Quasi-Newton methods in iterative image reconstruction for optical tomography

Alexander D. Klose [a,b *], Andreas H. Hielscher [a], and Jürgen Beuthan [b]

[a] State University of New York - Downstate Medical Center
Dept. of Pathology, 450 Clarkson Avenue, Brooklyn, NY 11203

[b] Freie Universität Berlin, Inst. Medizinische Physik, Krahmerstrasse 6–10, 12207 Berlin, Germany

ABSTRACT

Optical Tomography (OT) can provide useful information about the interior distribution of optical properties in various body parts, such as the brain, breast, or finger joints. This novel medical imaging modality uses measured transmission intensities of near infrared light that are detected on accessible surfaces. Image reconstruction schemes compute from the measured data cross sectional images of the optical properties throughout the body. The image quality and the computational speed largely depend on the employed reconstruction method. Of considerable interest are currently so-called model-based iterative image reconstruction schemes, in which the reconstruction problem is formulated as an optimization problem. The correct image equals the spatial distribution of optical properties that leads to a minimum of a user-defined objective function. In the past several groups have developed steepest-gradient-descent (SGD) techniques and conjugate-gradient (CG) methods, which start from an initial guess and search for the minimum. These methods have shown some good initial results, however, they are known to be only slowly converging. To alleviate this disadvantage we have implemented in this work a quasi-Newton (QN) method. We present numerical results that show that QN algorithms are superior to CG techniques, both in terms of conversion time and image quality.

Keywords: optical tomography, image reconstruction, optimization, quasi-Newton method, Broyden-Fletcher-Goldfarb-Shanno method, scattering media, transport theory.

1. INTRODUCTION

Photon propagation in turbid media has been investigated extensively in the recent years. Areas of interest are, for example, atmospheric science and oceanography,[1,2] scattering of stellar and interstellar media.[3] A major field of growing research is biomedical optics, in which near infrared (NIR) light is used to probe highly scattering biological tissues for diagnostical and interventional purposes. Optical tomography is a novel medical imaging technology that uses measurements of transmitted NIR intensities to reconstruct cross-sectional images of various body parts. The technology for making sensitive NIR measurement is nowadays readily available [4-7] and has been applied in a variety of pilot studies concerned with monitoring of blood oxygenation, hemorrhage detection, functional imaging of brain activities, Alzheimer diagnosis, early diagnosis and monitoring of rheumatoid arthritis and breast cancer detection.[8-17] Brain, breast or finger joints are illuminated and the transmitted light is detected on accessible surfaces. These measurements can be done with sources that are time-independent or time-dependent. Time-dependent sources can be further categorized in sources that use a short laser pulse, typically shorter than 100 ps (full-width-half-maximum) or sinusoidally amplitude modulated sources. The measured data is used in conjunction with image reconstruction algorithms to provide diagnostic information about the interior distribution of optical properties inside the human body.

Currently the most promising reconstruction algorithms for OT are so-called nonlinear model-based iterative image reconstruction schemes (MOBIIR).[18-23] In these schemes the image reconstruction problem is formulated as a numerical optimization problem, which consists of three major parts (Fig. 1). First, a forward model for light transport is used to predict the measured data, assuming a certain initial distribution of optical properties. Second, an objective function is evaluated to obtain a measure of difference between the predicted and measured data. In a third step the initial guess is

* aklose@netmail.hscbklyn.edu

updated in a way that reduces the difference between predicted and measured data as defined by the objective function. These steps are repeated until the objective function is minimized.

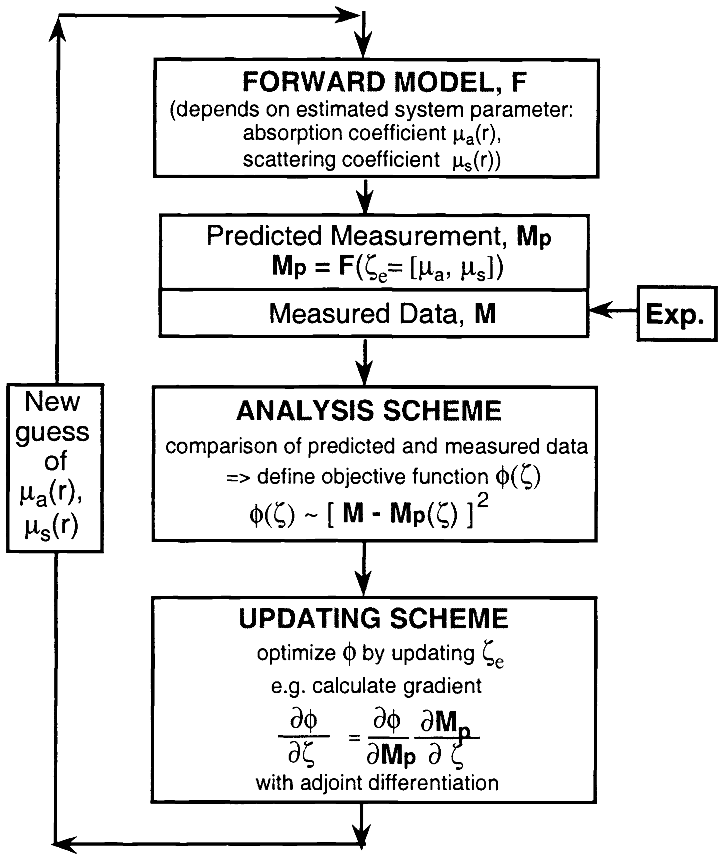


Figure 1. Flow chart of general model-based iterative image reconstruction scheme for optical tomography.

Gradient-based optimization schemes use the derivative of the error norm with respect to the optical parameters to find the minimum of the objective function. In the past several groups have developed steepest-gradient-descent (SGD) techniques, while our group has favored the conjugate-gradient (CG) method.[21,22] These methods have shown some good results, however they are known to be only slowly converging. To alleviate these disadvantages we have implemented a quasi-Newton (QN) method. The performance in terms of accuracy and computational efficiency of this new updating scheme is compared with the conjugate-gradient method.

2. FORWARD MODEL

In recent years many approaches were developed to model the photon migration in scattering tissue. All these methods are based on the theory of particle transport or one of their approximations. The fundamental quantity in radiative transport theory is the radiance $\Psi(r,\Omega)$ at the spatial position r, which is directed into a unit solid angle Ω. Its unit is W cm^{-2} sr^{-1}. The photon transport is described by the time-independent equation of radiative transfer, which is an integro-differential equation given by

$$\omega \cdot \nabla\Psi(r,\omega) + (\mu_a + \mu_s)\Psi(r,\omega) = S(r,\omega) + \mu_s \int_0^{2\pi} p(\omega,\omega')\Psi(r,\omega')d\omega' . \tag{1}$$

The integral of the radiance over all angles Ω at one point r yields the fluence rate

$$\Phi(r) = \int_0^{2\pi} \Psi(r,\omega)\ d\omega . \tag{2}$$

The optical parameters included in the transport equation are the scattering coefficient μ_s and the absorption coefficient μ_a in units of cm^{-1}. The phase function $p(\Omega,\Omega')$ denotes the scattering characteristics of the medium.
A commonly used phase function is the Henyey-Greenstein function, which is defined as [24]

$$p(\cos\theta) = \frac{1-g^2}{2(1+g^2-2g\cos\theta)^{1.5}} \tag{3}$$

where θ is the scattering angle and g the anisotropy factor, which characterizes the angular distribution of scattering. The factor $g = 0$ for isotropically scattering media, $-1 < g < 0$ for backscattering media, and $0 < g < 1$ for forward scattering media. The reduced scattering coefficient μ_s' is defined by $\mu_s' = (1 - g)\ \mu_s$. The boundary condition of the equation of radiative transfer can be described by Fresnel's law for interfaces of different refraction indices.

Analytical solutions to the equation of radiative transfer are only available for simple geometries such as infinite media and isotropic scattering with $g = 0$. For more complex geometries numerical solutions to the transport equation are required. For example, Monte Carlo (MC) methods are widely used in tissue optics. However, MC simulations are computationally very expensive and calculation times often exceed many hours for problems typically encountered in biomedical optics. Therefore, MC schemes are not suitable for iterative reconstruction algorithms in optical tomography, where many forward runs are necessary to minimize the objective function. The diffusion approximations to the transport equation yields solutions with much less computational effort. However, the diffusion approximation can only be applied for scattering media with $\mu_s' >> \mu_a$ and for large source-detector separations. The diffusion approximation is not applicable in cases of low-scattering tissues, high-absorbing tissues, and tissues that contain void-like areas such as the cerebrospinal-fluid-filled ventricles in the brain, or the synovial space in between joints.[25] To be widely applicable we use in this work a finite-difference discrete-ordinate method for photon propagation, which is based on the equation of radiative transfer. Details of this forward scheme are described elsewhere.[22]

We use the forward model to calculate the objective function which we define in this work as the χ^2 error norm

$$X = \chi^2(\mu_a,\mu_s) = \sum_s \sum_d \frac{(M_{s,d} - \Phi_{s,d}(\mu_a,\mu_s))^2}{\Phi_{s,d}^2(\mu_a,\mu_s)}, \tag{4}$$

where $M_{s,d}$ are the measured data and $\Phi_{s,d}$ are the predicted detector readings for a given source s and detector d. The objective function depends nonlinearly on the spatial distribution of optical parameters μ_a and μ_s. In the past we have used conjugate gradient techniques to minimizes this objective function.[21,22] Therefore, only information of the first derivative $\partial X/\partial(\mu_{a,s})$ was used in the minimization process. In this work we extend this approach to include higher derivatives.

3. NEWTON'S METHOD

The minimization of the nonlinear objective function can also be considered as finding a zero of the first derivative of the objective function with respect to the optical parameters. Then the first-order condition $\partial X/\partial(\mu_{a,s}) = 0$ for a local minimizer applies. As we will see, in this case the use of Newton's method in OT also includes second derivative information for the optimization process. This may promise a faster convergence towards to a minimum of the objective function than gradient methods (CG and SGD).

In general, Newton's method is an algorithm for finding a zero of a nonlinear function $g(x) = 0$.[26] It is based on solving a sequence of linear equations $q(x) = 0$ instead of solving the nonlinear equation $g(x) = 0$ directly. Given an estimate x_k of the solution, the nonlinear function $g(x)$ is approximated by the linear function $q(x)$ consisting of the first two terms of the Taylor series expansion of $g(x)$ at the point x_k. The resulting linear system linear system is solved to obtain a new estimate of the solution x_{k+1}. The Taylor series expansion for the nonlinear function $g(x)$ at the point x_k is with $x = x_k + p$:

$$g(x_k + p) \propto q(x_k + p) = g(x_k) + g'(x_k)p \tag{5}$$

$q(x_k + p)$ is a linear function on p. For $q(x_k + p) = 0$ we can solve the Newton equation for p:

$$g'(x_k)p = -g(x_k)\ . \tag{6}$$

Later on, a new estimate x_{k+1} to the solution is calculated with Newton's formula:

$$x_{k+1} = x_k + p\ . \tag{7}$$

We obtain an iteration algorithm for finding a solution:

1. Start with an initial guess x_0 of the solution
2. Test the optimality of the function $g(x)$ at x_k - if x_k is optimal, then stop.
3. Determine an improved estimate of the solution: $x_{k+1} = x_k + p_k$
4. go to 2.

Newton's method converges rapidly and if x_k is close to the solution x, the error is approximately squared at every iteration and the convergence rate is two, respectively. [26]

When Newton's method is employed to minimize the objective function in OT we have to find a solution to $\partial X/\partial(\mu_{a,s}) = f'(x) = 0$. With the Newton equation (6) we can calculate the update p of the optical parameters.

$$F(x_k)p = -f'(x_k)\,. \tag{8}$$

The Hessian matrix F consists of second derivatives of the objective function. The update p is called search direction or Newton direction and points in the general direction of the solution. The direction p is required to be a descent direction at x_k, that is, if $p^T x_k\ f'(x_k) < 0$ (or a stronger condition: the Hessian F has to be positive definite). After computing the update p a new solution of the optical parameters x_{k+1} is found with Newton's formula $x_{k+1} = x_k + p$.

Newton's method is rarely used in its "classical" form for nonlinear programming problems. It can diverge or fail if the initial guess x_o is not "close" enough to a solution x, because the Taylor series was expanded around the solution x. Furthermore, it can even converge to a solution, which is not a minimizer. Additionally, it requires the computation and the storage of the Hessian F, which is often difficult to do. To overcome some of these problems Newton's method is refined by quasi-Newton methods (QN).

4. QUASI-NEWTON METHOD

In the QN approach the Hessian matrix F in the Newton equation (8) is replaced by some approximation matrix B_k to yield $B_k(x)p = -f'(x_k)$. Furthermore, Newton's formula is now expressed by the adjusted equation:

$$x_{k+1} = x_k + \alpha p_k\ . \tag{9}$$

The scalar α is a step length that determines the new point x_{k+1}. We get the "classical" Newton's formula again with $\alpha = 1$. Additionally, Newton's method requires first and second derivatives for computing the search direction p, the solution of a linear system and the storage of the Hessian F. All these requirements have an impact on the computational costs. QN methods result in slower rates of convergence, and tend to use more but cheaper iterations to solve the nonlinear programming problem.

Therefore, calculating the new point x_{k+1} breaks up into two tasks:

1. calculating the search direction p
2. finding a proper step length α .

The step length α is chosen that $f(x_{k+1}) < f(x_k)$. The method, which determines the step length α for guaranteeing this condition is called line search. The line search guarantees the convergence of the minimization algorithm.[26]

The quasi-Newton methods are a compromise to Newton's method and they use different formulas to compute the search direction p_k. The Hessian matrix $F(x_k)$ of the Newton equation is approximated by a matrix B_k, which one can obtain at lower computational costs. Various quasi-Newton methods differ in the computation of B_k.

Quasi-Newton methods are based on generalizations of the secant method for one-dimensional problems to compute the matrix B_k. In doing so, the second derivative $f''(x_k)$ can be approximated only by first derivatives of subsequent iteration steps with

$$f''(x_k) \propto \frac{f'(x_k) - f'(x_{k-1})}{x_k - x_{k-1}}. \tag{10}$$

In the multi-dimensional case for a vector x_k the formula is re-written, because a division by a vector is not allowed:

$$F(x_k)(x_k - x_{k-1}) \propto f'(x_k) - f'(x_{k-1}) . \tag{11}$$

We can define the quasi-Newton approximation for the matrix B_k, which is also called secant condition:

$$B_k(x_k - x_{k-1}) \propto f'(x_k) - f'(x_{k-1}) . \tag{12}$$

Now, two additional vectors are defined:

$$s_k = x_{k+1} - x_k \tag{13}$$

$$y_k = f'(x_{k+1}) - f'(x_k) \tag{14}$$

The secant condition becomes now $B_k s_{k-1} = y_{k-1}$ or in a more appropriate form:

$$B_{k+1} s_k = y_k . \tag{15}$$

Performing a line search with $x_{k+1} = x_k + \alpha p_k$, the vector s_k is $s_k = \alpha_k p_k$. A new update of the matrix B_{k+1}, having the matrix B_k and the vectors s_k and y_k from the former iteration, is given by the Broyden-Fletcher-Goldfarb-Shanno formula (BFGS):

$$B_{k+1} = B_k - \frac{(B_k s_k)(B_k s_k)^T}{s_k{}^T B_k s_k} + \frac{y_k y_k{}^T}{y_k{}^T s_k} \tag{16}$$

Given an approximation to the inverse $H_k = B_k{}^{-1}$ it will be even easier to compute the search direction $p_k = -H_k f'(x_k)$ without having to solve a system of linear equations $B_k(x)p = -f'(x_k)$ at all.

$$H_{k+1} = H_k + \frac{(y_k - H_k s_k) y_k{}^T}{y_k{}^T s_k} - \frac{(y_k - H_k s_k)^T s_k}{(y_k{}^T s_k)^2} y_k y_k{}^T \tag{17}$$

5. NUMERICAL RESULTS

We compared the quasi-Newton methods to the already widely used conjugate-gradient method. The cross section to be reconstructed consists of three objects with a scattering coefficient of 6.0 cm^{-1}, embedded in a 4 x 4 cm^2 background medium with 5.0 cm^{-1}. The absorption coefficient of 1.0 cm^{-1} does not vary within the isotropically scattering medium (g = 0). The medium is surrounded by 4 equally spaced sources and 72 equally spaced detectors. The measured data $M_{s,d}$ at detector positions d for a given source s were generated by using the correct spatial distribution of optical properties and a time-independent finite-difference discrete-ordinate radiation transport code[22] that accurately models the light propagation inside the medium. The "measured data" was corrupted by 30dB Gaussian noise and are used as an input to the reconstruction algorithm. The reconstruction was started with a homogeneous medium $\mu_a = 1.0$ cm^{-1} and $\mu_s = 5.0$ cm^{-1} as an initial guess. The objective function was defined as the χ^2 error norm

$$X = \chi^2(\mu_s) = \sum_s \sum_d \frac{(M_{s,d} - \Phi_{s,d}(\mu_s))^2}{\Phi_{s,d}^2(\mu_s)}. \tag{18}$$

Therefore only μ_s was reconstructed, while μ_a was kept constant. To simulate actual measurements with glass fibers, which only measure the fluence within the acceptance angle (numerical aperture) of the fiber, the angular integration for Φ was limited to a typical numerical aperture of 45^0. The χ^2-error norm is iteratively minimized by the optimization scheme using the CG and QN method. Here, the reconstruction process was stopped after the sum of function evaluations and adjoint calculations reached a maximum of 103. That means, all gradient methods took the same time for reconstructing the final image.

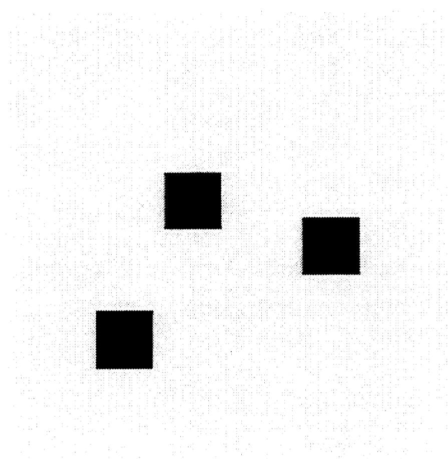
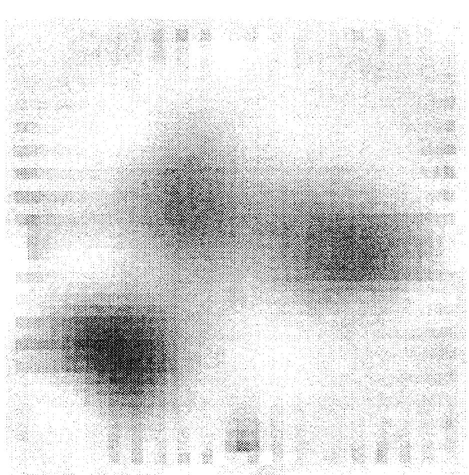
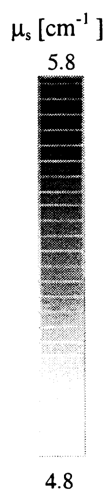


Figure 2. Original cross-section of the scattering coefficient μ_s





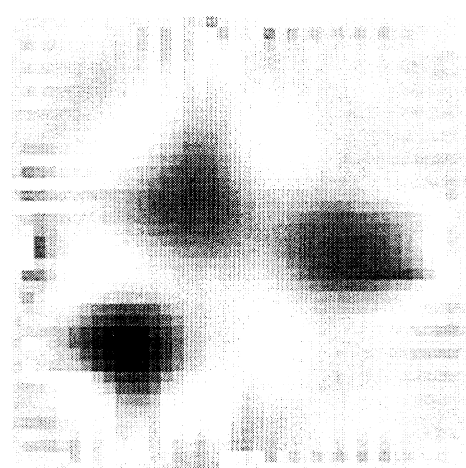


Figure 3. Reconstructed image using the CG-method. **Figure 4.** Reconstructed image using the BFGS-method.

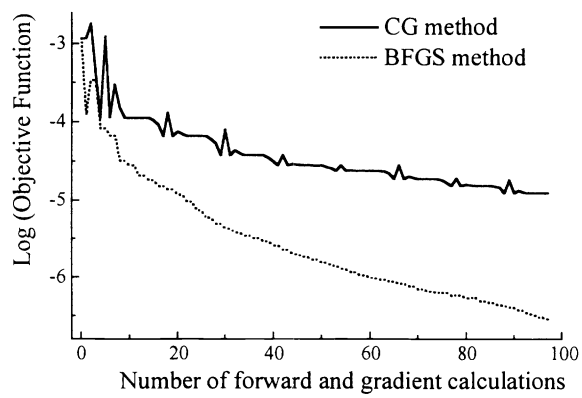


Figure 5. Value of the objective function (Eq. 18) as function of total number of forward and gradient calculations.

But, the absolute number of forward runs and adjoint calculations could differ, depending on the used method. Usually, the CG method used more forward runs and less adjoint calculations than the quasi-Newton method. The value of the objective function as a function of function evaluations is shown in Fig. 5. The reconstructed images of the scattering coefficient are shown in Figs. 3 and 4. The results show that the quasi-Newton algorithm is superior to the conjugate-gradient technique both in terms of conversion time and image quality. Both techniques eventually recover the position of three objects. However, for a fixed number of function evaluations the quasi-Newton method achieves higher contrast (Figs. 4) and a smaller objective function (Fig. 5) than conjugate-gradient method (Fig. 3).

6. SUMMARY

The image reconstruction problem in optical tomography can be considered as a nonlinear optimization problem. The goal is to find the minimum of an objective function that evaluates the difference between measured and predicted data. The predicted data is calculated with a forward model, which depends on the spatial distribution of optical properties within the medium. The majority of currently existing algorithm for OT use only first derivative information of the objective function with respect to the optical properties to find the minimum. In this work we presented a quasi Newton technique that employs approximations to the Hessian matrix, which contains second derivative information. Our results show that using the quasi-Newton method considerably reduces convergence time and improves image quality. The disadvantage of the quasi-Newton method is the large memory requirement for the storage of the Hessian matrix.

ACKNOWLEDGEMENTS

This work was supported in part by The Whitaker Foundation (grant # 98-0244), the New York City Council Speaker's Fund: Toward the Science of Patient Care, and the National Institute of Arthritis and Musculoskeletal and Skin Diseases, a part of the National Institute of Health (grant # R01 AR46255-01), and the Dean's Office of the College of Medicine at the State University of New York (SUNY) Downstate Medical Center.

REFERENCES

1. C. Vanbauce, J.C. Buriez, P. Dubuisson, and Y. Fouquart, "Determination of fog optical thickness over northern france using AVHRR imagery," *Annales Geophysicae - Atmospheres Hydrospheres and Space Science* **11**, pp. 160-172 (1993).
2. Z.H. Jin and K. Stamnes, "Radiative transfer in nonuniformly refracting layered media: Atmosphere ocean system," *Applied Optics* **33** 431-442 (1994)
3. K.M. Chick, J.B. Pollack, and P. Cassen, "The transport of thermal radiation in a prostellar envelope," *Astrophyisical Journal* **461** 956-971 (1996).
4. K. Wells, J. C. Hebden. F. E. W. Schmidt, and D. T. Delpy, " The UCL multichannel time-resolved system for optical tomography", in *Optical Tomograpgy and Spectroscopy of Tissue*, B. Chance and R.R. Alfano, eds., Proc. SPIE Int. Soc. Opt. Eng. **2979**, 599-607 (1997).

5. H. Jess, H. Erdl, K.T. Moesta, S. Fantini, M. A. Franceschini, E. Gratton, and M. Kaschke, "Intensity Modulated Breast Imaging: Technology and Clinical Pilot Study Results," in *Advances in Optical Imaging and Photon Migration*, OSA Trends in Optics and Photonics, Vol. II, R. R. Alfano and J. G. Fujimoto, eds., Optical Society of America, Washington, DC, 126-129 (1996).
6. M. Miwa and Y. Ueda, "Development of time-resolved spectroscopy system for quantitative noninvasive tissue measurement," in *Optical Tomography, Photon Migration, and Spectroscopy of Tissue and Model Media*, B. Chance and R.R. Alfano, eds., Proc. SPIE Int. Soc. Opt. Eng. **2389**, 142-149, (1995).
7. S. Fantini, M .A. Franceschini, J. S. Maier, S. A. Walker, B. Barbieri, and E. Gratton, "Frequency domain multichannel optical detector for noninvasive tissue spectroscopy and oximetry," *Optical Engineering* **34**, 32-42 (1995).
8. J. P. Vanhouten, D. A. Benaron, S. Spilman, and D. K. Stevenson, "Imaging brain injury using time-resolved near-infrared light scanning," Pediatric Research **39**, 470-476 (1996).
9. D. A. Benaron, J. P. Vanhouten, W. Cheong, E.L. Kermit, and R.A. King, "Ealry clinical results of time-of-flight optical tomography in a neonatal intensive care unit," Proc. SPIE Int. Soc. Opt. Eng. **2389**, 582-596 (1995).
10. S. P. Gopinath, C. S. Robertson, C. F. Contant, R. K. Narayan, R. G. Grossman, and B. Chance, "Early detection of delayed traumatic intracranial hematomas using near-infrared spectroscopy," Journal of Neurosurgery **83**, 438-444 (1995).
11. J. Beuthan, V. Prapavat, R.-D. Naber, O. Minet, and G. Müller, "Diagnosis of inflammatory rheumatic diseases with photon density waves," Proc. SPIE Int. Soc. Opt. Eng. **2676**, *Biomedical Sensing, Imaging, and Tracking Technologies I*, R. A. Lieberman, H. Podbielska, T. Vo-Dinh, eds., 43-53 (1996).
12. A. Klose, V. Prapavat, O. Minet, J. Beuthan, and G. Mueller, "RA diagnostics applying optical tomography in frequency-domain," Proc. SPIE Int. Soc. Opt. Eng. **3196**, 194-204 (1997).
13. A. D. Klose, A. H. Hielscher, K. M. Hanson, and J. Beuthan, "Three-dimensional optical tomography of a finger joint model for diagnostic of rheumatoid arthritis," Proc. SPIE Int. Soc. Opt. Eng. **3566**, 151-160 (1998).
14. S. Nioka, Y. Yung, M. Shnall, S. Zhao, S. Orel, C. Xie, B. Chance, and L. Solin, " Optical imaging of breast tumor by means of continuous waves," Advances in Experimental Medicine & Biology **411**, 227-232 (1997).
15. M. A. Franceschini, K. T. Moesta, S. Fantini, G. Gaida, E. Gratton, H. Jess, W. W. Mantulin, M. Seeber, P. M. Schlag, and M. Kaschke, "Frequency-domain techniques enhance optical mammography: initial clinical results," Proceedings of the National Academy of Sciences of the United States of America **94** (12), 6468-6473 (1997).
16. S.R. Arridge, "Optical tomography in medical imaging," Inverse Problems 15, pp. R41-R93 (1999).
17. M.A. Franceschini, K.T. Moesta, S. Fantini, G. Gaida, E. Gratton, H. Jess, W.W. Mantulin, M. Seeber, P.M. Schlag, and M. Kaschke, "Frequency-domain techniques enhance optical mammography: initial clinical results," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 94, pp. 6468-647 (1997).
18. S. S. Saquib, K. M. Hanson, and G. S. Cunningham, "Model-based image reconstruction from time-resolved diffusion data," in *Medical Imaging: Image Processing*, Proc. of the SPIE-The International Society for Optical Engineering, vol. 3034, pp. 369-380, 1997.
19. A.H. Hielscher, A. Klose, D.M. Catarious Jr., K.M. Hanson, "Tomographic imaging of biological tissue by time-resolved, model-based, iterative, image reconstruction," OSA Trends in Optics and Photonics: Advances in Optical Imaging and Photon Migration II, Vol. 21, R.R Alfano and J.G. Fujimoto, eds., Optical Society of America, Washington, DC, pp. 156-161, 1998.
20. S.R. Arridge and M. Schweiger, "A gradient-based optimisation scheme for optical tomography," *Optics Express* **2** (6), 213-226 (1998).
21. A.H. Hielscher, A.D. Klose, K.M. Hanson, "Gradient-based iterative image reconstruction scheme for time-resolved optical tomography," *IEEE Transactions on Medical Imaging* **18**, 262-271 (1999).
22. A.D. Klose and A.H. Hielscher, "Iterative reconstruction scheme for optical tomography based on the equation of radiative transfer," *Medical Physics* **26** (8), 1698-1707 (1999).
23. R. Roy and E.M. Sevick-Muraca,"Truncated Newton's optimization scheme for absorption and fluorescence optical tomography: Part I Theory and formulation, *Optics Express* **4** (10), 353-371 (1999).
24. A.J. Welsh and M.J.C. van Gemert, *Optical-thermal response of laser-irradiated tissue*, Plenum Press, New York, 1995.
25. A.H. Hielscher, R.E. Alcouffe, R.L. Barbour, "Comparison of finite-difference transport and diffusion calculations for photon migration in homogeneous and heterogeneous tissues," *Phys. Med. Biol.* **43**, 1285-1302 (1998).
26. Stephen G. Nash, *Linear and Nonlinear Programming*, McGraw-Hill