Chapter 11

Outlook

So far we presented in this work a MOBIIR scheme for reconstructing optical parameters in tissue. The forward model predicted the detector readings based on the time-independent ERT given a source and the optical parameters. The inverse model determined the distribution of optical parameters within the tissue by using gradient-based optimization techniques given the detector predictions on the tissue boundary. Both, the forward and inverse model, can be extended in such a way that, for example, other data types can be used as detector readings [Schweiger97]. Time-resolved measurements of the fluence $\phi(\boldsymbol{r}, t)$ in conjunction with a time-dependent forward model to calculate the predictions $\boldsymbol{p}(t)$ lead, in turn, to a different objective function Φ. It has been shown that this objective function might be better suited for gradient-based optimization techniques [Schweiger99].

Additionally, the inverse model can be extended in such a way, that the gradient $\nabla_\mu \Phi$ of the objective function is not used for determining a search direction. This is very important in those cases where, for example, the search direction does not point to the minimum (see Subsection 5.2.5) or the gradient is very small resulting in an extraordinary slow convergence towards the solution [Schweiger97]. Stochastic optimization techniques,

such as evolution strategies, promise to address these problems.

In this last chapter we give a brief outlook to future works. We extend the forward model so that it is now based on the time-dependent ERT. Furthermore, we introduce evolution strategies as a means of solving the inverse problem in OT.

11.1 Time-Dependent Forward Model

OT can also be carried out in the time-domain by using a time-dependent light source. A light pulse propagates through the scattering tissue and a detector measures the time-dependent fluence on the tissue boundary. A time-dependent forward model predicts the detector readings on the tissue boundary. It has the advantage over time-independent forward models that additional information, such as the mean time of flight of photons or the shape of the measured light pulse, is available. Subsequently, time-dependent predictions $\boldsymbol{p}(t)$ and measurements $\boldsymbol{m}(t)$ lead to a different objective function Φ, which in turn might be more suitable for gradient-based optimization schemes. We discuss in this section a time-dependent forward model based on the ERT.

The two-dimensional time-dependent ERT is given as

$$\frac{1}{c}\frac{\partial\psi(\boldsymbol{r},\tilde{\boldsymbol{\omega}},t)}{\partial t}+\tilde{\boldsymbol{\omega}}\cdot\nabla\psi(\boldsymbol{r},\tilde{\boldsymbol{\omega}},t)+(\mu_a(\boldsymbol{r})+\mu_s(\boldsymbol{r}))\psi(\boldsymbol{r},\tilde{\boldsymbol{\omega}},t)= \tag{11.1}$$
$$S(\boldsymbol{r},\tilde{\boldsymbol{\omega}},t)+\mu_s(\boldsymbol{r})\int_{2\pi}\tilde{p}(\tilde{\boldsymbol{\omega}},\tilde{\boldsymbol{\omega}}')\psi(\boldsymbol{r},\tilde{\boldsymbol{\omega}}',t)d\tilde{\boldsymbol{\omega}}'.$$

The quantity c is the speed of light in the scattering medium with units of cm ns^{-1} (e.g. $c_0 \approx 30$ cm ns^{-1}, and $c \approx 20$ cm ns^{-1} for refractive index $n = 1.5$). We discretize the spatial and angular variables of Equation 11.1 as previously shown in Chapter 2.3. We get

for example for the ordinates $\xi_k > 0$, $\eta_k > 0$:

$$\frac{1}{c}\frac{\partial\psi_{kij}(t)}{\partial t} + \xi_k\frac{\psi_{kij}(t) - \psi_{ki-1j}(t)}{\Delta x} + \eta_k\frac{\psi_{kij}(t) - \psi_{kij-1}(t)}{\Delta y} + ([\mu_s]_{ij} + [\mu_a]_{ij})\psi_{kij}(t) = \quad (11.2)$$

$$S_{kij}(t) + [\mu_s]_{ij}\sum_{k'} a_{k'}\tilde{p}_{kk'}\psi_{k'ij}(t).$$

In addition, we have to discretize the temporal derivative. The time discretization is performed by using *forward Euler differencing* [Press92]:

$$\frac{\partial\psi(t)}{\partial t} \approx \frac{\psi^{z+1} - \psi^z}{\Delta t}. \quad (11.3)$$

Considering an explicit upwind scheme [Sewell88], we get for the ordinates $\xi_k > 0$, $\eta_k > 0$:

$$\frac{1}{c}\frac{\psi^{z+1}_{kij} - \psi^z_{kij}}{\Delta t} + \xi_k\frac{\psi^z_{kij} - \psi^z_{ki-1j}}{\Delta x} + \eta_k\frac{\psi^z_{kij} - \psi^z_{kij-1}}{\Delta y} + ([\mu_s]_{ij} + [\mu_a]_{ij})\psi^z_{kij} = \quad (11.4)$$

$$S^z_{kij} + [\mu_s]_{ij}\sum_{k'} a_{k'}\tilde{p}_{kk'}\psi^z_{k'ij},$$

and solve for ψ^{z+1}_{kij}:

$$\psi^{z+1}_{kij} = \psi^z_{kij} + c\Delta t S^z_{kij} - c\Delta t\xi_k\frac{\psi^z_{kij} - \psi^z_{ki-1j}}{\Delta x} - c\Delta t\eta_k\frac{\psi^z_{kij} - \psi^z_{kij-1}}{\Delta y} \quad (11.5)$$

$$-c\Delta t([\mu_s]_{ij} + [\mu_a]_{ij})\psi^z_{kij} + [\mu_s]_{ij}\sum_{k'} a_{k'}\tilde{p}_{kk'}\psi^z_{k'ij}.$$

The explicit upwind scheme is conditionally stable for [Sewell88]

$$\Delta t \leq \Delta x/2c. \quad (11.6)$$

The time-dependent fluence $\phi(z\Delta t)$ at one grid point (i, j) on the tissue boundary after the time $z\Delta t$ has elapsed is given by

$$\phi_{ij}(z\Delta t) = \sum_{k\in AP} a_k\psi^z_{kij}. \quad (11.7)$$

We sum over all ordinates with indices k that enter the detector aperture AP. The coefficients a_k are weighting factors, which are determined by the quadrature formula, such as the extended trapezoidal rule [Press92].

11.1.1 Numerical Example

We demonstrate on an example the calculation of the predicted detector readings of a scattering medium with known optical parameters. The numerical model of the scattering medium had dimensions of 3 cm × 3 cm. The optical parameters were $\mu_s = 11.6$ cm^{-1}, $\mu_a = 0.001$ cm^{-1}, $g = 0$, and $n = 1.5$. We considered non-reentry boundary conditions. The source was placed at the center of one side of the phantom. The time-dependent source $S(t)$ generated a light pulse over a time intervall of 0.01 ns. The detectors were positioned on the side opposite the source at positions A, B, and C, and on the side adjacent to the source at positions D, E, and F (see Figure 11.1).

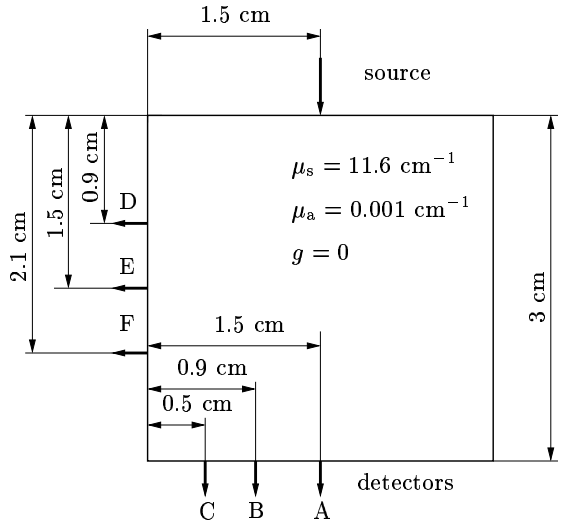
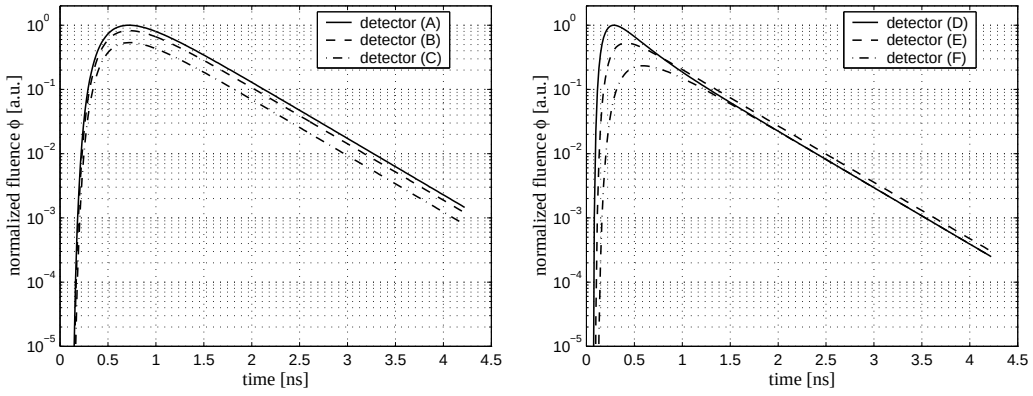


Figure 11.1: Schematic and source-detector configuration of the test medium.

Subsequently, the time-dependent ERT was solved on a 361×361 grid with a spatial step size of $\Delta x = \Delta y = 0.00833$ cm for 32 discrete ordinates. 26000 time points were used with a temporal step size $\Delta t = 0.000163$ ns covering a total time interval of 4.238 ns. The temporal step size had to be relatively small because of the stability criterion (see

Equation 11.6). The calculation time was approximately 107 hours using a PENTIUM III XEON ® processor. The normalized time-dependent fluence along the x-axis is presented in Figure 11.2(a) and along the y-axis in Figure 11.2(b). For better comparison we normalized the calculated data to its largest value of all 3 profiles in each figure.



(a) Fluence for detector positions A, B, and C

(b) Fluence for detector positions D, E, and F

Figure 11.2: Calculated fluence profiles based on the time-dependent ERT. The medium had dimensions of 3 cm × 3 cm and optical parameters $\mu_s = 11.6$ cm^{-1}, $\mu_a = 0.001$ cm^{-1}, $g = 0$, and $n = 1.5$.

We see that the overall trend of the calculated data is correct. The fluence profile for detector A is largest because of its smallest distance to the source position when compared to detectors B and C (Figure 11.2(a)). The same holds for detector D that was closest to the source when compared to detectors E and F (Figure 11.2(b)).

11.1.2 Conclusion

Time-dependent measurement data in conjunction with a time-dependent forward model for light transport might lead to a different objective function, which in turn facilitates the numerical optimization to find the minimum. Therefore, we developed a forward model

for light propagation in scattering media based on the time-dependent ERT. We employed an explicit upwind scheme for calculating iteratively the time-dependent fluence. We found that the calculation time is relatively long (107 hours) compared to the processing time that was required by the time-independent forward model (15 minutes on a 361×361 grid). Consequently, the main focus of future applications of a time-dependent forward model based on the ERT is to utilize other numerical solution methods than the explicit upwind scheme to reduce the computational cost. Furthermore, an experimental evaluation of the predicted detector readings will be necessary as well as the development of the adjoint differentiation scheme.

11.2 Stochastic Optimization Methods

Numerical optimization techniques can be divided into deterministic and stochastic methods. The gradient-based techniques used throughout this work are deterministic methods, because the update of the optical parameters (see Equation 5.4) is based on deterministic rules (e.g. Equations 5.3 and 5.20). However, gradient-based techniques have certain drawbacks that will be briefly discussed.

First, they find local minima of the objective function Φ. A local minimum $\boldsymbol{\mu}^*$ is considered if there is a neighborhood L of $\boldsymbol{\mu}^*$ such that $\Phi(\boldsymbol{\mu}^*) \leq \Phi(\boldsymbol{\mu})$ for $\boldsymbol{\mu} \in L$. The vector $\boldsymbol{\mu}^*$ is a global minimum if $\Phi(\boldsymbol{\mu}^*) \leq \Phi(\boldsymbol{\mu})$ for all $\boldsymbol{\mu}$ [Luenberger84] [Nash96]. The global minimum is difficult to find, because our knowledge of Φ is only local. In turn, the reconstruction result strongly depends on its initial guess $\boldsymbol{\mu}_0$ that was closest to one of the local minima. Second, gradient-based optimization methods use the first derivative $\nabla_{\boldsymbol{\mu}}\Phi$ for calculating a search direction $\boldsymbol{u}$ in the vicinity of a minimum (see Chapter 5.2). However, if the gradient is close to zero ($\|\nabla_{\mu}\Phi\| = 0$), we are not able

to compute a search direction. We observed this type of behavior, when, for example, the minimum of the objective function was inside a long narrow valley (see Figure 11.4). Gradient techniques had difficulties descending inside the long narrow valley of the objective function and the optimization process led to a premature convergence. This result was also observed by Schweiger *et al* [Schweiger97] in diffusion-theory-based OT when using gradient-based optimization techniques. Third, the determination of the scaling factor χ, which was empirically chosen to re-scale the gradient $\nabla_\mu \Phi$ (see Section 6.5), may also lead to a premature convergence when χ, which depends on the initial guess, was not chosen appropriately.

Stochastic optimization methods do not rely on the gradient $\nabla_{\boldsymbol{\mu}} \Phi$ of the objective function for finding the minimum. These methods vary the function variables according to stochastic instead of deterministic rules, and sample the entire search space in order to find the global minimum. Therefore, stochastic optimization methods may provide a way to overcome problems related to deterministic optimization techniques. However, a disadvantage of random strategies is that they are computationally more costly than deterministic methods, especially when large number of unknowns are used.

Commonly used stochastic optimization methods are, for example, simulated annealing (SA) [Kirkpatrick+83], evolutionary programming (EP) [Fogel94], genetic algorithms (GA) [Michalewicz99], and evolution strategies (ES) [Bäck93] [Winter96]. In this outlook we give an introduction to ES as an example for stochastic optimization techniques in OT.

11.2.1 Evolution Strategy

ESs were first introduced by Rechenberg and are algorithms, which imitate the principles of evolution seen in nature [Rechenberg73]. In natural evolution, the mechanism

of variation is the occurrence of random exchanges in the transfer of genetic information to the next generation of individuals. The selection criterion favors individuals best adapted to their environment, which is also known as the *survival of the fittest*. This natural strategy has been adapted to solving numerical optimization problems. Therefore, we exploit a multi-membered ES for reconstructing optical parameters in OT.

A multi-membered population consists of $\nu > 1$ individuals $\boldsymbol{v}^{\mathrm{i}}$. Each individual of a population is composed of a pair of real-valued vectors $\boldsymbol{v}^{\mathrm{i}} = (\boldsymbol{\mu}^{\mathrm{i}}, \boldsymbol{\sigma}^{\mathrm{i}})$. Here, $\boldsymbol{\mu}$ represents the unknown scattering coefficient μ_s and the unknown absorption coefficient μ_a, which is called the *object variable*. The second vector $\boldsymbol{\sigma}$ is a vector of standard deviations σ_{μ_s} and σ_{μ_a}, which is called the *strategy parameter*. Furthermore, each generation consists of a parent population with ν members and an offspring population with λ members. Some individuals of the offspring population in the current generation t become part of a new parent population in the next generation t + 1. The ES repeatedly applies three different operations to the individuals of the parent or offspring population at each successive generation. These operations are called *recombination*, *mutation*, and *selection* [Schwefel95] [Bäck96] [Michalewicz99].

The recombination is applied to the object variables and the strategy parameters of the parent population. Besides *discrete recombination* we applied an *intermediate recombination*. Each new individual $\boldsymbol{v}^{\mathrm{i}}_{new}$ results from taking the average of the corresponding strategy parameters and object variables of two randomly chosen parents $\boldsymbol{v}^{\mathrm{a}}$ and $\boldsymbol{v}^{\mathrm{b}}$ with $\mathrm{a}, \mathrm{b} \in [1, \nu]$:

$$\boldsymbol{v}^{\mathrm{i}}_{new} = \frac{\boldsymbol{v}^{\mathrm{a}} + \boldsymbol{v}^{\mathrm{b}}}{2} \tag{11.8}$$

The recombination process also changes the population size from ν to λ individuals, where the recombination process is applied λ times to the parent population. This results in λ

new individuals $\boldsymbol{v}^{\mathrm{i}}_{new}$ with $\mathrm{i} \in [1, \lambda]$.

Subsequently, the mutation is applied to all new members $\boldsymbol{v}^{\mathrm{i}}_{new}$ producing the offspring population. First, normally distributed random numbers with expectation zero and standard deviations $\boldsymbol{\sigma}^{\mathrm{i}}_{new}$ are added to the object variables $\boldsymbol{\mu}^{\mathrm{i}}_{new}$, indicated by the expression $N(0, \boldsymbol{\sigma}^{\mathrm{i}}_{new})$

$$\boldsymbol{\mu}^{\mathrm{i}} = \boldsymbol{\mu}^{\mathrm{i}}_{new} + N(0, \boldsymbol{\sigma}^{\mathrm{i}}_{new}). \tag{11.9}$$

Second, the strategy parameters $\boldsymbol{\sigma}^{\mathrm{i}}_{new}$ are changed by applying the *1/5 success rule* to optimize the convergence rate [Michalewicz99]. This rule is only applied to every l generations:

$$\boldsymbol{\sigma}^{\mathrm{i}} = \begin{cases} \alpha\boldsymbol{\sigma}^{\mathrm{i}}_{new}, & \text{if } \gamma(\mathrm{l}) < 1/5 \\ \beta\boldsymbol{\sigma}^{\mathrm{i}}_{new}, & \text{if } \gamma(\mathrm{l}) > 1/5 \\ \boldsymbol{\sigma}^{\mathrm{i}}_{new}, & \text{if } \gamma(\mathrm{l}) = 1/5 \end{cases}$$

The term $\gamma(\mathrm{l})$ is the ratio of the number of successful mutations to the number of total mutations during the last l generations. The parameters α and β regulate the increase and decrease rates for the standard deviations. We used following values: $\alpha = 0.82$ and $\beta = 1/0.82 = 1.22$ [Michalewicz99]. The updated individuals $\boldsymbol{v}^{\mathrm{i}}_{new}$ now become the new offspring population.

The selection process determines ν individuals out of the λ members of the offspring population. Only those members are chosen, who have the smallest objective function $\Phi(\boldsymbol{\mu}^{\mathrm{i}})$ for the given object variables $\boldsymbol{\mu}^{\mathrm{i}}$. The selected individuals become members of the parent population of the next generation t+1. An ES that uses this particular selection process is also called a (ν,λ)-ES. Other selection processes can be found, for example, in [Schwefel95].

The iterative process is terminated after the relative change of the objective function in two successive generations $|(\Phi(\boldsymbol{\mu}_{\mathrm{i}})^{\mathrm{t}} - \Phi(\boldsymbol{\mu}_{\mathrm{i}})^{\mathrm{t}+1})/\Phi(\boldsymbol{\mu}_{\mathrm{i}})^{\mathrm{t}}|$ is smaller than a defined ϵ for an individual member $\boldsymbol{v}_{\mathrm{i}}$ of the parent population.

11.2.2 Numerical Example

As an easily tractable example, which illustrates the performance of the ES, we chose to reconstruct a homogeneous medium represented by only two optical parameters (μ_s and μ_a). We generated synthetic measurement data by using the time-independent forward model based on the ERT with Fresnel boundary conditions. The homogeneous test medium with a size of 4 cm $\times$ 4 cm was isotropically scattering and had the optical parameters $\mu_{\mathrm{s}} = 10\ \mathrm{cm}^{-1}$ and $\mu_{\mathrm{a}} = 0.6\ \mathrm{cm}^{-1}$. The refractive index was $n = 1.54$. The calculations were performed for 16 ordinates on a finite-difference mesh of 40 $\times$ 40 grid points and $\Delta x = \Delta y = 0.10256$ cm. We used two different source positions on one side and placed 54 detectors on the other three sides with a separation of 0.2 cm yielding 108 source-detector pairs. The measured data were normalized as given by Equation 3.1.

The reconstruction of the optical parameters μ_s and μ_a was performed by using the synthetic measurements and a (ν, λ)-ES. The anisotropy factor was not reconstructed and was constant ($g = 0$) throughout the reconstruction process. We started with a population of $\nu = 3$ parents having the initial optical parameters $\boldsymbol{\mu}^1 = (\mu_{\mathrm{s}_0} = 4\ \mathrm{cm}^{-1}, \mu_{\mathrm{a}_0} = 0.2\ \mathrm{cm}^{-1})$, $\boldsymbol{\mu}^2 = (\mu_{s_0} = 7\ \mathrm{cm}^{-1}, \mu_{a_0} = 0.8\ \mathrm{cm}^{-1})$, and $\boldsymbol{\mu}^3 = (\mu_{s_0} = 16\ \mathrm{cm}^{-1}, \mu_{a_0} = 0.4\ \mathrm{cm}^{-1})$. The initial standard deviations were set to $\boldsymbol{\sigma}^1 = (\sigma_{\mu_s} = 4, \sigma_{\mu_a} = 0.4)$, $\boldsymbol{\sigma}^2 = (\sigma_{\mu_s} = 4, \sigma_{\mu_a} = 0.4)$, and $\boldsymbol{\sigma}^3 = (\sigma_{\mu_s} = 4, \sigma_{\mu_a} = 0.4)$.

Using the recombination and mutation rules described above throughout the optimization process, the parents produced $\lambda = 20$ offspring at each generation. The recombination process calculated the average value of the optical parameters of two randomly chosen members of the parent generation. These new optical parameters were mutated by applying the mutation rule, which subsequently resulted in a new offspring. After each generation with $\mathrm{l} = 1$ the maximum rate of progress $\gamma(\mathrm{l})$ was calculated. Three offspring with the smallest objective function Φ became parents of the next generation. The ES

located the minimum after 48 generations (960 forward calculations) where the stopping criterion with $\epsilon = 10^{-5}$ was satisfied. The minimum was found at the optical parameters $\mu_s = 9.74$ cm^{-1} and $\mu_a = 0.61$ cm^{-1} with $\log_{10}(\tilde{\varphi}) = -4.43$. We show values $\tilde{\varphi}$ of the objective function for 100 generations in Figure 11.3. The surface of the objective function is displayed in Figure 11.4.

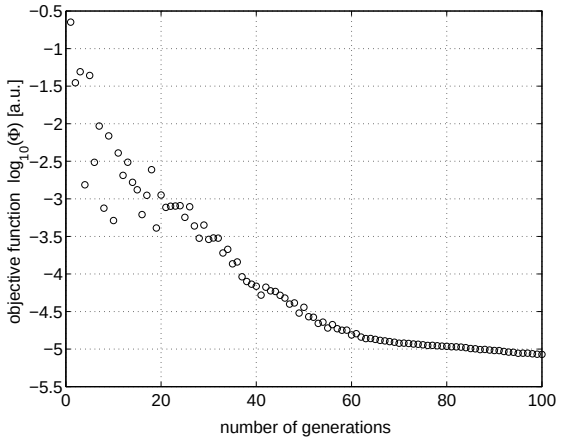


Figure 11.3: Objective function $\log_{10}(\Phi)$ for 100 generations of the (ν, λ)-ES. Each generation t is represented by its smallest value $\tilde{\varphi}$ of the objective function. The minimum was found after 48 generations (stopping criterion $\epsilon = 10^{-5}$).

11.2.3 Conclusion

We applied a (ν, λ)-ES within a MOBIIR scheme to the reconstruction problem of two optical parameters of a homogeneous medium. An ES is a stochastic optimization technique that mimics the evolutionary process found in nature. A population of ν parents, each parent is represented by the optical parameters μ_s and μ_a and strategy parameters σ_{μ_s} and σ_{μ_a}, produces λ offspring by a recombination and mutation process. After selection of offspring with smallest values $\tilde{\varphi}$ of the objective function, a new parent generation is created. This process is repeated for many generations until the stopping criterion is satisfied. In our

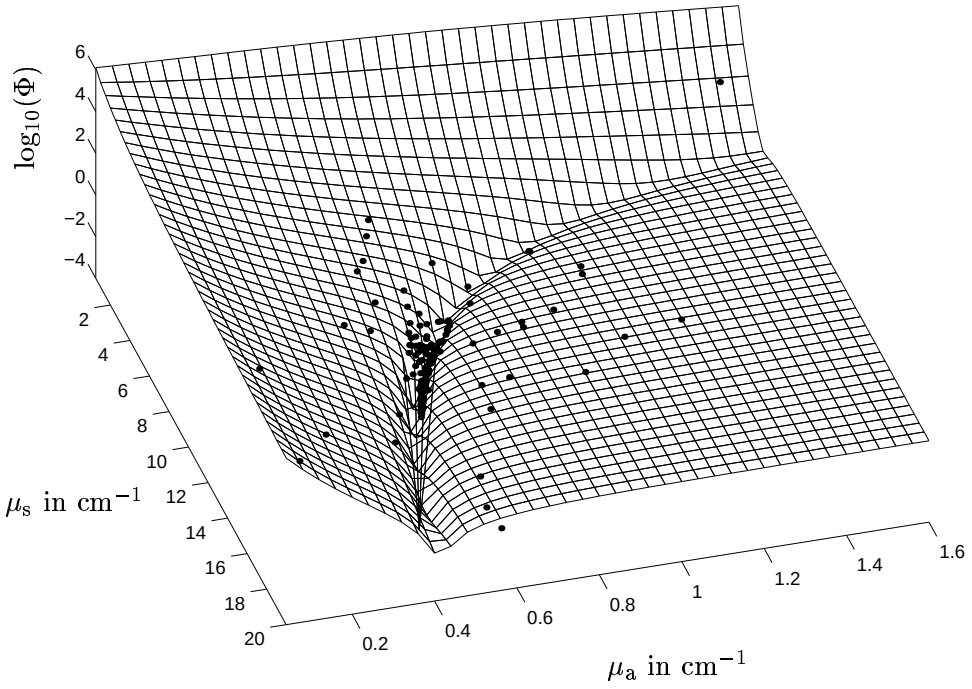


Figure 11.4: Objective function $\log_{10}(\Phi)$ of the optical parameters $\mu_s = 0.5..20$ cm^{-1}, $\mu_a = 0.04..1.6$ cm^{-1}, and $g = 0$. The minimum is at $\mu_s = 10$ cm^{-1} and $\mu_a = 0.6$ cm^{-1}. The search space was sampled by the (ν, λ)-ES, which is displayed by 200 dots. The minimum $\log_{10}(\tilde{\varphi}) = -4.43$ was found at $\mu_s = 9.74$ cm^{-1} and $\mu_a = 0.61$ cm^{-1}.

numerical example we found the minimum after 48 generations (960 forward calculations). The reconstructed optical parameters ($\mu_s = 9.74$ cm^{-1} and $\mu_a = 0.61$ cm^{-1}) deviate by 2.6% from the original scattering coefficient and by 1.7% from the original absorption coefficient.

The numerical example shows that the ES can be employed for reconstructing homogeneous scattering media. The ES might be a promising technique for determining an initial guess of optical parameters μ_{s_0} and μ_{a_0}. These optical parameters could subsequently be used in gradient-based MOBIIR schemes to reconstruct heterogeneous media. This novel stochastic optimization technique we have introduced in OT still needs to be experimentally verified in future studies and has to be applied to practical problems.