcal{P}_{j,i}[\mu, \kappa]}{\sigma_{j,i}^2} \right) \left(-\frac{\partial \mathcal{P}_{j,i}[\mu, \kappa]}{\partial x_k} \right) \quad (21)$$

leading to the vector form

$$\vec{z} = -\sum_{j=1}^S \mathcal{P}'_j{}^T \mathbf{R}_j^{-1} \vec{b}_j = -\vec{\mathcal{P}}'{}^T \mathbf{R}^{-1} \vec{\mathbf{b}} \quad (22)$$

$$= -\sum_{j=1}^S \mathbf{J}_j{}^T \vec{b}_j = -\mathbf{J}^T \vec{\mathbf{b}} \quad (23)$$

The Jacobian \mathbf{J} of the forward problem is a $M_{\text{TOT}} \times N_{\text{TOT}}$ matrix where $M_{\text{TOT}} = \sum M_j$ and N_{TOT} is the number of coefficients of μ and κ expressed in some basis. In this paper we take the basis to be the element values, although other choices are possible. \mathbf{J} has the structure

$$\begin{pmatrix} \vec{b}_1 \\ \vec{b}_2 \\ \cdot \\ \cdot \\ \cdot \\ \vec{b}_S \end{pmatrix} = \begin{pmatrix} \mathbf{J}_{1,(\mu)} & \mathbf{J}_{1,(\kappa)} \\ \mathbf{J}_{2,(\mu)} & \mathbf{J}_{2,(\kappa)} \\ \cdot & \cdot \\ \cdot & \cdot \\ \cdot & \cdot \\ \mathbf{J}_{S,(\mu)} & \mathbf{J}_{S,(\kappa)} \end{pmatrix} \begin{pmatrix} \Delta \mu \\ \Delta \kappa \end{pmatrix} \quad (24)$$

where we define J_j as the $M_j \times N_{\text{TOT}}$ matrix that is the Jacobian for the objective function corresponding to source j and $J_{j,(\mu)}, J_{j,(\kappa)}$ as the sub-matrices corresponding to the two variables μ, κ .

The Jacobian may be found in a row-by-row fashion if we establish the notation that the vector corresponding to the i^{th} row of J_j is the *Photon Measurement Density Function* $\text{PMDF}_{(j,i)}^{\text{T}}$, as defined in [19,20] which we represent :

$$J_{j_i} = \text{PMDF}_{(j,i)}^{\text{T}} = \left[\text{PMDF}_{(j,i),(\mu)}^{\text{T}} \text{PMDF}_{(j,i),(\kappa)}^{\text{T}} \right] \quad (25)$$

Previously reported reconstruction algorithms are of Newton-type, a prototypical example being the Levenberg-Marquardt algorithm [24,25]. The feature of such methods is that J needs to be explicitly created, and repeatedly inverted. Whenever the number of parameters of the optimisation problem is large, explicit calculation and inversion of the linearised step is regarded as intractable. To overcome this difficulty we have previously reported the use of ART (Algebraic Reconstruction Technique) methods that operate in a row-by-row fashion using only the implicit representation of J given by Eq. 25. However these methods still have a storage overhead that is quite costly [29]. Instead, gradient methods require only the computation of the gradient \vec{z} defined in Eq. 22, the proto-typical example being conjugate gradients. In this method a set of conjugate search directions is generated to find the minimum of the objective function. At each iteration step a one-dimensional line minimisation along the current search direction is performed. Conjugate gradient methods are well-established in nonlinear optimisation. See for example [21]. A brief outline is given in Algorithm 1.

Algorithm 1 Nonlinear conjugate-gradient method

Set search direction $\vec{d}^{(0)} = -\vec{z}(\vec{x}^{(0)})$
Set residual $\vec{p}^{(0)} = \vec{d}^{(0)}$
Define termination criterion ε
Set iteration counter $n = 0$
repeat
Find $\gamma^{(n)}$ that minimises $\Psi(\vec{x}^{(n)} + \gamma^{(n)} \vec{d}^{(n)})$
 $\vec{x}^{(n+1)} = \vec{x}^{(n)} + \gamma^{(n)} \vec{d}^{(n)}$
 $\vec{p}^{(n+1)} = -\vec{z}(\vec{x}^{(n+1)})$
 $\beta^{(n+1)} = \max \left(\frac{\vec{p}^{(n+1)\text{T}} (\vec{p}^{(n+1)} - \vec{p}^{(n)})}{\vec{p}^{(n)\text{T}} \vec{p}^{(n)}}, 0 \right)$
 $\vec{d}^{(n+1)} = \vec{p}^{(n+1)} + \beta^{(n+1)} \vec{d}^{(n)}$
 $n = n + 1$
until $\|\vec{z}(\vec{x}^{(n)})\| < \varepsilon$

Different choices for β exist. Above we used the Polak-Ribière method. Note that this requires restarting of the conjugate gradient method whenever $\beta < 0$ to guarantee convergence. A variety of line search methods can be used, either utilising gradient information, such as the secant method, or using only function evaluations such as the quadratic fit method. Often an exact line search is too computationally expensive due to the large number of function or derivative computations. Inexact line search methods such as Armijo's Rule define the bounds for acceptable step lengths which guarantee convergence. There the line search is terminated when the step length is within the valid range.

7. Gradient calculation

If the gradient is found by the explicit form of Eq. 23 then no advantage is gained. The use of an adjoint scheme to calculate \vec{z} has been discussed previously for the time-domain diffusion model using a Finite Difference Scheme [26], the time-domain transport model using a Finite Difference Scheme [27], and the steady-state DC case using a Finite Element scheme [28]. A summary of these approaches and their relation to other reconstruction methods is given in [29]. Here we give a derivation for the normalised measurement operators that we use in our application.

Consider first the frequency domain problem. We have that

$$\vec{z} = -\mathbf{J}^T \vec{\mathbf{b}} \quad (26)$$

$$= -\sum_{j=1}^S \mathbf{J}_j^T \vec{b}_j \quad (27)$$

$$= -\sum_{j=1}^S \sum_{i=1}^{M_j} \text{PMDF}_{(j,i)}^T b_{j,i} \quad (28)$$

If we define the *Adjoint problem*

$$(\mathbf{K} + \mathbf{C} + \alpha\mathbf{A} - \omega\mathbf{B})\vec{\Phi}_m^+(\omega) = \vec{Q}_i^+(\omega) \quad (29)$$

where $\vec{Q}_i^+(\omega)$ is the *adjoint source* [20], then we have that $\text{PMDF}_{(j,i)}^T$ is a vector in two parts, whose k^{th} components are given by $\sigma_{j,i}^{-1} \vec{\Phi}_i^+(\omega) \times \vec{\Phi}_j(\omega)$ and $\sigma_{j,i}^{-1} \nabla \vec{\Phi}_i^+(\omega) \times \nabla \vec{\Phi}_j(\omega)$ respectively. Here the notation $\vec{a} \times \vec{b}$ as defined in [20] is taken to mean the vector of samples of the product of the fields a, b and $\nabla \vec{a} \times \nabla \vec{b}$ is the vector of samples of the scalar product of the fields $\nabla a \cdot \nabla b$. We may now write

$$\vec{z}_{(\mu)} = -\sum_{j=1}^S \sum_{i=1}^{M_j} \frac{b_{j,i}}{\sigma_{j,i}} \vec{\Phi}_i^+(\omega) \times \vec{\Phi}_j(\omega) \quad (30)$$

$$\vec{z}_{(\kappa)} = -\sum_{j=1}^S \sum_{i=1}^{M_j} \frac{b_{j,i}}{\sigma_{j,i}} \nabla \vec{\Phi}_i^+(\omega) \times \nabla \vec{\Phi}_j(\omega) \quad (31)$$

But $\vec{\Phi}_m^+(\omega)$ is the solution to Eq. 29, so we form a vector

$$\vec{v}_j^+(\omega) = \sum_{i=1}^{M_j} \frac{b_{j,i}}{\sigma_{j,i}} \vec{Q}_i^+(\omega) \quad (32)$$

and solve

$$(\mathbf{K} + \mathbf{C} + \alpha\mathbf{A} - \omega\mathbf{B})\vec{\eta}_j^+(\omega) = \vec{v}_j^+(\omega) \quad (33)$$

leading to:

$$\vec{z}_{(\mu)} = -\sum_{j=1}^S \vec{\eta}_j^+(\omega) \times \vec{\Phi}_j(\omega) \quad (34)$$

$$\vec{z}_{(\kappa)} = -\sum_{j=1}^S \nabla \vec{\eta}_j^+(\omega) \times \nabla \vec{\Phi}_j(\omega) \quad (35)$$

In the following we extend the notation of §2. by adding a superscript to indicate the measurement type. Thus $\mathcal{P}_j^{\mathcal{M}}$ is the projection operator for measurement \mathcal{M} , and $F_{j,i}^{\mathcal{M}}$, $b_{j,i}^{\mathcal{M}}$, $\sigma_{j,i}^{\mathcal{M}}$ are the forward data, residual data, and standard deviation of the error for the i^{th} measurement from source j of type \mathcal{M} . To find the gradient for the normalised measurement types, we use the result from [20] that the Jacobian for $\bar{\mathcal{M}}$ can be constructed from

$$J_{j_i}^{\bar{\mathcal{M}}} = \frac{1}{F_{j,i}^{\mathcal{E}}} \left(J_{j_i}^{\mathcal{M}} - F_{j,i}^{\bar{\mathcal{M}}} J_{j_i}^{\mathcal{E}} \right) \quad (36)$$

from which we can write

$$\mathbf{J}^{\bar{\mathcal{M}}} \mathbf{b}^{\bar{\mathcal{M}}} = \sum_j^S \sum_i^{M_j} J_{j_i}^{\bar{\mathcal{M}}} b_{j,i}^{\bar{\mathcal{M}}} \quad (37)$$

$$= \sum_j^S \sum_i^{M_j} \frac{1}{F_{j,i}^{\mathcal{E}}} \left(J_{j_i}^{\mathcal{M}} - F_{j,i}^{\bar{\mathcal{M}}} J_{j_i}^{\mathcal{E}} \right) b_{j,i}^{\bar{\mathcal{M}}} \quad (38)$$

which leads to the gradient :

$$\bar{z}_{(\mu)}^{\bar{\mathcal{M}}} = \sum_j^S \sum_i^{M_j} \frac{b_{j,i}^{\bar{\mathcal{M}}}}{\sigma_{j,i}^{\bar{\mathcal{M}}} F_{j,i}^{\mathcal{E}}} \left(\mathcal{T} \left[\vec{\Phi}_i^+ \times \vec{\Phi}_j \right] - F_{j,i}^{\bar{\mathcal{M}}} \vec{\Phi}_i^+ \times \vec{\Phi}_j \right) \quad (39)$$

$$\bar{z}_{(\kappa)}^{\bar{\mathcal{M}}} = \sum_j^S \sum_i^{M_j} \frac{b_{j,i}^{\bar{\mathcal{M}}}}{\sigma_{j,i}^{\bar{\mathcal{M}}} F_{j,i}^{\mathcal{E}}} \left(\mathcal{T} \left[\nabla \vec{\Phi}_i^+ \times \nabla \vec{\Phi}_j \right] - F_{j,i}^{\bar{\mathcal{M}}} \nabla \vec{\Phi}_i^+ \times \nabla \vec{\Phi}_j \right) \quad (40)$$

We therefore need to construct two adjoint sources :

$$\vec{v}_j^{\bar{\mathcal{M}}^+}(0) = \sum_{i=1}^{M_j} \frac{b_{j,i}^{\bar{\mathcal{M}}} F_{j,i}^{\bar{\mathcal{M}}}}{\sigma_{j,i}^{\bar{\mathcal{M}}} F_{j,i}^{\mathcal{E}}} \vec{Q}_i^+ \quad (41)$$

$$\vec{v}_j^{\bar{\mathcal{M}}^+}(1) = \sum_{i=1}^{M_j} \frac{b_{j,i}^{\bar{\mathcal{M}}}}{\sigma_{j,i}^{\bar{\mathcal{M}}} F_{j,i}^{\mathcal{E}}} \vec{Q}_i^+ \quad (42)$$

From here we proceed as before, finding two solutions $\vec{\eta}_j^{\bar{\mathcal{M}}^+}(0)$, $\vec{\eta}_j^{\bar{\mathcal{M}}^+}(1)$ and computing the gradient as :

$$\bar{z}_{(\mu)}^{\bar{\mathcal{M}}} = - \sum_j^S \mathcal{T} \left[\vec{\eta}_j^{\bar{\mathcal{M}}^+}(1) \times \vec{\Phi}_j \right] - \vec{\eta}_j^{\bar{\mathcal{M}}^+}(0) \times \vec{\Phi}_j \quad (43)$$

$$\bar{z}_{(\kappa)}^{\bar{\mathcal{M}}} = - \sum_j^S \mathcal{T} \left[\nabla \vec{\eta}_j^{\bar{\mathcal{M}}^+}(1) \times \nabla \vec{\Phi}_j \right] - \nabla \vec{\eta}_j^{\bar{\mathcal{M}}^+}(0) \times \nabla \vec{\Phi}_j \quad (44)$$

This method is applicable for measurement types (18) and (20). The n^{th} central moment given by (19) can be constructed as a linear combination of the first n temporal moments and so the gradient requires n adjoint sources, constructed in a straightforward manner.

8. Results

We present simultaneous reconstructions of μ_a and μ_s using the nonlinear conjugate gradient algorithm in conjunction with the adjoint source scheme for two different test cases: a circular object with three absorbing and/or scattering objects embedded in a homogeneous background, and the model of a slice of the head of a neonate, with a complex distribution of different tissue types and embedded lesions.

The forward data used in the reconstruction of these test cases were simulated by the FEM light transport model, using highly resolved meshes to ensure good quality of the data. Coarser meshes were used for the inverse solution to speed up computation times. For all following reconstructions we used a combination of skew and Laplace transform data. We have argued previously that this yields the best results among the data types listed above [31].

8.1 Circular test object

The test object was a circle of radius 25 mm and three embedded objects, each with a radius of 1.5 mm. The optical parameters of the background medium and the objects are listed in Table 1. The two images in the left column of Fig. 1 show the locations of the perturbations.

Forward data were calculated for 32 source and 32 detector positions, equally spaced along the circumference. Noise was simulated with a noise model presented previously [30][31], assuming that 10^4 photons were collected for each measurement. The meshes for the forward and inverse solvers contained 7200 and 2808 triangular elements, respectively. Figure 1 shows the target images, the reconstruction results obtained with the CG solver, and, for comparison, reconstructions obtained with a steepest descent and an ART (algebraic reconstruction technique) solver. All reconstructions started from the homogeneous background, and were terminated after 100 iterations. The ART solver used was Block-ART with random source ordering as described in [29].

All three algorithms manage to distinguish between absorption and scatter features with relatively little cross-talk, with the exception of the ART algorithm which shows a strong artefact in the absorption image in the position of the scattering feature. None of the methods recovers the absolute values of the perturbations very well, but the CG and ART methods obtain higher peak values than the steepest descent method after a given number of iterations.

Table 1. Optical parameters of circular test object.

	μ_a [mm ⁻¹]	μ_s [mm ⁻¹]	n
background	0.025	2	1.4
objects			
top	0.05	2	1.4
bottom right	0.025	4	1.4
bottom left	0.05	4	1.4

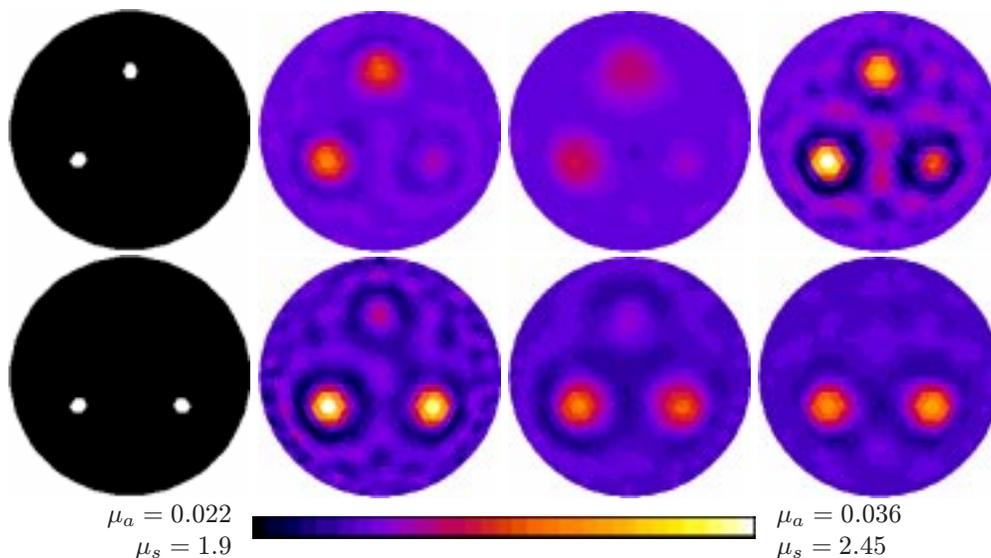


Figure 1. Target image (col. 1), reconstruction after 100 iterations with conjugate gradient method (col. 2), with steepest descent method (col. 3) and with ART method (col. 4). In all cases, the top image is μ_a , and the bottom image is μ_s . The animations linked to the figure show the first 50 reconstruction iterations, and the final 50 in steps of 10.

The L_2 norms of the data errors, $\|\mathbf{b}^{-2}\|$, for the three algorithms are shown in Fig. 2, as a function of the iteration number (left graph) and as a function of the reconstruction runtime (right graph). We find that ART performs better than the CG method at the beginning of the reconstruction, but levels out very quickly. After 100 iterations both reconstructions arrive approximately at the same error norm. With respect to runtime, the CG solver performs better than ART, because a single ART iteration is computationally more expensive than a CG iteration.

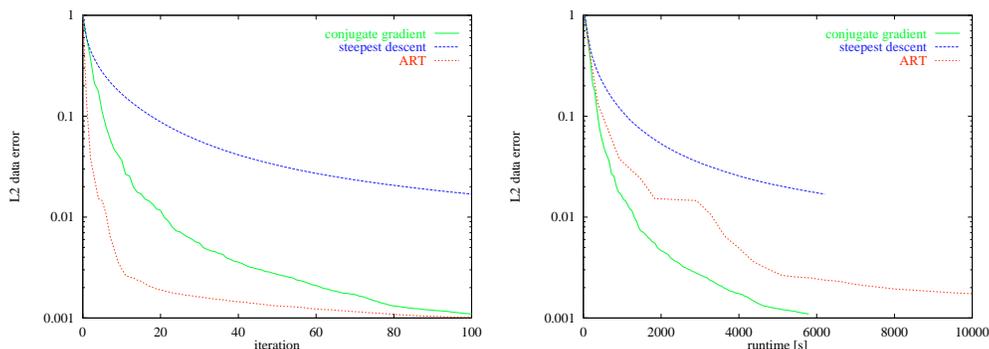


Figure 2. L_2 data norms as a function of iteration number, for different algorithms, as a function of iteration number (left) and of runtime (right).

Figure 3 shows the L_2 norms of the reconstructed images, $\|\vec{x} - \vec{x}_{\text{target}}\|^2$, for both absorption and scatter as a function of the iteration number. The results are similar to the data errors. Again we find that ART performs well during the first few iterations but quickly levels out, in particular for μ_s . For higher iterations the CG algorithm becomes superior. The steepest descent method performs poorly for μ_a and μ_s .

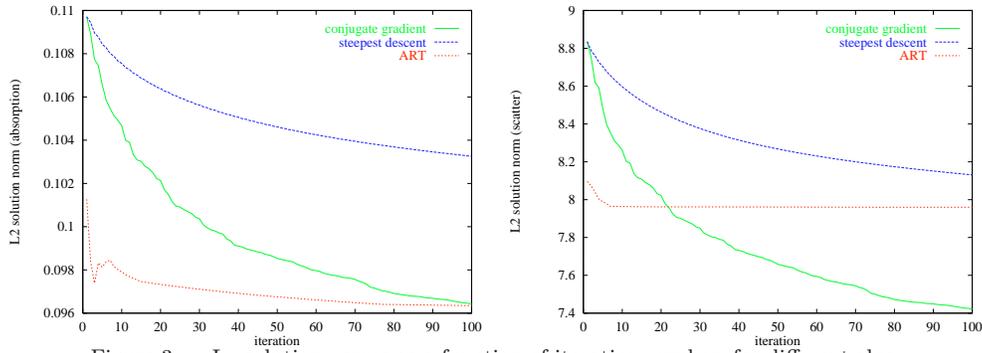


Figure 3. L_2 solution norms as a function of iteration number, for different algorithms. Left: absorption, right: scatter.

8.2 Neonatal head model

To investigate a more complex case we use a mesh that simulates the parameter distribution in a cross-section of a neonatal head. The outline and internal boundaries between tissue types were derived from an MRI image. The sagittal diameter is 80 mm. We distinguish four tissue types and assigned literature values for their optical properties (Table 2). In addition we placed absorbing and scattering perturbations in random locations to simulate lesions or similar high-contrast features.

Skew and Laplace transform forward data for 32 sources and 32 detectors, equidistantly spaced at the boundary, were generated with a high-resolution mesh (41500 elements). For the reconstruction a coarser mesh of 10500 elements was used. The reconstruction was started from a homogeneous distribution, corresponding to the skin parameters ($\mu_a = 0.022$, $\mu_s = 1$). This is likely to be a worst-case scenario, since in practice it may be possible to assume more prior knowledge.

Figure 4 shows the target images and CG reconstructions after 100 iterations. The high contrast objects were generally well localised, in particular for μ_s , while on the other hand the central μ_a object is very diffuse, and the small μ_a object at the lower right was not recovered. There is also some cross-talk from the μ_s object at the upper left into the μ_a image. The fine detail of the background tissues was not recovered. This example demonstrates the difficulties one might expect in the reconstruction of clinical data obtained from complex tissue structures, and it shows that simplistic models of blobs in a homogeneous background, as presented in the previous example, may not be very relevant in the assessment of reconstruction performance of clinical images.

Table 2. Optical parameters of neonatal head model.

		μ_a [mm^{-1}]	μ_s [mm^{-1}]	n
normal tissue	skin	0.022	1	1.4
	bone	0.025	2	1.4
	grey matter	0.02	0.55	1.4
	white matter	0.01	2	1.4
lesions	centre	0.1	0.55	1.4
	centre left	0.05	2	1.4
	top left	0.02	6	1.4
	top right	0.05	4	1.4
	bottom right	0.05	0.55	1.4
	bottom left	0.01	4	1.4

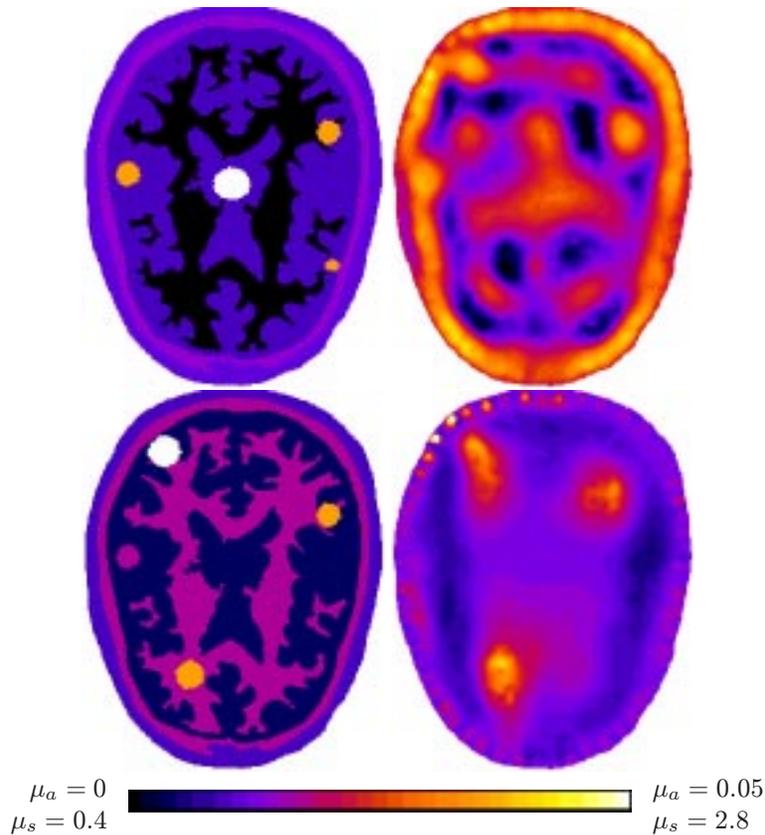


Figure 4. Target image (left col.), reconstruction after 100 iterations with conjugate gradient method (right col.). Top row is μ_a , bottom row is μ_s . The animations linked to the figure show the first 50 reconstruction iterations, and the final 50 in steps of 10.

9. Conclusions

Often it is thought that iterative optimisation based approaches to image reconstruction poses an insurmountable computational burden, especially when the fully three-dimensional and time-dependent problem is addressed. In this paper we have presented a method for deriving the gradient of the objective function in a Finite Element Method based scheme for optical tomography. We presented only the conjugate gradient optimisation method. The use of higher order methods such as Truncated-Newton, is being investigated as well as the choice of appropriate regularisation schemes.

We have presented simultaneous reconstructions of absorption and scatter from simulated data for two cases: a simple homogeneous circle with embedded absorbing and scattering features, and a complex neonatal head model with an inhomogeneous background and several high-contrast lesions. We have presented a comparison between conjugate gradient (CG), steepest descent and ART implementations and shown that after the first iterations, where ART is favourable, CG performs equivalently or even better, if runtime is considered.

We have also demonstrated that the reconstruction of complex problems such as the neonatal head model in the second example, is fundamentally more difficult than the often used blobs in a homogeneous background, and we suggest that such complex cases must be considered in evaluating reconstruction algorithms that are to be employed in

clinical imaging problems.

Acknowledgements

We would like to thank Ken Hanson of Los Alamos National Laboratory, Frank Natterer of Münster University, and Ranadhir Roy of University College London for many useful discussions concerning adjoint methods in optimisation. This work was supported by funding from Action Research, and the Wellcome Trust.