

Evolution strategies for optical tomographic characterization of homogeneous media

Andreas H. Hielscher,^{1,2} Alexander D. Klose,^{1,3} and J. Beuthan³

¹State University of New York Downstate Medical Center, Dept. of Pathology, Brooklyn, NY 11203

²Polytechnic University, Dept. of Electrical Engineering, Brooklyn, NY 11001

³Freie Universität Berlin, Dept. of Medical Physics and Laser Medicine, 14195 Berlin, Germany

ahielscher@downstate.edu

Abstract: The reconstruction problem in diffuse optical tomography can be formulated as an optimization problem, in which an objective function has to be minimized. Current model-based iterative image reconstruction schemes commonly use information about the gradient of the objective function to locate the minimum. These gradient-based search algorithms often find local minima close to an initial guess, or do not converge if the gradient is very small. If the initial guess is too far from the solution, gradient-based schemes prove inefficient for finding the global minimum. In this work we introduce evolution-strategy (ES) algorithms for diffuse optical tomography. These algorithms seek to find the global minimum and are less sensitive to initial guesses and regions with small gradients. We illustrate the fundamental concepts by comparing the performance of gradient-based schemes and ES algorithms in finding optical properties (absorption coefficient μ_a , scattering coefficient μ_s , and anisotropy factor g) of a homogenous medium.

©2000 Optical Society of America

OCIS codes: (170.3010) Image reconstruction techniques; (290.7050) Turbid media

References

1. *Optical Tomography and Spectroscopy of Tissue III*, B. Chance, R.R. Alfano, B. J. Tromberg, eds. Proc. of the SPIE Vol. 3597 (1999).
2. *Optical Tomography and Spectroscopy of Tissue II: Theory, Instrumentation, Model, and Human Studies*, B. Chance, R.R. Alfano, A. Katzir, eds. Proc. of the SPIE Vol. 2979 (1997).
3. *Optical Tomography, Photon Migration, and Spectroscopy of tissue and Model Media: Theory, human Studies, and Instrumentation*, B. Chance, R.R. Alfano, A. Katzir, eds. Proc. of the SPIE Vol. 2389 (1995).
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 Tomography and Spectroscopy of Tissue*, B. Chance and R.R. Alfano, Eds., Proc. of the SPIE Vol. 2979, pp. 599-607 (1997).
5. 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).
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. of the SPIE Vol. 2389, pp.142-149 (1995).
7. 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," Proc. of the National Academy of Sciences of the USA, vol. 94, pp. 6468-647 (1997).
8. S.R. Arridge, "Optical tomography in medical imaging," *Inverse Problems* **15**, pp. R41-R93 (1999).
9. S.R. Arridge and J.C. Hebden, "Optical imaging in medicine: II. Modelling and reconstruction," *Phys. Med. Biol.* **42**, pp. 841-853 (1997).
10. R.L. Barbour, H.L. Graber, J.W. Chang, S.L. S. Barbour, P.C. Koo, R. Aronson, "MRI-guided optical tomography: Prospects and computation for a new imaging method," *IEEE Computational Science & Engineering* **2**, pp. 63-77 (1995).
11. Y.Q. Yao, Y. Wang, Y.L. Pei, W. W. Zhu, R.L. Barbour, " Frequency-domain optical imaging of absorption and scattering distributions by Born iterative method," *J. Opt. Soc. Am. A* **14**, pp. 325-342 (1997).
12. K.D. Paulsen and H. Jiang, "Enhanced frequency domain optical image reconstruction in tissues through total variation minimization," *Appl. Opt.* **35**, pp. 3447-3458 (1996).

13. B.W. Pogue, T.O. McBride, J. Prewitt, U.L. Osterberg, K.D. Paulsen, "Spatially variant regularization improves diffuse optical tomography," *Appl. Opt.* **38** (13), 2950-2961 (1999).
14. H. Jiang, K.D. Paulsen, and U.L. Osterberg, "Optical image reconstruction using DC data: simulations and experiments," *Phys. Med. Biol.* **41**, pp.1483-1498 (1996).
15. M.A. O'Leary, D.A. Boas, B. Chance, and A.G. Yodh, "Experimental images of heterogeneous turbid media by frequency-domain diffusion-photon tomography," *Opt. Letters* **20**, 426-428 (1995).
16. D.Y. Paithankar, A.U. Chen, B.W. Pogue, M.S. Patterson, and E.M. Sevick-Muraca, "Imaging of fluorescent yield and lifetime from multiply scattered light reemitted from random media," *Appl. Opt.* **36**, pp. 2260-2272 (1997).
17. M.V. Klivanov, T.R. Lucas, and R.M. Frank, "A fast and accurate imaging algorithm in optical diffusion tomography," *Inverse Problems*, vol. 13, pp. 1341-1361, 1997.
18. O. Dorn, "A transport-backtransport method for optical tomography," *Inverse Problems* **14**, 1107-1130 (1998).
19. 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).
20. 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 Vol. 3034, pp. 369-380, 1997.
21. S.R. Arridge and M. Schweiger, "A gradient-based optimisation scheme for optical tomography," *Opt. Express* **2** (6), 213-226 (1998). <http://www.opticsexpress.org/oearchive/source/4014.htm>
22. 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).
23. R. Roy and E.M. Sevick-Muraca, "Truncated Newton's optimization scheme for absorption and fluorescence optical tomography: Part I Theory and formulation," *Opt. Express* **4** (10), 353-371 (1999). <http://www.opticsexpress.org/oearchive/source/9268.htm>
24. H.P. Schwefel, *Evolution and Optimum Seeking*, John Wiley & Sons, New York, NY 1995.
25. T. Bäck, H.P. Schwefel, "An overview of evolutionary algorithms for parameter optimization," *Evolutionary Computation* **1** (1), 1-23 1993.
26. Z. Michlewicz, *Genetic Algorithms + Data Structures = Evolution Programs*, Springer, New York, NY 1999.
27. W.M. Star, J.P.A. Marijnissen, H. Jansen, M. Keijzer, M.J.C. van Gemert, "Light dosimetry for photodynamic therapy by whole bladder wall irradiation," *Photochem. Photobiol.* **46**, 619-624 (1987).
28. C.J.M. Moes, M.J.C. van Gemert, W.M. Star, J.P.A. Marijnissen, S.A. Prahl, "Measurements and calculations of the energy fluence rate in a scattering and absorbing phantom at 633 nm," *Appl. Opt.* **28**, 2292-2296 (1989).
29. H.J. van Staveren, C.J.M. Moes, J. van Marle, S.A. Prahl, M.J.C. van Gemert, "Light scattering in Intralipid-10% in the wavelength range of 400-1100 nm," *Appl. Opt.* **30**, 4507-4514 (1991).
30. S.T. Flock, S.L. Jacques, B.C. Wilson, W.M. Star, M.J.C. van Gemert, "Optical properties of Intralipid: A phantom medium for light propagation studies," *Lasers Surg. Med.* **12**, 510-519 (1992).
31. A.J. Welch, M.J.C. van Gemert, *Optical-Thermal Response of Laser-Irradiated Tissue*, (New York, NY, Plenum Press 1995).
32. D.A. Pierre, *Optimization Theory with Applications* (Mineola, NY, Dover Publication 1986).
33. W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical Recipes in C*, (New York, NY Cambridge University Press 1992), pp. 420-425.
34. G. Schneider, J. Schuchhardt, P. Wrede, "Evolutionary optimization in multimodal space," *Biological Cybernetics* **74**, pp. 203-207 (1996).
35. W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, *Numerical Recipes in C*, (New York, NY Cambridge University Press 1992), (pp. 288-290).

1. Introduction

In recent years considerable progress has been made towards the development of tomographic imaging systems that rely on strongly scattered near-infrared (NIR) light. [1-3] In this novel medical imaging technique, commonly referred to as diffuse optical tomography (DOT), one attempts to reconstruct the spatial distribution of optical properties (e.g., absorption coefficient $\mu_a(\mathbf{r})$, scattering coefficients $\mu_s(\mathbf{r})$, and scattering anisotropy factor g) within the body from surface measurements of transmitted light intensities. The technology for making such measurements on human subjects is readily available and has been applied in a variety of non-clinical and clinical pilot studies. [4-7] The development of algorithms that accurately and efficiently transform these measurements into cross-sectional images of the internal optical properties remains a major challenge.

Most of the currently employed reconstruction techniques in DOT are model-based iterative image reconstruction schemes (MOBIIR). [8-17] These schemes rely on a model of photon propagation in tissue to predict expected detector readings. All algorithms in References 8-17 employ as a forward model the diffusion equation, which is an approximation to the equation of radiative transfer. While the diffusion equation is easier to solve, it considers only the two parameters μ_a and $\mu_s' = (1-g)\mu_s$. Use of the equation of radiative transfer is required when, besides μ_a and μ_s' , the anisotropy factor g is to be included

as an independent variable.[18,19] Therefore, depending on what equation is used, the input parameters for the forward model are either given by the vector $\xi(\mathbf{r}) = (\mu_a(\mathbf{r}), \mu_s(\mathbf{r}))$ or $\xi(\mathbf{r}) = (\mu_a(\mathbf{r}), \mu_s(\mathbf{r}), g(\mathbf{r}))$ with \mathbf{r} being a position inside the medium under consideration. Starting with an initial guess $\xi_0(\mathbf{r})$ the forward model generates an initial prediction of detector readings $P_{s,d}(\xi(\mathbf{r}))$ for each specific source s and detector d . These predictions are then compared to physical measurement data $M_{s,d}$ using an appropriately defined objective function Φ . Typically the χ^2 -error norm, which is given by

$$\Phi(\xi) = \chi^2(\xi) \equiv \sum_s \sum_d \frac{(M_{s,d} - P_{s,d}(\xi))^2}{2\eta_{s,d}^2} \quad (1)$$

is used. The parameter $\eta_{s,d}$ is a normalization constant, which for example can be set to $M_{s,d}(\xi)$. In this case the error norm minimizes the sum of the squared percentage difference between measured and theoretical data. The challenge remains to find efficient ways of updating the initial guess $\xi_0(\mathbf{r})$ such that the differences between predicted and measured data become smaller and the value of the objective function decreases.

A host of different updating schemes have been devised over the past 8 years.[8-18] Of increasing interest are gradient-based iterative image reconstruction (GIIR) algorithms. [19-23] These codes make use of $d\Phi/d\xi(\mathbf{r})$, which is the gradient of the objective function with respect to the spatial distribution of the optical properties. In the gradient approach the image reconstruction problem is interpreted as a nonlinear optimization problem, in which the distribution $\xi_{min}(\mathbf{r})$ that minimizes the objective function is sought. These schemes have somewhat alleviated the computational burden of the image reconstruction problem in DOT. However, it is well known that gradient-based minimization methods are prone to become trapped in a minimum close to the initial guess. They also have difficulties escaping from parameter regions in the search space with gradients close to zero. In general, gradient methods are not well suited to find the global minimum.

To overcome problems related to gradient-based schemes, we investigate in this work the use of so-called evolution-strategy (ES) algorithms in DOT. ES algorithms borrow from ideas in biological evolution theory to find global minima of objective functions. [24-26] They don't make use of gradient information during the optimization process and are therefore not sensitive to areas with gradients close to zero. In general, ES algorithms have four major components. (see also Fig. 1) Firstly, a parent population is randomly chosen (*initialization*). Each of the N members of the parent population is characterized by its phenotype consisting of n variables. In the next step, pairs of parents are formed, which produce a fixed number of offspring so that a total of M new individuals are available. The offspring are generated by so-called *recombination* and *mutation* rules. Finally, the fitness of each member of the offspring population is evaluated (*selection*). The members with the highest fitness become the new parent generation and go on to generate offspring by recombination and mutation.

To demonstrate how the concepts of ES algorithms can work in DOT we focus in this study on the simple yet significant problem of finding the optical properties of a homogenous medium based on measurements at the boundary of the medium. While this appears to be the simplest of all possible reconstruction problems, it surprisingly has attracted very little attention in the past. Indeed we are not aware of a single publication that addresses this problem. Currently there does not exist a unique and widely accepted method to determine optical properties of homogenous samples. It is noteworthy that neither the National Institute of Standards (NIST) in the US nor any other comparable organization in the world supplies standards with known optical properties for diffuse optical tomography. This poses a particular problem for researchers in optical tomography who want to validate their reconstruction codes. Scientists often use Intralipid suspensions to validate their algorithms. However, over the last 15 years several groups have measured the optical properties of Intralipid with different outcomes. [27-30] Flock et al [30] summarizes these studies by

giving the means and standard deviations of all works combined. They find that the optical properties of 10% Intralipid at 633nm are $\mu_a = (0.027 \pm 0.154) \text{ cm}^{-1}$, $\mu_s = (144 \pm 9) \text{ cm}^{-1}$, and $g = (0.75 \pm 0.18)$. Given these huge standard deviations, especially for μ_a and g , it is clear that code validation is problematic, and that the accurate determination of optical properties of homogeneous media is still a challenging problem in itself.

2. Methods

We compare gradient-based MOBIIR schemes with evolution-strategy-based MOBIIR schemes. Both approaches use the same forward model to predict detector readings and the same definition of the objective function. The forward model used in this study is an upwind-finite-difference, discrete-ordinate code that solves the equation of radiative transfer. [19] This algorithm allows specifying the absorption coefficients $\mu_a(\mathbf{r})$, the scattering coefficient $\mu_s(\mathbf{r})$, and the angle-dependent scattering phase functions. We use in this work the widely applied Henyey-Greenstein phase function

$$p(\cos \theta) = \frac{1 - g^2}{2(1 + g^2 - 2g \cos \theta)^{3/2}}, \quad (2)$$

where g is the already mentioned anisotropy factor, which characterizes the angular distribution of scattering. [31] We apply Fresnel boundary conditions that take the refractive-index mismatch at the air-medium boundary into account. For more details concerning the code, we refer readers to Reference 19. The forward model is used to numerically obtain predictions of the detector readings, $P_{s,d}(\boldsymbol{\xi})$. These predictions depend on the source position, the detector position, and the optical properties $\boldsymbol{\xi} = (\mu_a(\mathbf{r}), \mu_s(\mathbf{r}), g)$. The predicted data is compared to measured data using the objective function defined in Eq. 1. The goal is to find the minimum of this objective function.

2.1. Gradient-Based Optimization Scheme

To minimize Φ the gradient of the objective function $\boldsymbol{\gamma}_o = d\Phi(\boldsymbol{\xi}_o)/d\boldsymbol{\xi}$ at the initial guess $\boldsymbol{\xi}_o$ is calculated. To do this we use in this work a technique called adjoint differentiation. Adjoint differentiation provides an efficient means to calculate the gradient of a numerically evaluated function in a time comparable to one forward calculation.[19,20,22,23] The gradient information is then used within a Polak-Ribieri conjugate-gradient scheme to find the minimum of Φ . [32,33] This approach consists of an *inner* and an *outer* iteration (Fig. 2). During the *inner iteration* a line minimization along a search direction d_k is performed according to

$$\boldsymbol{\xi}_{k+1} = \boldsymbol{\xi}_k + \alpha \cdot d_k \quad (3)$$

where α is a real number representing the step size. The inner iteration consists of several forward calculations in which the parameters $\boldsymbol{\xi}$ are updated until the minimum along this direction is found. Once the minimum is found, a new gradient is calculated at this minimum (*outer iteration*) and another line-minimization along a new direction d_{k+1} is performed. In the Polak Ribieri conjugate gradient scheme this new direction is calculate as follows:

$$d_{k+1} = -\boldsymbol{\gamma}_{k+1} + \frac{\boldsymbol{\gamma}_{k+1}^T (\boldsymbol{\gamma}_{k+1} - \boldsymbol{\gamma}_k)}{|\boldsymbol{\gamma}_k|^2} d_k \quad (4)$$

with $d_0 = \boldsymbol{\gamma}_o$. These steps are repeated until a distribution $\boldsymbol{\xi}_k$ is found for which $|\Phi(\boldsymbol{\xi}_k) - \Phi(\boldsymbol{\xi}_{k-1})|$ is smaller than a user defined value ϵ . A more detailed description of the GIIR algorithm used in this work can be found elsewhere.[19]

2.2. Evolution Strategies(ES)

Evolution strategies do not require the gradient calculation and the search for a minimum is not performed along certain directions. To illustrate the concept we have focused on a simple but important case of determining the optical properties of a homogeneous medium. In this case we deal with only three unknowns, $\xi = (\xi_1 = \mu_a, \xi_2 = \mu_s, \xi_3 = g)$. We start the search process by randomly choosing three ξ with values of $0.1 < \mu_a < 1$, $2.0 < \mu_s < 12$, and $0 < g < 1$. This is our parent population. The vectors $(\xi_1, \xi_2, \xi_3)^k = (\mu_a, \mu_s, g)^k$, $k=1,2,3$, represent the phenotypes of each member of the parent population.

Next we need to generate the offspring generation that correspond to new vectors ξ . In our case, this is done by randomly pairing 12 couples from the parent population. The following six different pairs of parents are equally likely to be chosen: $[\xi^1 - \xi^2]$, $[\xi^1 - \xi^3]$, $[\xi^2 - \xi^3]$, $[\xi^2 - \xi^1]$, $[\xi^3 - \xi^1]$, $[\xi^3 - \xi^2]$. From each pair of parents one child is generated in a two-step process. In the first step properties of the two parents are combined in a procedure generally referred to as recombination. In our case we chose a simple recombination rule by just averaging the optical properties of both parents.

$$\xi_i^{new*} = (\xi_i^{k_1} + \xi_i^{k_2}) / 2 \quad (5)$$

In the second step, the average optical properties are further modified by a process referred to as mutation. We define the mutation rule as

$$\xi_i^{new} = \xi_i^{new*} + \sigma_i y \text{ for } i \in \{1,2,3\}, \quad (6)$$

where y is a normally (Gaussian) distributed random number, which is obtained from

$$y = \sqrt{-2 \log(v_1)} \sin(2\pi v_2) \quad (7)$$

with v_1 and v_2 being independent and uniformly distributed between 0 and 1. [34,35] The values of σ_i where set to $\sigma_i = 1/3 \xi_{i,old}$. Therefore the optical properties of the offspring are Gaussian distributed around the average optical properties of the parents. The width of the Gaussian distribution is given by 1/3 of the actual value of μ_a , μ_s , and g respectively. Other values for σ and other distributions are also possible [24, 26] but have not been explored by us at this point.

Once the offspring are generated, the fitness off all members of the offspring generation is evaluated to select the three fittest members as the new parent generation. The selection is done by performing a set of forward calculations using the finite-difference transport code. The input parameters for the forward code are source and detector positions, and the optical properties $(\mu_a, \mu_s, g)_i$ that characterize each member. The results of the forward calculation are used to determine the value of the objective function as given in Eq. 2. The lower the value of the objective function, the "fitter" is that member of the offspring population. Only the three children that result in the lowest objective-function values survive, and they become the new parents for the next generation. This procedure is iterated until the optical properties of the next generation converge within 2%. The flowchart of the ES scheme is displayed in Fig. 1. A schematic comparison of the gradient-based scheme and the ES schemes is depicted in Fig. 2.

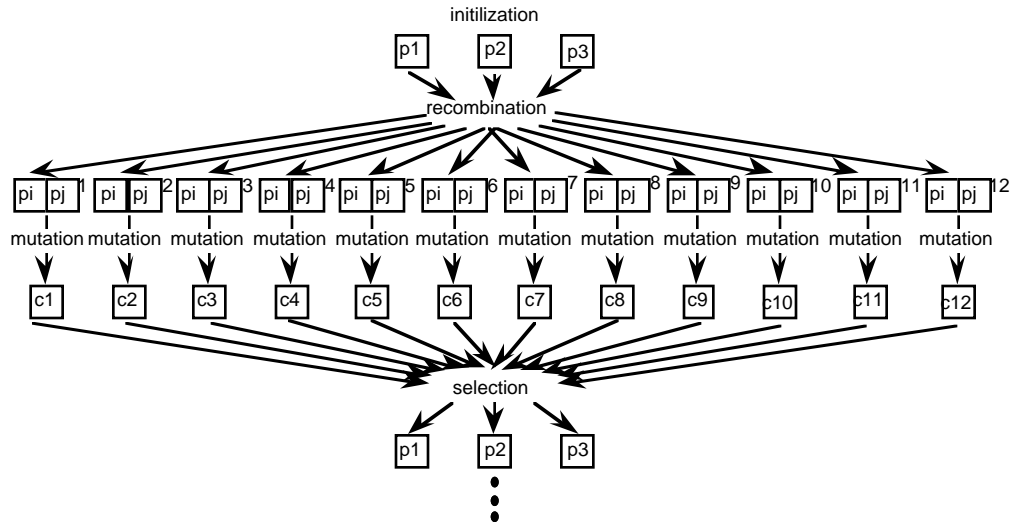


Fig. 1. Basic scheme of the evolution strategy employed in this work. The parent population is initialized by randomly choosing three parents, p_1 , p_2 , p_3 , each with different optical properties μ_a , μ_s , and g . During the recombination process 12 parent pairs $(p_i, p_j)^1$ to $(p_i, p_j)^{12}$ are randomly generated. For each pair the average optical properties are calculated (Eq. 5 recombination rule) and one child (c_1 to c_{12}) is produced by mutation in which optical properties are randomly drawn from a Gauss distribution centered at the average optical properties of the parents (Eq. 6). In the selection process the fittest members of the offspring population are determined by evaluating the objective function for each child (Eq. 2). The three children with the lowest objective function become the three new parents for the next iteration.

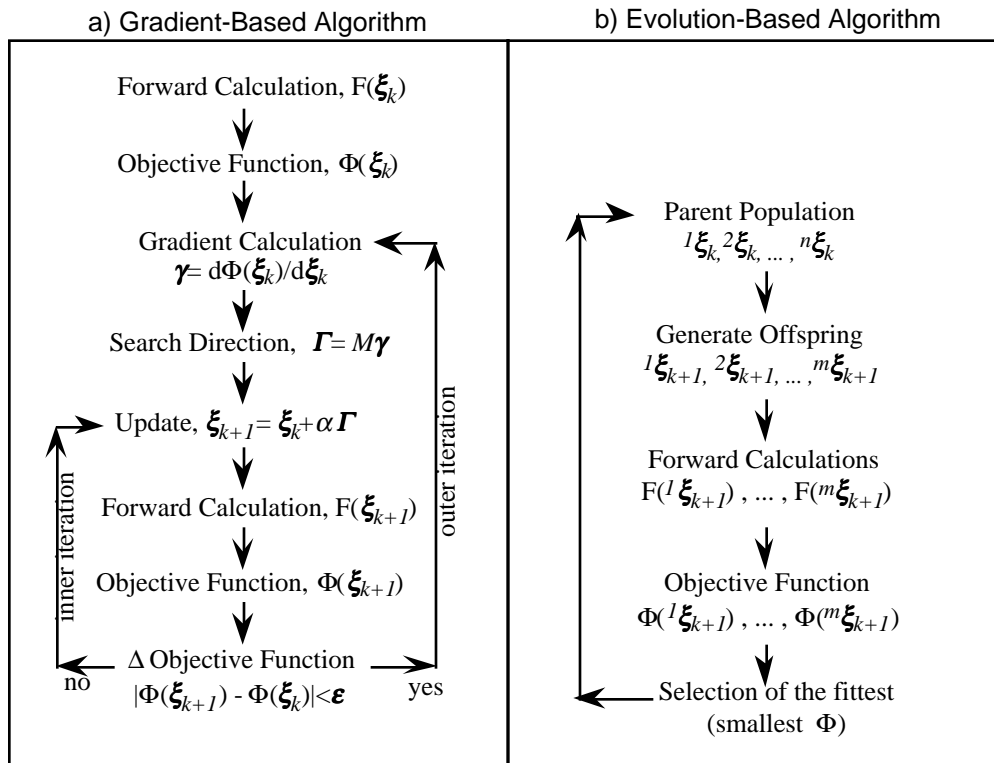


Fig. 2. Comparison of gradient-based reconstruction algorithms (a-left) and evolution-strategy algorithms (b-right) for diffuse optical tomography. The lower index k indicates different updates and generations. The upper indices n and m indicate different parents and children.

3. Results

To illustrate the use of ES schemes in optical tomography we consider the following example. A 4cm x 4cm homogeneous medium is discretized with a spatial resolution of 0.1 cm resulting in a 40x40 finite-difference grid. The optical properties of the medium are $\mu_a = 0.6 \text{ cm}^{-1}$, $\mu_s = 6.0 \text{ cm}^{-1}$, and $g = 0.5$. Synthetic measurements $M_{s,d}$ were generated for 2 source positions and 32 detectors (Fig. 3), using the aforementioned optical parameters and the transport-equation-based forward model. [19]

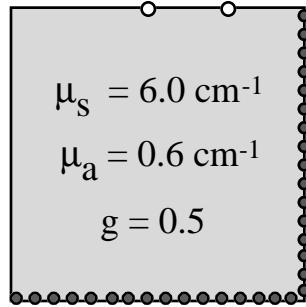


Fig. 3. Setup and measurement geometry for the numerical model. Synthetic data were generated for 2 sources (open circles) and 32 detectors (gray circles) with a time-independent transport-theory-based algorithm.

Using different sets of optical properties (μ_a, μ_s, g), we generated predicted data $P_{s,d}$ and calculated the value of the objective function Φ (Eq. 2). The absorption coefficients μ_a ranged from 0.1 to 1.0 cm^{-1} and g values ranged from 0.3 to 0.8. The μ_s values were chosen so that $1 \text{ cm}^{-1} \leq \mu_s' = (1-g)\mu_s \leq 6 \text{ cm}^{-1}$ in each image. The results are displayed in Fig. 4, where values of $\text{Log}(\Phi)$ are shown as a function of μ_a and μ_s for six different g values. For each g -value we observe an elongated "blue" valley in which the objective function is smallest. The minimum within the valley shifts from the lower left quadrant to the upper right quadrant as g is increased. For $g = 0.3$ the minimum is at ($\mu_a = 0.53 \text{ cm}^{-1}$, $\mu_s = 4.37 \text{ cm}^{-1} \Rightarrow \mu_s' = 3.06 \text{ cm}^{-1}$), and for $g = 0.8$ the minimum is at ($\mu_a = 0.76 \text{ cm}^{-1}$, $\mu_s = 23.4 \text{ cm}^{-1} \Rightarrow \mu_s' = 4.68 \text{ cm}^{-1}$). (The strong influence of the g value on the position of the minimum is noteworthy. Diffusion-equation-based codes that consider only μ_s' cannot distinguish between pairs of g and μ_s values that result in a same μ_s' value. Since the μ_s' range is the same in all 6 maps in Fig. 4, using a diffusion model to predict detector readings and calculate the objective function would yield 6 identical maps!) The global minimum, with $\Phi = 0$, is at $\mu_a = 0.6 \text{ cm}^{-1}$, $\mu_s = 6.0 \text{ cm}^{-1}$, and $g = 0.5$ (Fig. 4c). However, as can be seen from Fig. 4 many combinations of optical properties (μ_a, μ_s, g) lead to similar low values. Assuming that one does not know where the true minimum is, the questions of what algorithm is best suited to find the global minimum arises?

To answer this question, we first employed the gradient-based scheme starting with three different initial guesses (μ_a, μ_s), while keeping the anisotropy factor fixed at $g = 0.5$. The reconstruction scheme found different pairs of (μ_a, μ_s) values for each initial guess. Starting from ($\mu_a = 0.8 \text{ cm}^{-1}$, $\mu_s = 7.0 \text{ cm}^{-1}$) the recovered values were ($\mu_a = 0.52 \text{ cm}^{-1}$, $\mu_s = 6.96 \text{ cm}^{-1}$), starting from ($\mu_a = 0.2 \text{ cm}^{-1}$, $\mu_s = 4 \text{ cm}^{-1}$) the algorithm found ($\mu_a = 0.59 \text{ cm}^{-1}$, $\mu_s = 6.04 \text{ cm}^{-1}$), and starting from ($\mu_a = 0.4 \text{ cm}^{-1}$, $\mu_s' = 10 \text{ cm}^{-1}$) the code determined values of ($\mu_a = 0.39 \text{ cm}^{-1}$, $\mu_s = 9.9 \text{ cm}^{-1}$). Figure 5 shows how the value of the objective function monotonically decreases for all three initial guesses during the first 10 outer iterations. It can also be seen that starting from ($\mu_a = 0.8 \text{ cm}^{-1}$, $\mu_s = 7.0 \text{ cm}^{-1}$) the value of the objective function decreases considerably after the first iteration, but remains almost constant thereafter. Starting from

($\mu_a = 0.2 \text{ cm}^{-1}$, $\mu_s = 4 \text{ cm}^{-1}$) the objective function reaches a plateau value after 7 iterations. Starting from ($\mu_a = 0.4 \text{ cm}^{-1}$, $\mu_s = 10 \text{ cm}^{-1}$), which is a point inside the valley, the value of the objective function changes very little.

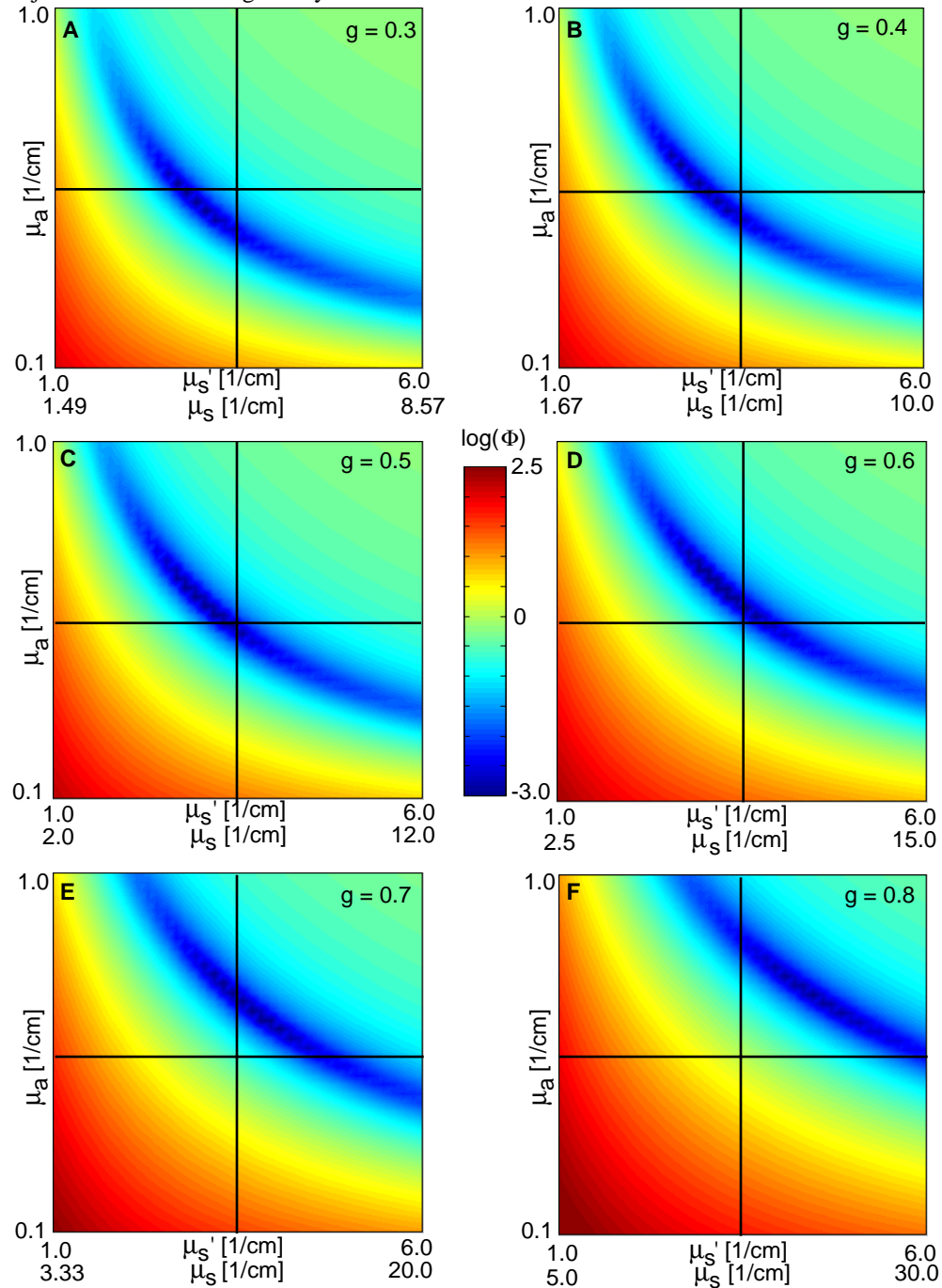


Fig. 4. Contour-map representations of 3-dimensional slices through the 4-dimensional objective function (Eq. 2). Slices are taken for different g values, which results in $\text{Log}(\Phi)$ versus μ_a and μ_s maps. The minimum of $\text{Log}(\Phi)$ is at $\mu_a = 0.6 \text{ cm}^{-1}$, $\mu_s = 6.0 \text{ cm}^{-1}$, $g = 0.5$. Note that the x-axis is chosen so that the reduced scattering coefficient $\mu_s' = (1-g)\mu_s$ ranges from 1 to 6 cm^{-1} in all figures. It can be seen that many combinations of μ_a , μ_s , g lead to similar low values in the $\log(\Phi)$.

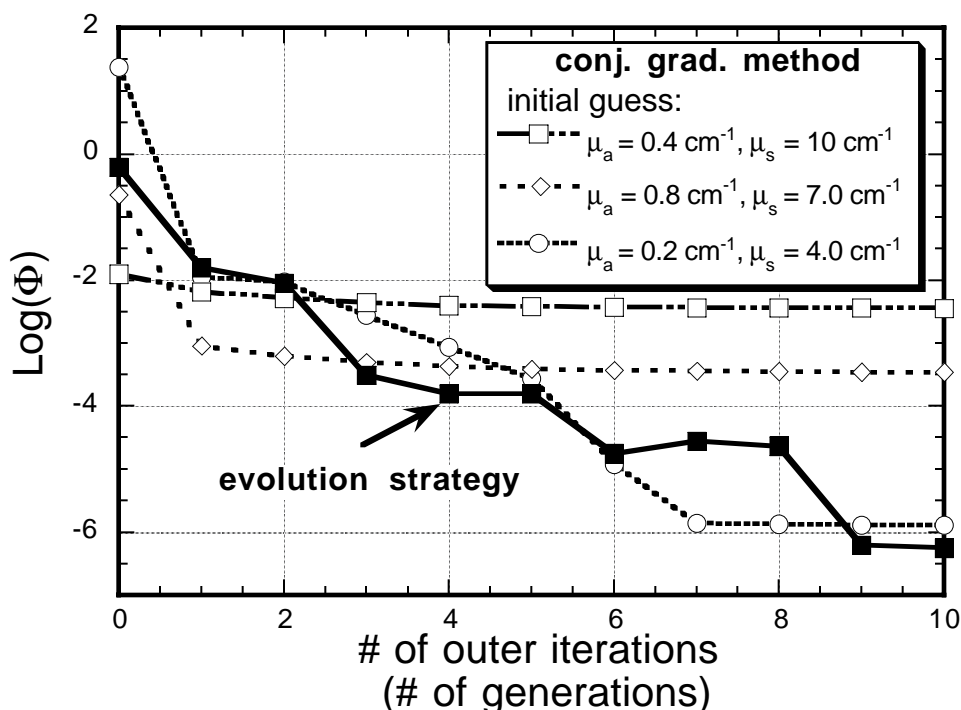


Fig. 5. Objective function (Eq. 2) as a function of outer iterations (conjugate-gradient scheme) and number of generations (evolution-strategy scheme). For the evolution algorithm only the member with the smallest objective function for each generation is shown. The time required to complete all calculations for one generation equals the time it takes to complete one outer iteration within $\pm 10\%$.

Figures 6 and 7 display the path of the iterations on the surface of the objective function in a 3-dimensional and 2-dimensional graph, respectively. Fig. 7 is identical to Fig. 4c, except that in Fig. 7 the μ_a -axis has not been stretched. Therefore the units on both axes are the same, which further emphasizes the narrow character of the valley. It can be seen that the first iteration step in the gradient-based algorithm takes the parameters along the gradient direction directly into the elongated "blue" valley. Once inside the valley only points to the left of the minimum at $(\mu_a = 0.6 \text{ cm}^{-1}, \mu_s = 6.0 \text{ cm}^{-1})$ are significantly improving and get close the minimum. Starting from $(\mu_a = 0.8 \text{ cm}^{-1}, \mu_s = 7.0 \text{ cm}^{-1})$ the search basically stops as soon as the valley is reached. Starting from a point inside the valley at $(\mu_a = 0.4 \text{ cm}^{-1}, \mu_s = 8.0 \text{ cm}^{-1})$ only negligible improvement towards the minimum is observed. In general we found that the gradient-based scheme can find the real minimum only for a limited range of initial guesses. In particular, initial guesses that are located in the lower left quadrant of Fig. 4c fare the best, while initial guesses located in the upper right quadrant perform the worst. This can be understood by taking a closer look at the objective function along the bottom of the valley (Fig. 8). The gradient within the valley is steeper to the left of the minimum than to the right. Furthermore, as can be seen in Figs. 4c, Fig. 6, and Fig. 7, the gradient below the valley is in general larger than the gradient above the valley. Therefore starting at a point below the valley to the left (lower left quadrant) the search algorithm will encounter steeper gradients than starting from a point above the valley to the right (upper right quadrant). The steeper the gradient, the better the gradient based algorithm will perform.

For the ES-based algorithm we randomly chose 3 initial sets of (μ_a, μ_s) representing the initial parents generation. Each subsequent offspring generation had 12 members, which is equivalent to 12 sets of (μ_a, μ_s) . We found that for almost any pair of three initial parents we were able to recover the global minimum within 5% after less than 10 generations. Only in the pathological case when the randomly chosen 3 initial sets of (μ_a, μ_s) , where within 10% of each other would the algorithm not converge to the true minimum. Choosing the same

three initial guesses used for the gradient based schemes in Fig. 5, we plotted the objective function with the lowest value for the each generation alongside the values for the gradient-based schemes. In addition to the lowest value of Φ for each generation, we show in Figs. 6 and 7, the values of all sample points in the ES scheme (blue diamonds).

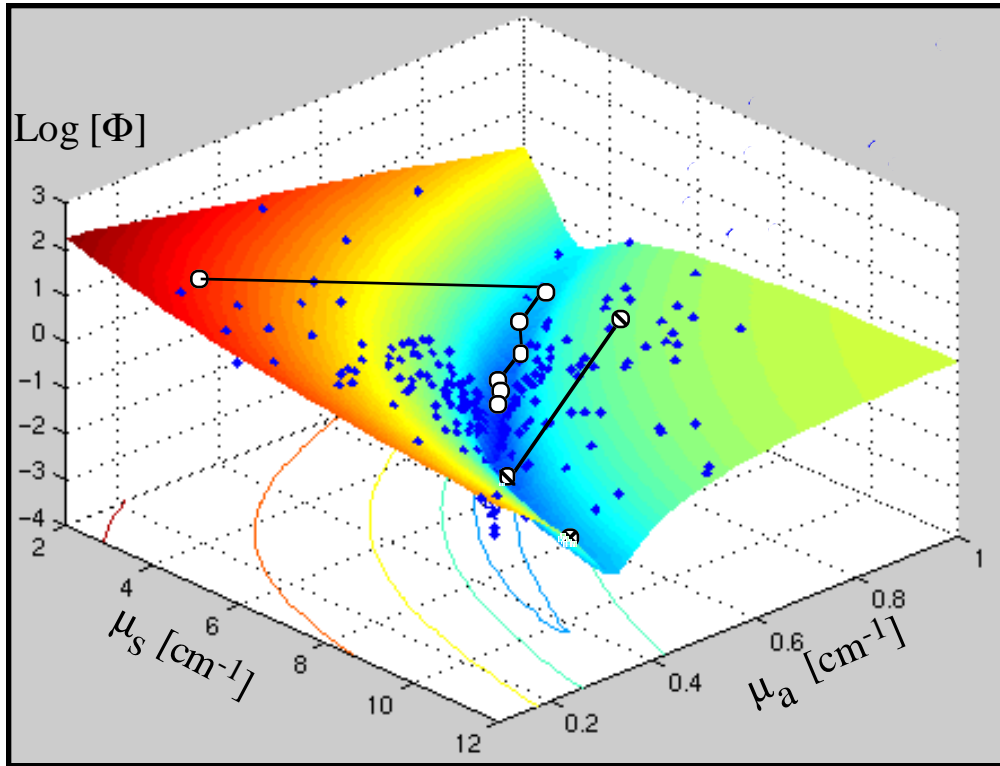


Fig. 6: Three-dimensional depiction of objective function Φ (see Fig. 4c). The μ_a -axis has been stretched by a factor of 10. Open circles indicate values of Φ for outer iterations during the gradient-based algorithm starting from $(\mu_s = 0.2 \text{ cm}^{-1}, \mu_a = 4 \text{ cm}^{-1})$, circles with a line show values of Φ starting from $(\mu_s = 0.8 \text{ cm}^{-1}, \mu_a = 7 \text{ cm}^{-1})$, circles with a cross show values of Φ starting from $(\mu_s = 0.4 \text{ cm}^{-1}, \mu_a = 10 \text{ cm}^{-1})$. The diamonds show sampling points during the ES algorithms.

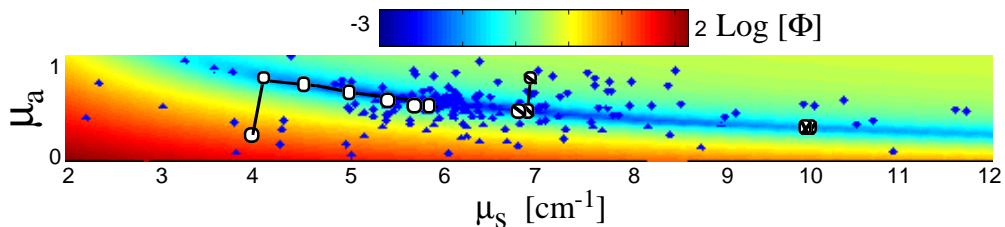


Fig. 7: Same as Fig. 6 and Fig. 4c, only displayed as a 2-dimensional map and without a stretched μ_a -axis. The minimum is located at $(\mu_s = 0.6 \text{ cm}^{-1}, \mu_a = 6.0 \text{ cm}^{-1})$.

The time it takes to complete 10 generations of the ES scheme is on the same order as finding a minimum using 1 initial guess and the gradient-based scheme. In the ES scheme the objective function is evaluated 120 times during 10 generations. The gradient-based scheme requires about the same number of objective function evaluations, since one outer iteration requires about 10-15 inner iterations. Each inner iteration requires one forward

calculation. To assure that a global minimum has been found using a gradient-based optimization scheme one could perform several minimizations starting from different initial guesses and compare the final values of the objective function. In the examples studied so far, we have found that this always takes longer than using the ES algorithm.

Evolution-strategy algorithms do not require more memory than gradient-based schemes. Once the forward calculations for different sets of optical properties are done, only the values of the objective functions for these sets of data need to be stored to perform the recombination, mutation, and selection. In gradient based schemes that use adjoint differentiation for the calculation of the gradient all intensities as calculated by the forward model need to be stored. [19, 22] This typically requires more memory than storage of only the value of the objective function.

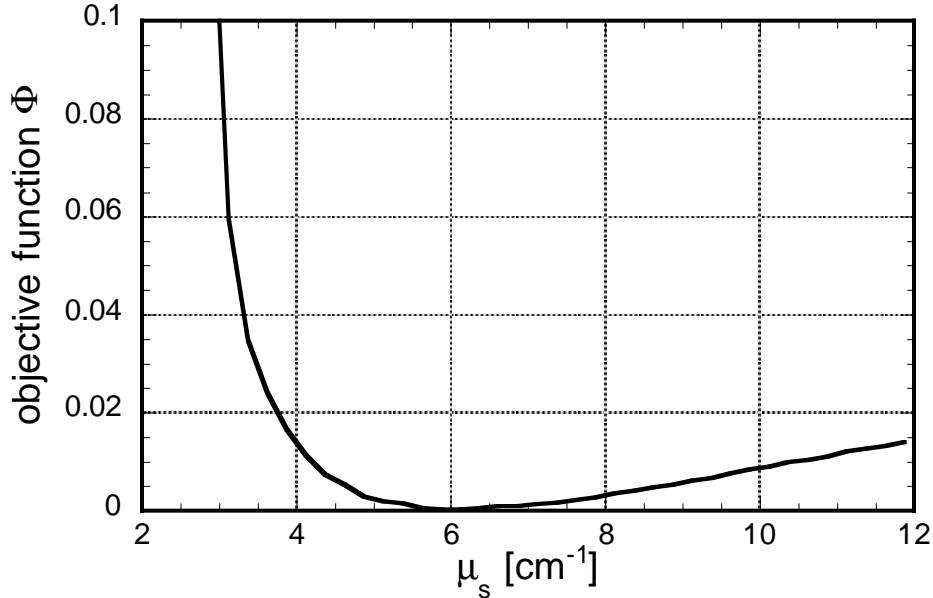


Fig.8: Values of the objective function at the bottom of the elongated, curved (blue) valley in Fig. 6. See also blue valleys in Fig. 4c and 7. Note that in Fig. 4c, 6, and 7 $\text{Log}(\Phi)$ is displayed while here Φ is plotted. It can be seen that the gradient of the objective function is much steeper to the left of the minimum at $\mu_s = 6 \text{ cm}^{-2}$ than to the right of the minimum

4. Summary and Outlook

We have introduced the concept of evolution-strategy for diffuse optical tomography. Unlike currently widely employed gradient-based reconstruction algorithms, evolution-strategy codes don't make use of gradient information during the reconstruction process. Interpreting the reconstruction process as a minimization problem, gradient schemes often get trapped in local shallow valleys, in which the gradient is very small. Therefore these schemes have to rely on a good initial guesses close to the true minimum to be successful. Evolution-strategy algorithms promise to overcome this problem, since they are less dependent on good initial guesses and sample the solution space more effectively. Using the example of finding the correct optical properties of a homogenous medium, we have shown that an evolution-strategy algorithm finds the global minimum more reliably than a conjugate gradient-based reconstruction scheme.

Applying evolution-strategy-based schemes to heterogeneous media, which contain a larger number of unknowns, will result in computationally more intensive operations. For example the convergence speed of an evolution strategy is commonly described by the rate of progress φ . The rate of progress is the ratio of the number of successful mutations to the total

number of mutations. The maximum rate of progress ϕ_{max} is inversely proportional to the number of unknown variables n with $\phi_{max} \propto 1/n$ using Rechenberg's sphere model [24, page 127-141]. That means if the number of variables is increased the maximum rate of progress is decreased, which leads to fewer successful mutations and therefore to a lower convergence speed. In general that means that more generations are required, which in turn leads to a larger number of forward calculations. Furthermore, larger sets of parents are necessary, which will further slow the optimization process. The particular performance of evolution-strategies in optical tomographic problems that involve a larger number of unknowns remains to be studied. It appears likely that some hybrid method that employs gradient-based techniques as well as evolution-strategies may be most successful.

Finally, it should be mentioned that other global minimization techniques exist that are similar to the evolution-strategy (ES) algorithm presented in this work. These other approaches are commonly referred to as evolutionary programming (EP) and genetic algorithms (GAs). The main difference between these three types of approaches (ES, EP and GA) lies in the way the recombination, mutation, and selection rules are formulated and weighted in the search algorithm. Furthermore, GAs represent the object variables (e.g. optical properties) in bit strings that can be thought of as genotypic level of information. In contrast, ES algorithms work on a phenotypic level represented by real-valued vectors. A detailed discussion of the differences among these various approaches and their potential benefits for diffuse optical tomography is beyond the scope of this work, and the interested readers may be directed to References 24-26. However, we believe that the present study clearly encourages further exploration into the many different aspects of global minimization schemes and their potential use in diffuse optical tomography.

Acknowledgements

We would like to thank Prof. Paul Wrede from the Free University Berlin, Dept. of Medical Physics, for many useful discussions concerning evolution strategies for optimization. Furthermore, we thank Mr. Avraham Bluestone SUNY Downstate Medical Center, for his helpful comments concerning this manuscript. This work was supported in part by the National Institute of Arthritis and Musculoskeletal and Skin Diseases, a part of the National Institute of Health (grant # R01 AR46255-01), the Whitaker Foundation, and the Dean of the College of Medicine at the State University of New York - Downstate Medical Center.