Stochastic Optimisation for Coupled Physics Imaging

Simon R. Arridge[1]

[1]Department of Computer Science, University College London, UK

New Ideas in Computational Inverse Problems
Banff International Research Station,
24th-28th October 2022

Outline

Outline

Main collaborators for the work presented in this talk

UCL	P. Beard, M. Betcke, B. Cox, A. Hauptmann, R. Huchli, B. Jin. Z. Kerata, C. Macdonald, B. Treeby
UNottingham	S. Powell
UeF	A. Pulkkinen, T. Tarvainen
UOulu	A. Hauptman
CWI	F. Lucka

(Non linear) inverse problems of the form

$$y^{\text{obs}} = A(x) + e$$

Typically solved as a (regularised) optimisation problem

$$x_* = \arg\min_x \left[\mathcal{D}(y^{\text{obs}}, A(x)) + \lambda R(x)\right] \quad \equiv \quad -\log\left(\underbrace{e^{-\mathcal{D}(y^{\text{obs}}, A(x))}}_{\text{Likelihood}} \underbrace{e^{-R(x)^\lambda}}_{\text{Prior}}\right)$$

In general iterative methods are required e.g

$$x_n = x_{n-1} - \alpha_n \left(\nabla\mathcal{D}(y^{\text{obs}}, A(x_{n-1})) + \lambda\nabla R(x_{n-1})\right)$$

or proximal methods

$$x_n = \text{prox}_{\alpha_n\lambda, R}\left[x_{n-1} - \alpha_n \left(\nabla\mathcal{D}(y^{\text{obs}}, A(x_{n-1}))\right)\right]$$

In discrete setting $A : \mathbb{R}^N \to \mathbb{R}^M$ where M is size of the data. In the case of medical imaging, M maybe large e.g. $O(10^6 - 10^8)$ which leads to computationally demanding reconstruction times.

Consider a partition $S = \{S_1, \ldots, S_{N_s}\}$ of the set $[M]$, i.e. a collection of (sub)sets such that $\emptyset \neq S_t \subset [M]$; $S_{t_1} \cap S_{t_2} = 0$ for $t_1 \neq t_2$ and $\cup_{t=1}^{N_s} S_t = [M]$. Given the partition S we can subdivide the log likelihood term $F(x) = \mathcal{D}(y^{\text{obs}}, A(x))$ into

$$F(x) = \sum_{t=1}^{N_S} F_{S_t}(x) \quad \text{with} \quad F_{S_t}(x) = \sum_{m \in S_t} \mathcal{D}(y_m^{\text{obs}}, A_m(x))$$

In *Ordered* Subset Methods, the partition S is carefully chosen in order to optimise the accucracy and convergence of the iterative algorithm[1]. For example in Positron Emission Tomography (PET) partitions are based on projections at different angles ordered so as to be as nearly "orthogonal" as possible[2] (the 'OSEM' algorithm).

[1] H.M. Hudson and R Larkin, *IEEE Trans Med Im*, 1994.
[2] G. T. Herman and L. B. Meyer, *IEEE Trans Med Im*, 1993

Alternatively, subsets can be chosen *stochastically*, which is the basis for *Stochastic Gradient Descent* (SGD).
These methods have risen to prominence for training Neural Networks algorithms, where the objective is referred to as a *Loss Function*, e.g.

$$F = \frac{1}{N_{\text{train}}} \sum_{i=1}^{N_{\text{train}}} L(y^{(i)}, f_\theta(x^{(i)}))$$

representing the empiral mean estimate of the error in predicting a mapping f_θ between pairs of data in spaces X and Y. use of subsets in this context is referred to as *minibatching*.
Variance reduction methods for accelearating SGD have been extensively investigated (SAGA[3], SVRG[4]) in these applications.
in this talk we consider the use of SGD for some nonlinear inverse problems where the data is naturally considered as a set with an arbitrary random partitioning.

[3] A. Defazio, F Bach, and S. Lacoste-Julien, NIPS 27, 2014
[4] R Johnson and T Zhang, NIPS 26, 2013

Introduction

Imaging from Coupled Physics (ICP)

- High Resolution Imaging Modalities (CT, MRI, Ultrasound) are well established.
- Reconstruction algorithms (usually) simple $\Leftrightarrow$ Inverse Problem well posed (or "mildly ill-posed")
- Contrast often low, precisely because objects of interest **do not** "interfere with" probing wave or radiation

- High Resolution Imaging Modalities (CT, MRI, Ultrasound) are well established.
- Reconstruction algorithms (usually) simple $\Leftrightarrow$ Inverse Problem well posed (or "mildly ill-posed")
- Contrast often low, precisely because objects of interest **do not** "interfere with" probing wave or radiation
- Novel Imaging Modalities (EIT, DOT, Microwave Imaging etc.) provide contrast in novel parameters (conductivity, permitivity, optical absorption etc.).
- Reconstruction algorithms (usually) complex $\Leftrightarrow$ Inverse Problem strongly ill-posed (and usually non-linear)
- Contrast high, because objects of interest **do** "interfere with" probing wave

Introduction

Imaging from Coupled Physics (ICP)

- ICP methods seek to probe with one "wave" and readout with another.
- Contrast from first wave is readout (high resolution) with the second.
- Not the same as Multimodality

Introduction

Imaging from Coupled Physics (ICP)

- ICP methods seek to probe with one "wave" and readout with another.
- Contrast from first wave is readout (high resolution) with the second.
- Not the same as Multimodality

Partial List

- Photo-Acoustic Imaging
- Thermo-Acoustic Imaging
- Acousto-Optic Imaging
- Acoustic-Modulated EIT
- MR-EIT
- Current density EIT
- MR-Elastography
- Ultrasound Elastography

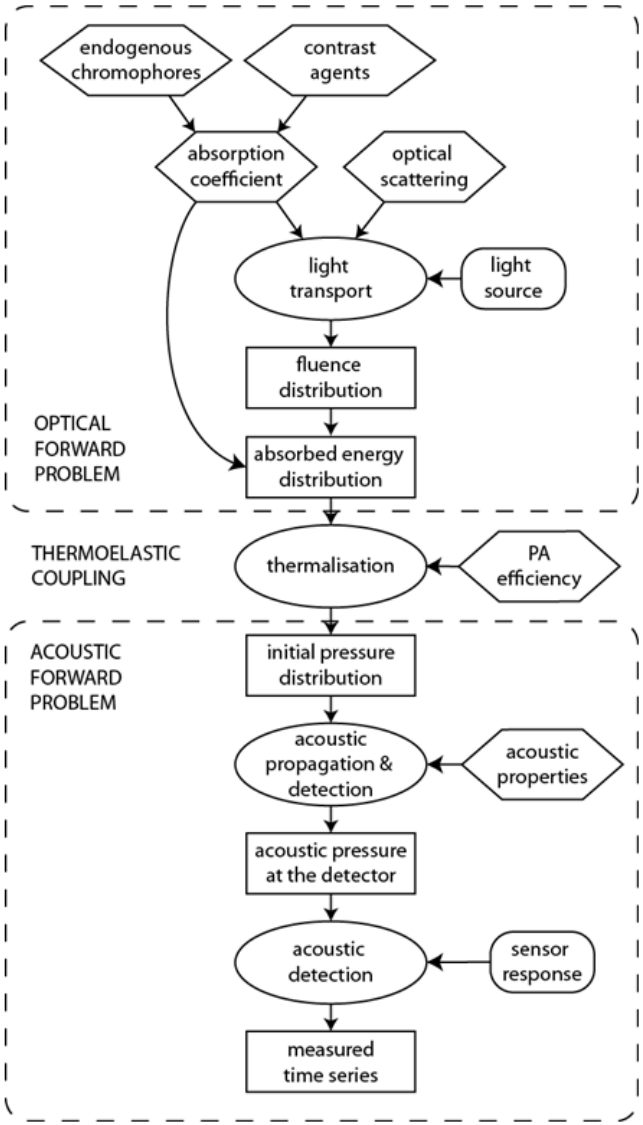
Introduction : PhotoAcoustic Tomography

Motivation

- Optical Imaging : Pros
  - High intrinsic contrast based upon optical absorption and scattering
  - Spectroscopic specificity – chemical information
  - Functional imaging of physiological parameters – blood oxygenation
- Optical Imaging : Cons
  - Imaging depth/spatial resolution limited by strong optical scattering
- Ultrasound Imaging : Pros
  - Images of soft tissue anatomy
  - High spatial resolution: scalable with depth 100's $\mu m \rightarrow \sim mm$
  - Large penetration depth: $\sim 10cm$
  - Physiological information via measurement of blood flow
- Ultrasound Imaging : Cons
  - Weak contrast provided by certain important targets – e.g. the microvasculature
  - Limited specificity: weak sensitivity to chemical differences

PhotoAcoustic Tomography

PhotoAcoustic Signal Generation



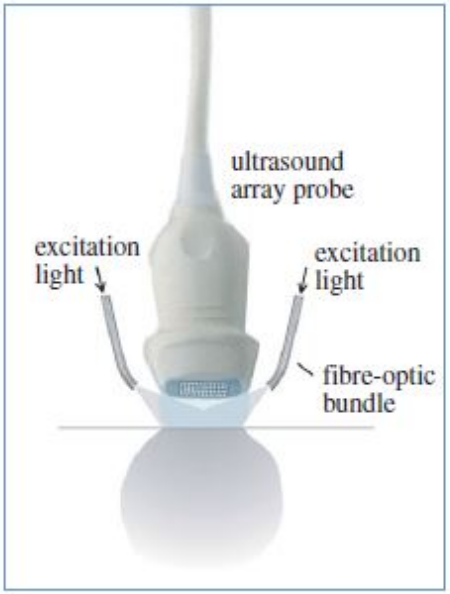
[slide courtesy of Ben Cox]

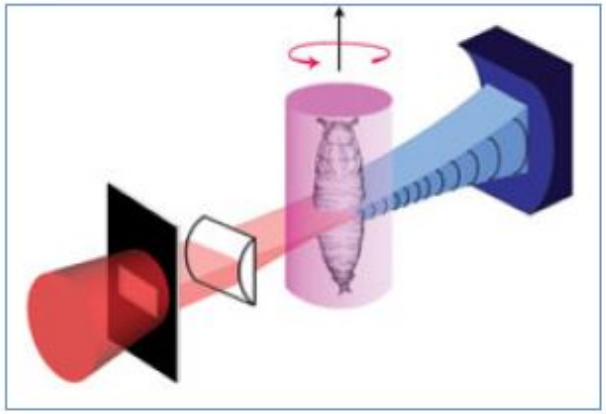
- Naturally occuring contrast agents (chromphores) give rise to optical absorption in the medium.
- The absorption and scattering coefficients μ_a and μ_s determine the fluence distribution Φ,
- $\mu_a \Phi \mapsto H$ (deposited energy).
- $\Gamma H \mapsto p_0$ (pressure distribution) via thermalisation,
- p_0 propagates as an acoustic pulse (elasticity of tissue).
- Sensor detects PA time series $p(t)$.

Cox, A. Laufer, Beard, 2012.

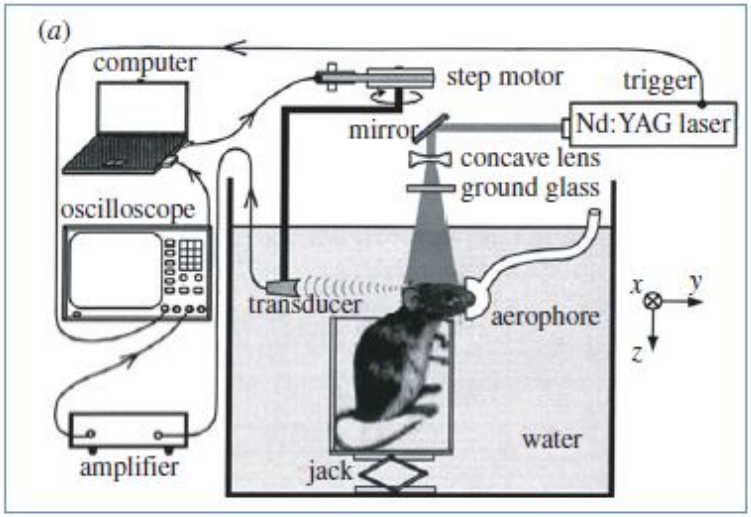
Introduction : PhotoAcoustic Tomography

Some PAT systems

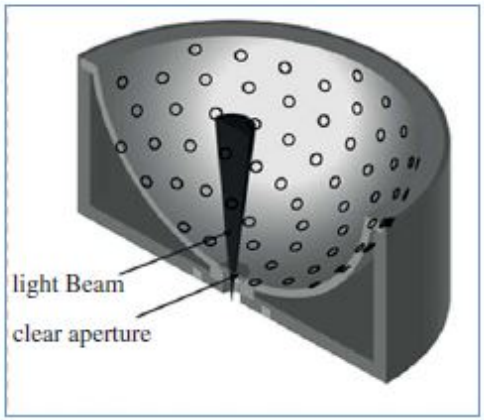




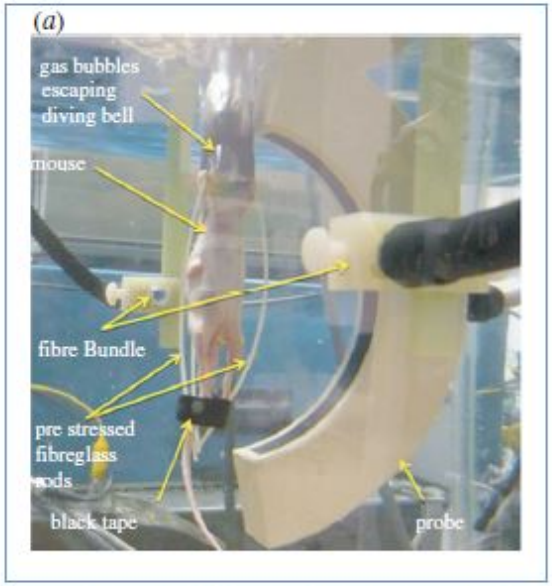
Razansky et al (2009) Nat. Phot. **3**, 412



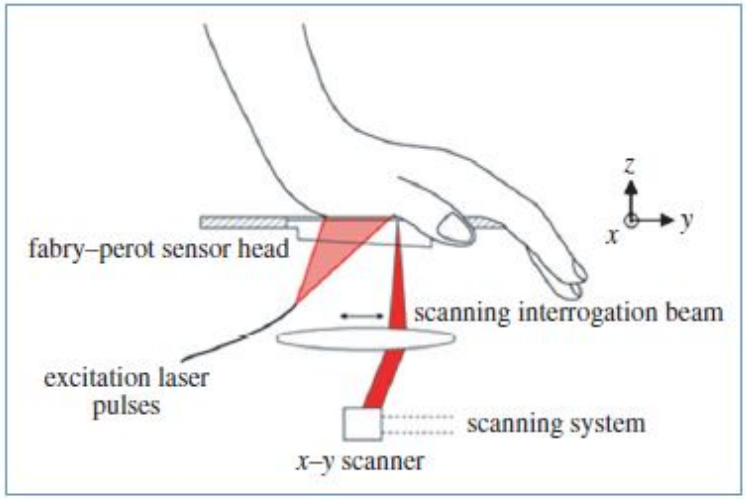
Wang et al. (2003) Nat. Biotech., **21**, 803 – 806.



Kruger et al. (2010) Med. Phys. **37**, 6096



Brecht et al. (2009) JBO, **14**, 064007



Zhang et al (2008) Appl. Opt., **47**, 561-577

Introduction

Ultrasound Modulated Optical Tomography

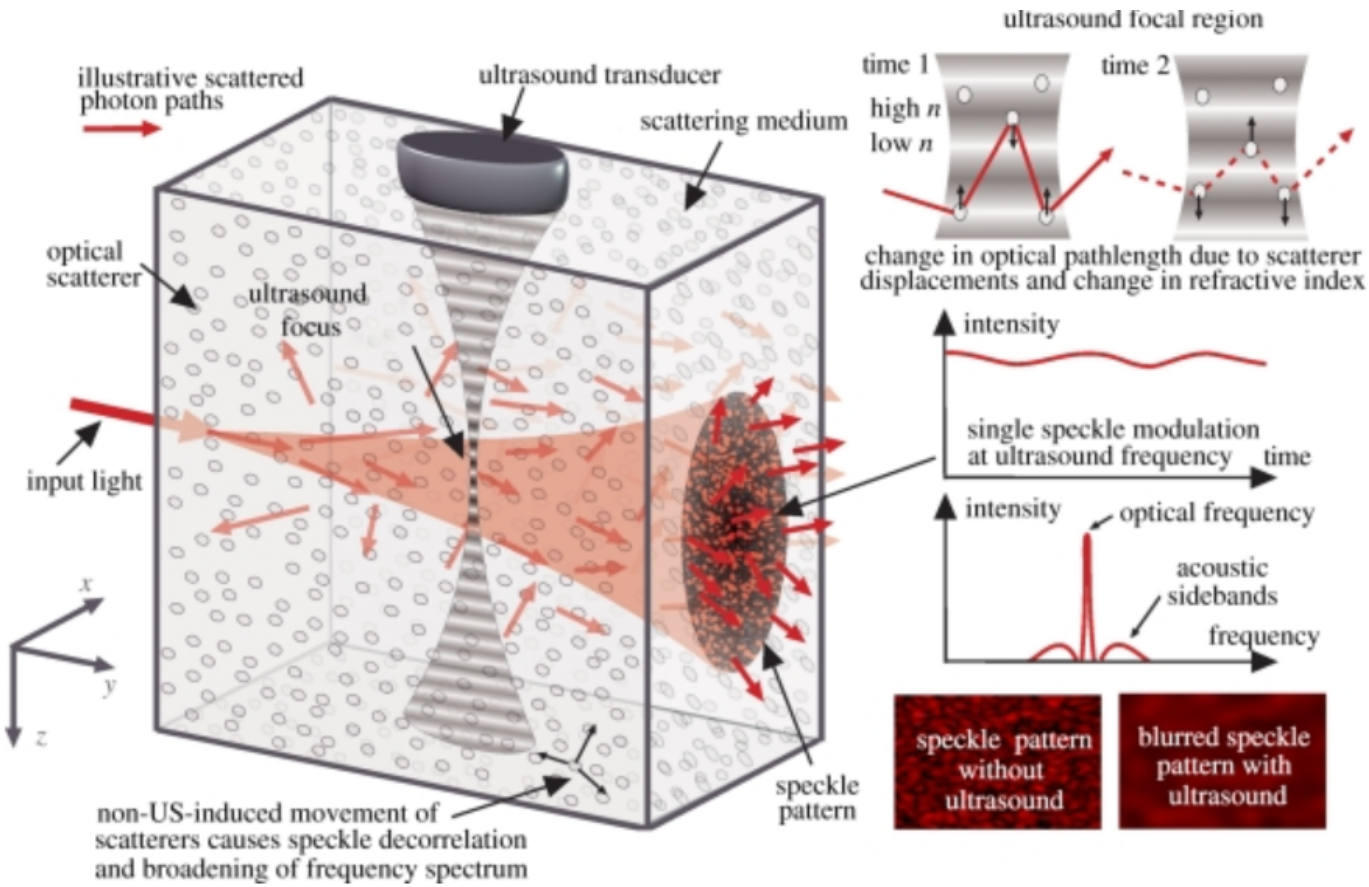


Figure: Schematic showing propagation of highly scattered photons through a biological tissue in the presence of a focused ultrasound beam. The resulting speckle pattern at the output face is illustrated together with the modulation in intensity of a single speckle grain. From : [Elson *et.al.*, Interface Focus, 2011]

Outline

The Radiative Transfer Equation (RTE) is a natural description of light considered as photons. It represents a balance equation where photons in a constant refractive index medium, in the absence of scattering, are propagated along rays $\boldsymbol{l} := \boldsymbol{r}_0 + l\hat{\boldsymbol{s}}$

$$\hat{\boldsymbol{s}} \cdot \nabla\phi + \mu_{\mathrm{a}}\phi = 0 \quad \equiv \quad \mathcal{T}_{\mu_{\mathrm{a}}}\phi = 0 \tag{1}$$

whose solution

$$\phi = \phi_0 \exp\left[-\int_{\boldsymbol{l}} \mu_{\mathrm{a}}(\boldsymbol{r}_0 + l\hat{\boldsymbol{s}})\mathrm{d}l\right] \tag{2}$$

is the basis for the definition of the *Ray Transform*

$$g_{\hat{\boldsymbol{s}}}(p) := -\ln\left[\frac{\phi}{\phi_0}\right] = \int_{-\infty}^{\infty} \mu_{\mathrm{a}}(p\hat{\boldsymbol{s}}_{\perp} + l\hat{\boldsymbol{s}})\mathrm{d}l \quad \equiv \quad g_{\hat{\boldsymbol{s}}} = \mathcal{R}_{\hat{\boldsymbol{s}}}\mu_{\mathrm{a}} \tag{3}$$

In the presence of scattering, and with source terms q, eq.(1) becomes

$$\begin{aligned}(\hat{\boldsymbol{s}}\cdot\nabla+\mu_{\mathrm{a}}(\boldsymbol{r})+\mu_{\mathrm{s}}(\boldsymbol{r}))\,\phi(\boldsymbol{r},\hat{\boldsymbol{s}}) &= \mu_{\mathrm{s}}\int_{S^{n-1}}\Theta(\hat{\boldsymbol{s}},\hat{\boldsymbol{s}}')\phi(\boldsymbol{r},\hat{\boldsymbol{s}}')d\hat{\boldsymbol{s}}'+q(\boldsymbol{r},\hat{\boldsymbol{s}})\\ \equiv \underbrace{[\mathcal{T}_{\mu_{\mathrm{tr}}}-\mu_{\mathrm{s}}\mathcal{S}]}_{\mathcal{L}}\phi &= q \end{aligned} \tag{4}$$

$\mu_{\mathrm{tr}}=\mu_{\mathrm{s}}+\mu_{\mathrm{a}}$ is the attenuation coefficent
$\mathcal{S}$ is the scattering operator, (local, non propagating).
Method of successive approximation (Sobolev 1963) :

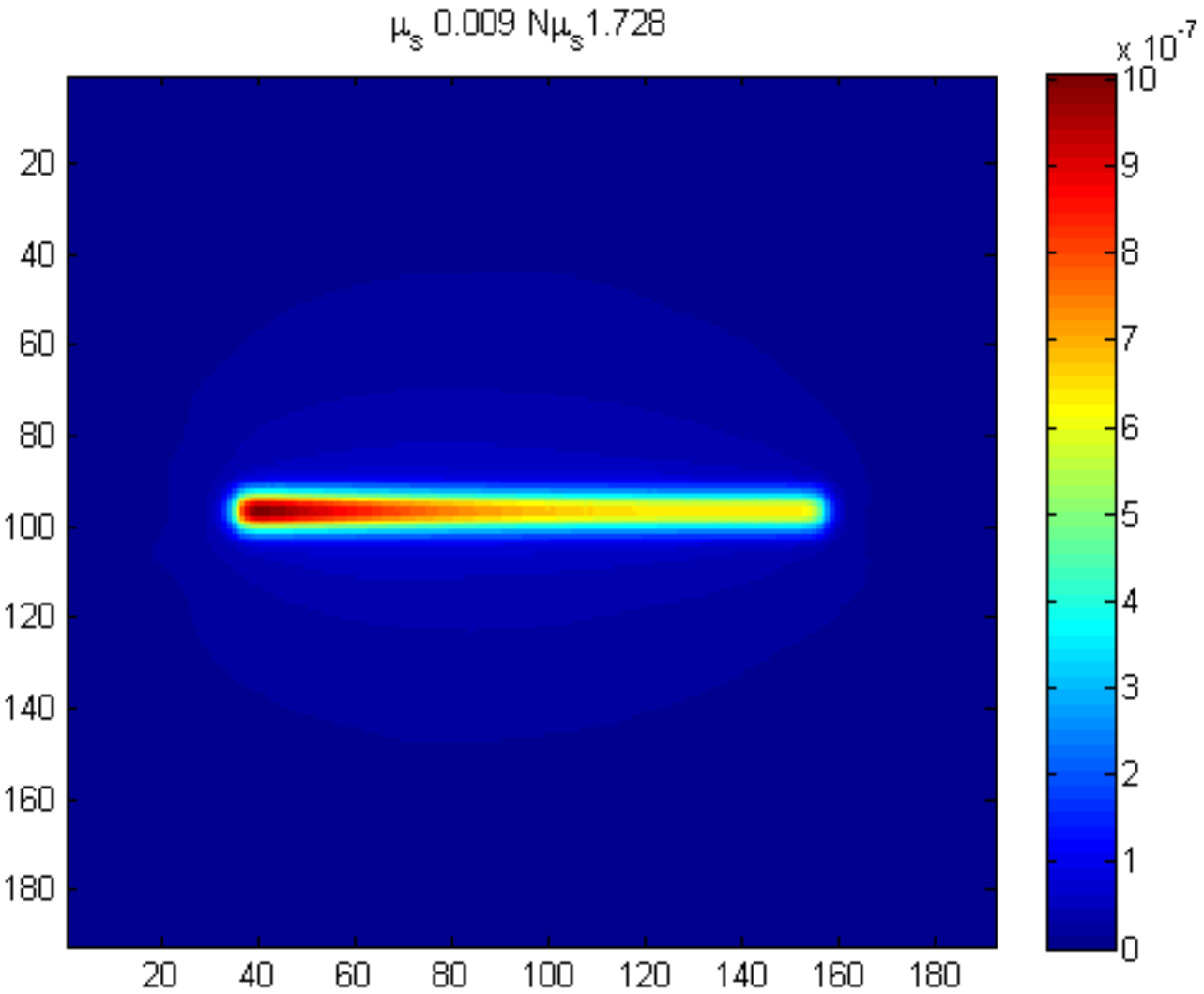
$$\phi=\left[\mathcal{T}_{\mu_{\mathrm{tr}}}^{-1}+\mathcal{T}_{\mu_{\mathrm{tr}}}^{-1}\mu_{\mathrm{s}}\mathcal{S}\mathcal{T}_{\mu_{\mathrm{tr}}}^{-1}+\ldots\left(\mathcal{T}_{\mu_{\mathrm{tr}}}^{-1}\mu_{\mathrm{s}}\mathcal{S}\right)^{k}\mathcal{T}_{\mu_{\mathrm{tr}}}^{-1}\ldots\right]q \tag{5}$$

The first term may be found from the Ray Transform, giving an alternative equation for the *collided flux*

$$[\mathcal{T}_{\mu_{\mathrm{tr}}}-\mu_{\mathrm{s}}\mathcal{S}]\,\phi_{\mathrm{collided}}=\mu_{\mathrm{s}}\mathcal{S}\underbrace{\mathcal{T}_{\mu_{\mathrm{tr}}}^{-1}q}_{\text{uncollided}} \tag{6}$$

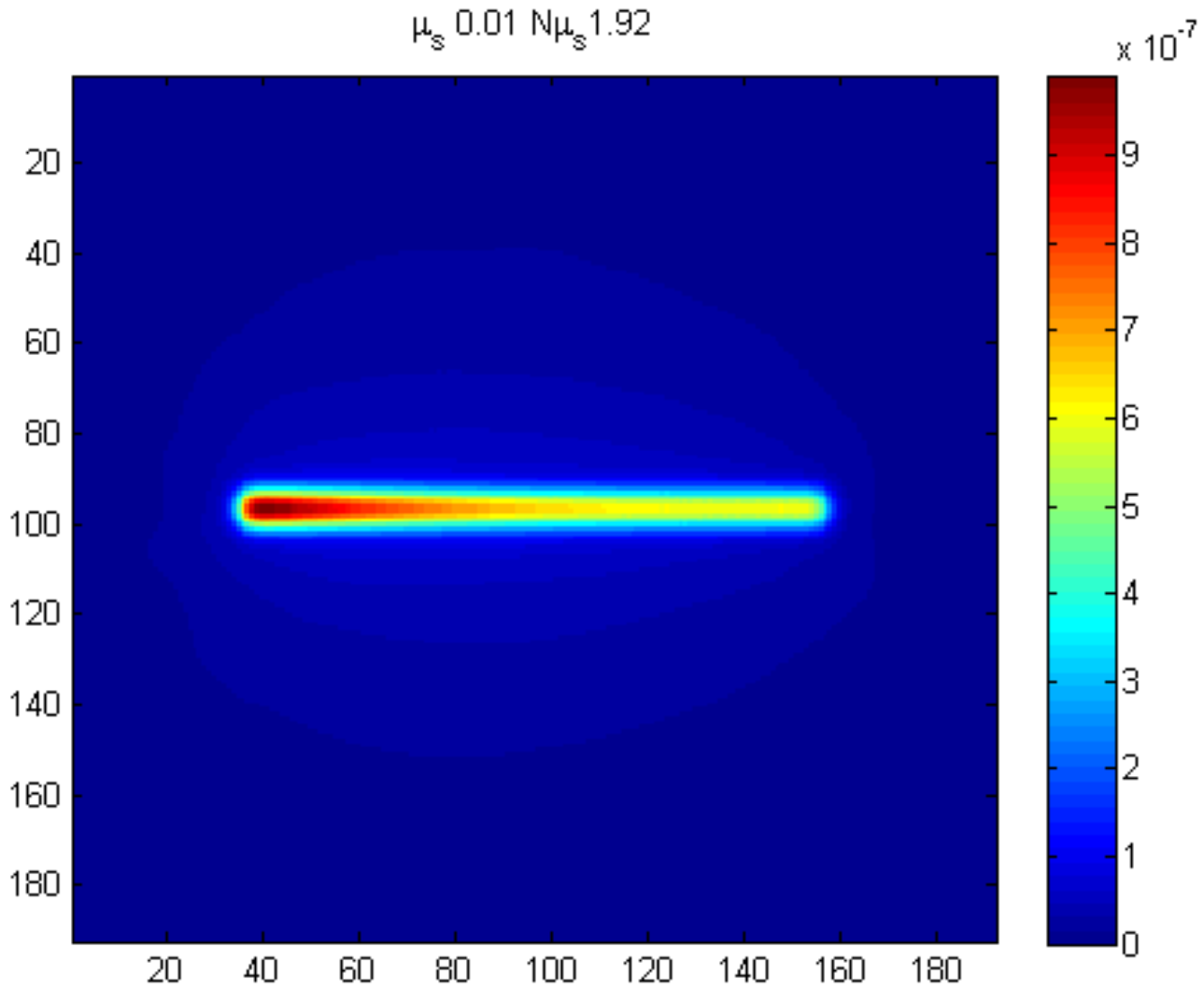
Modelling in Optical Tomography

RTE solutions



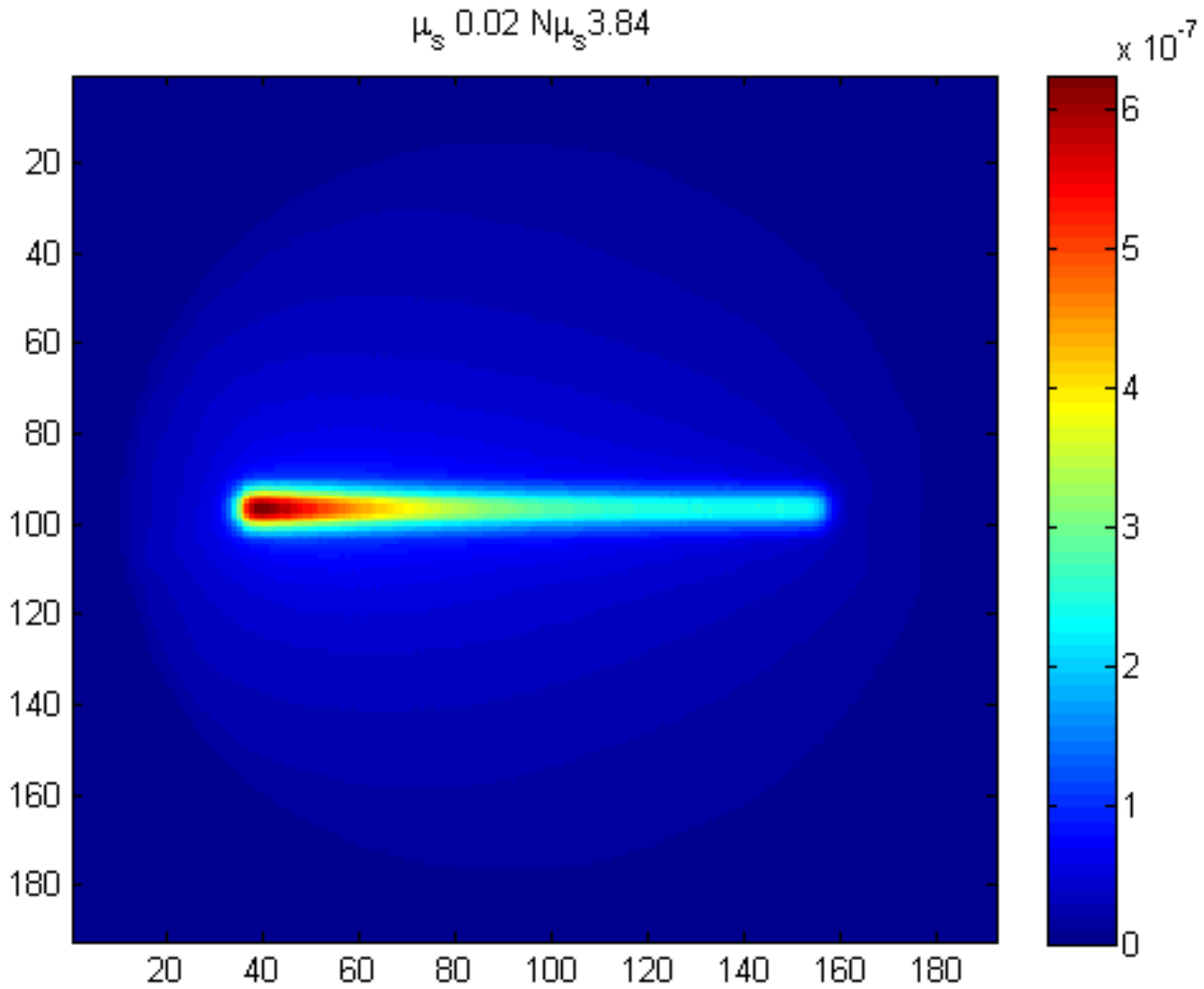
Modelling in Optical Tomography

RTE solutions



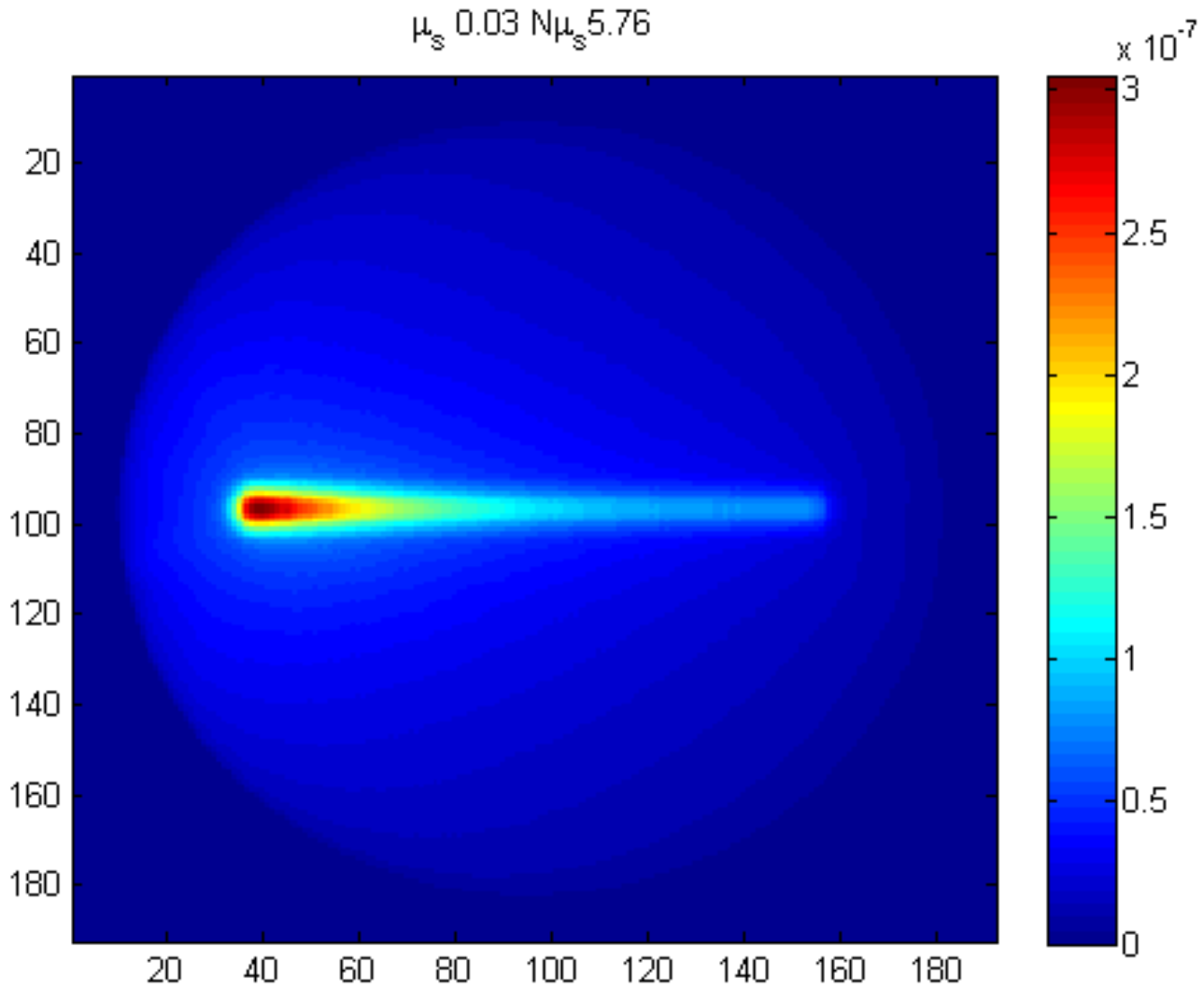
Modelling in Optical Tomography

RTE solutions



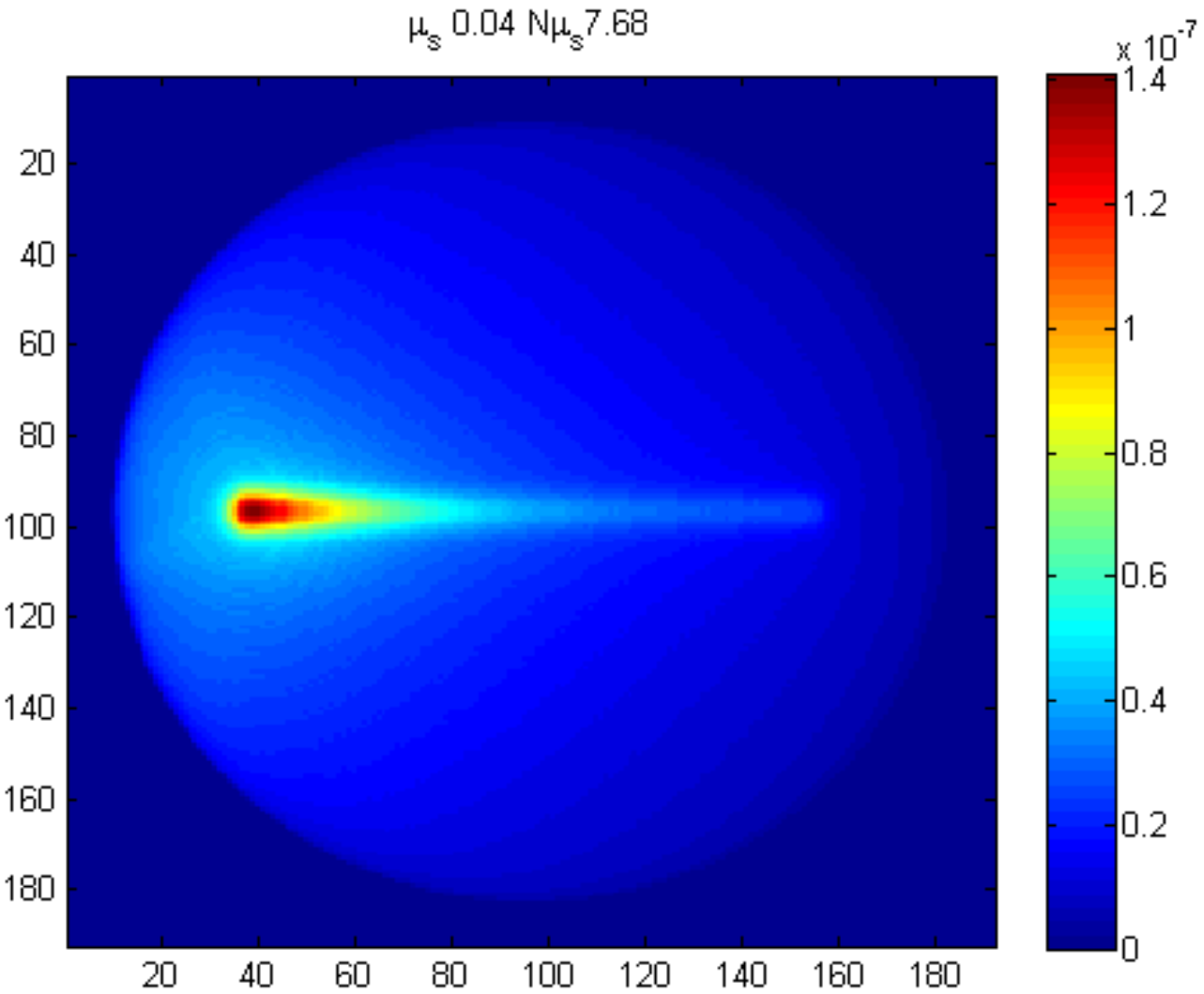
Modelling in Optical Tomography

RTE solutions



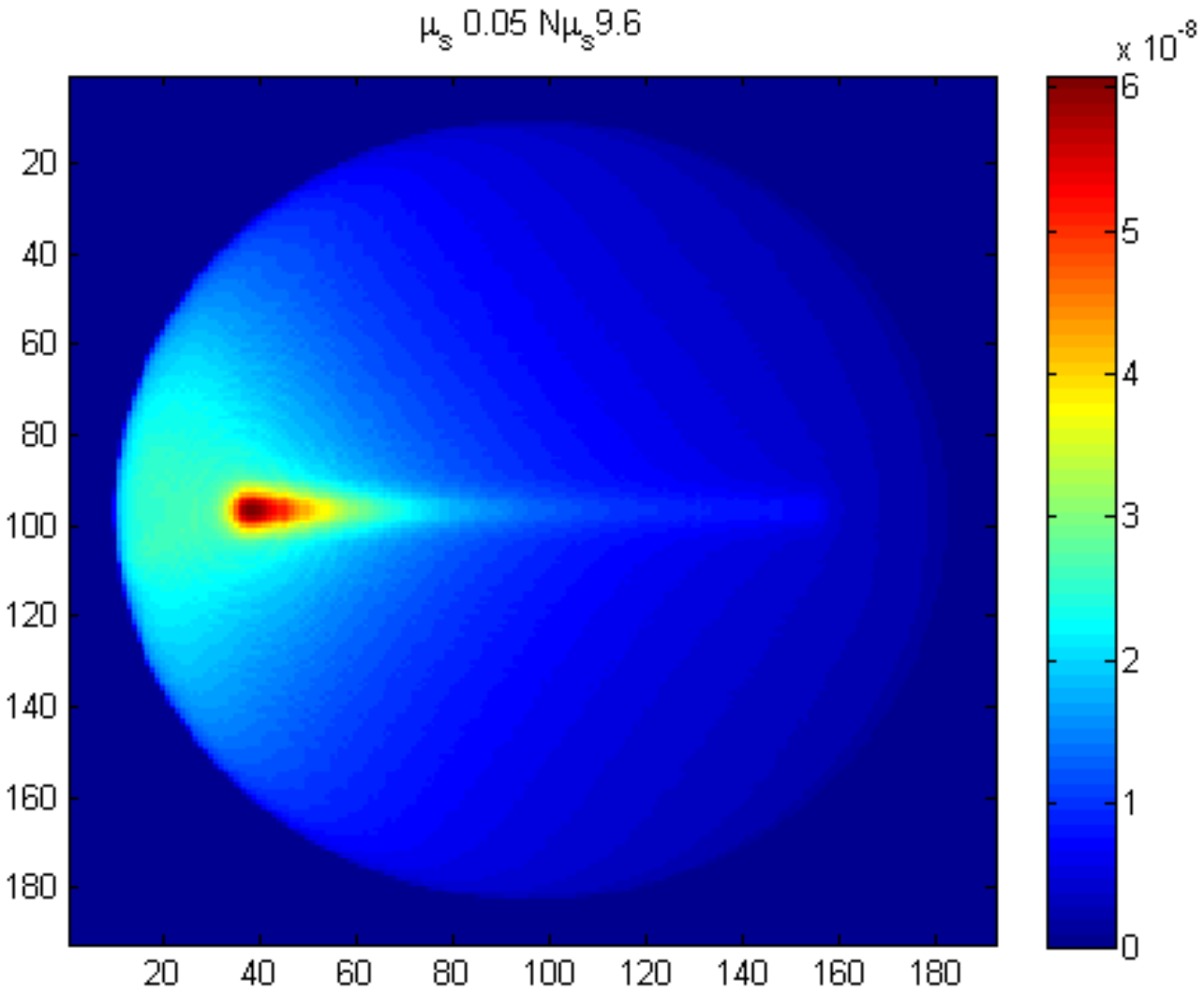
Modelling in Optical Tomography

RTE solutions



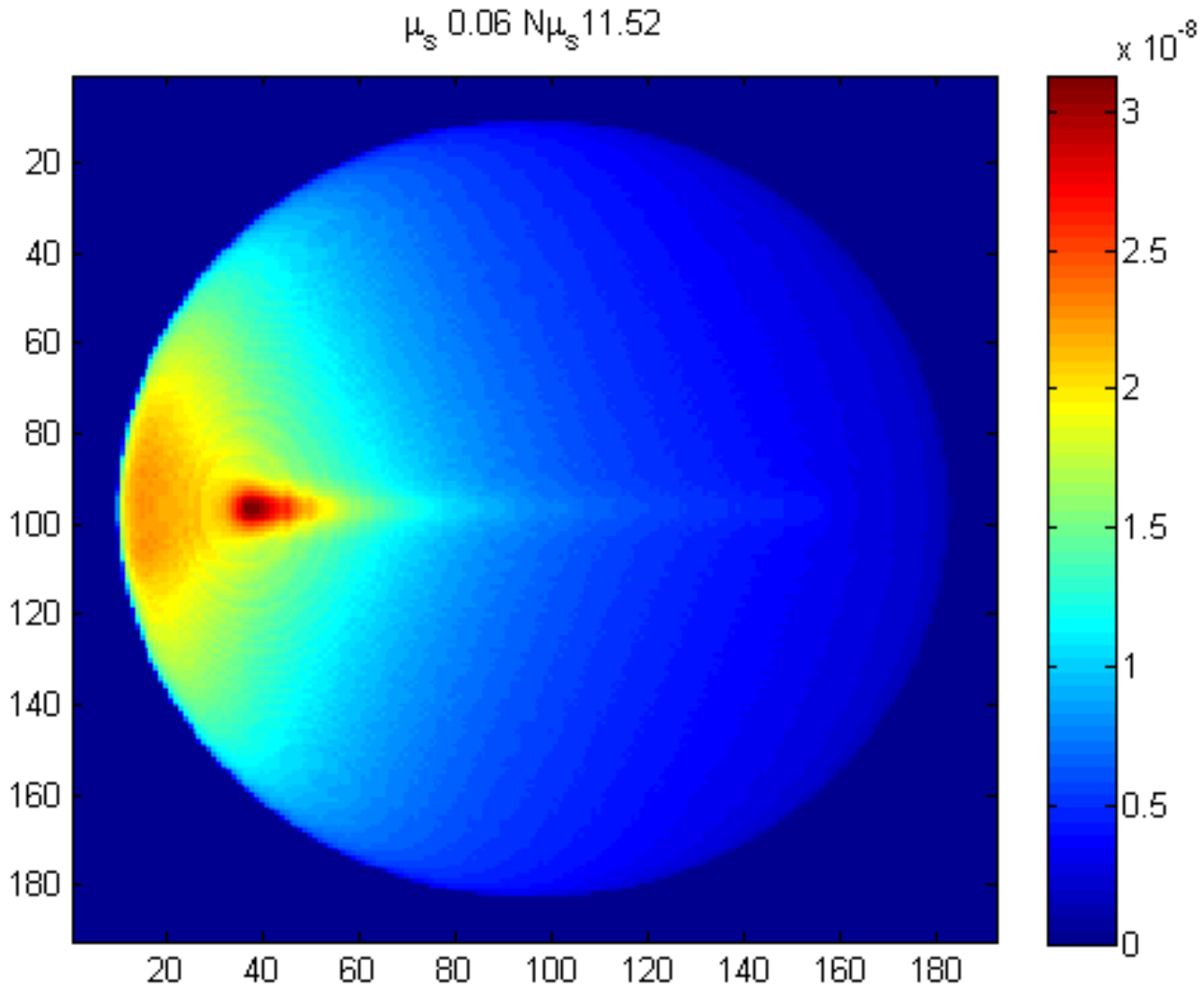
Modelling in Optical Tomography

RTE solutions



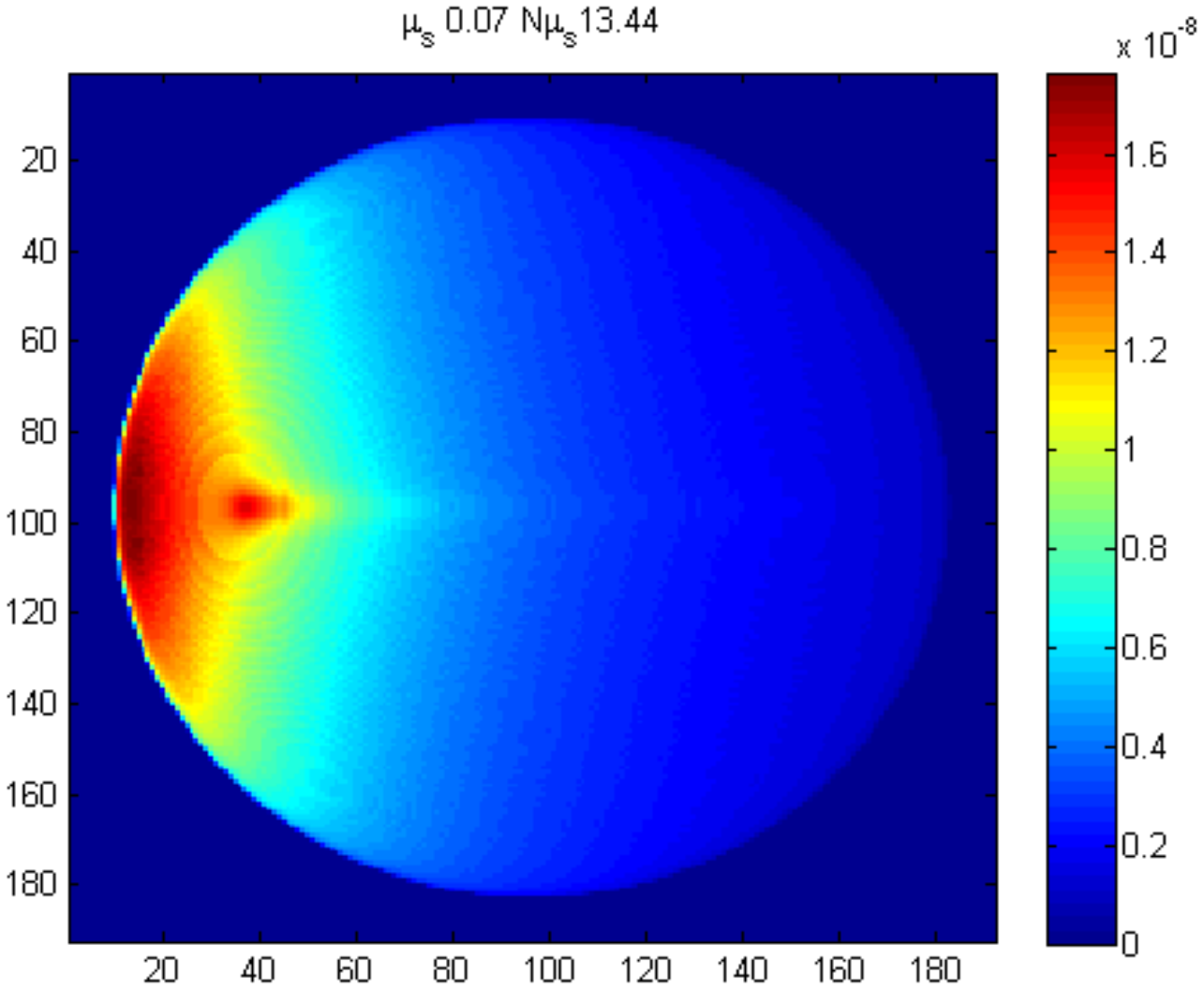
Modelling in Optical Tomography

RTE solutions



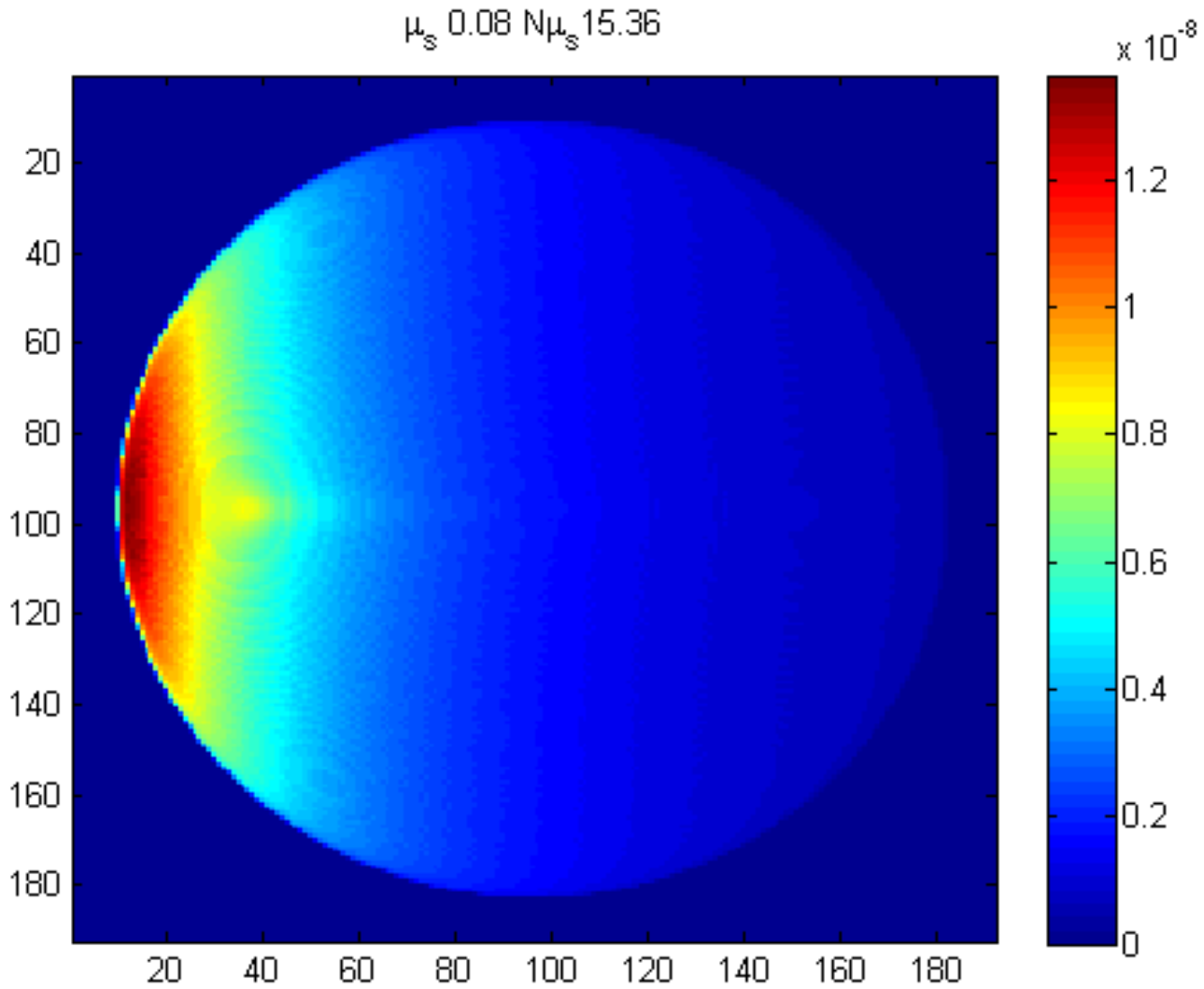
Modelling in Optical Tomography

RTE solutions



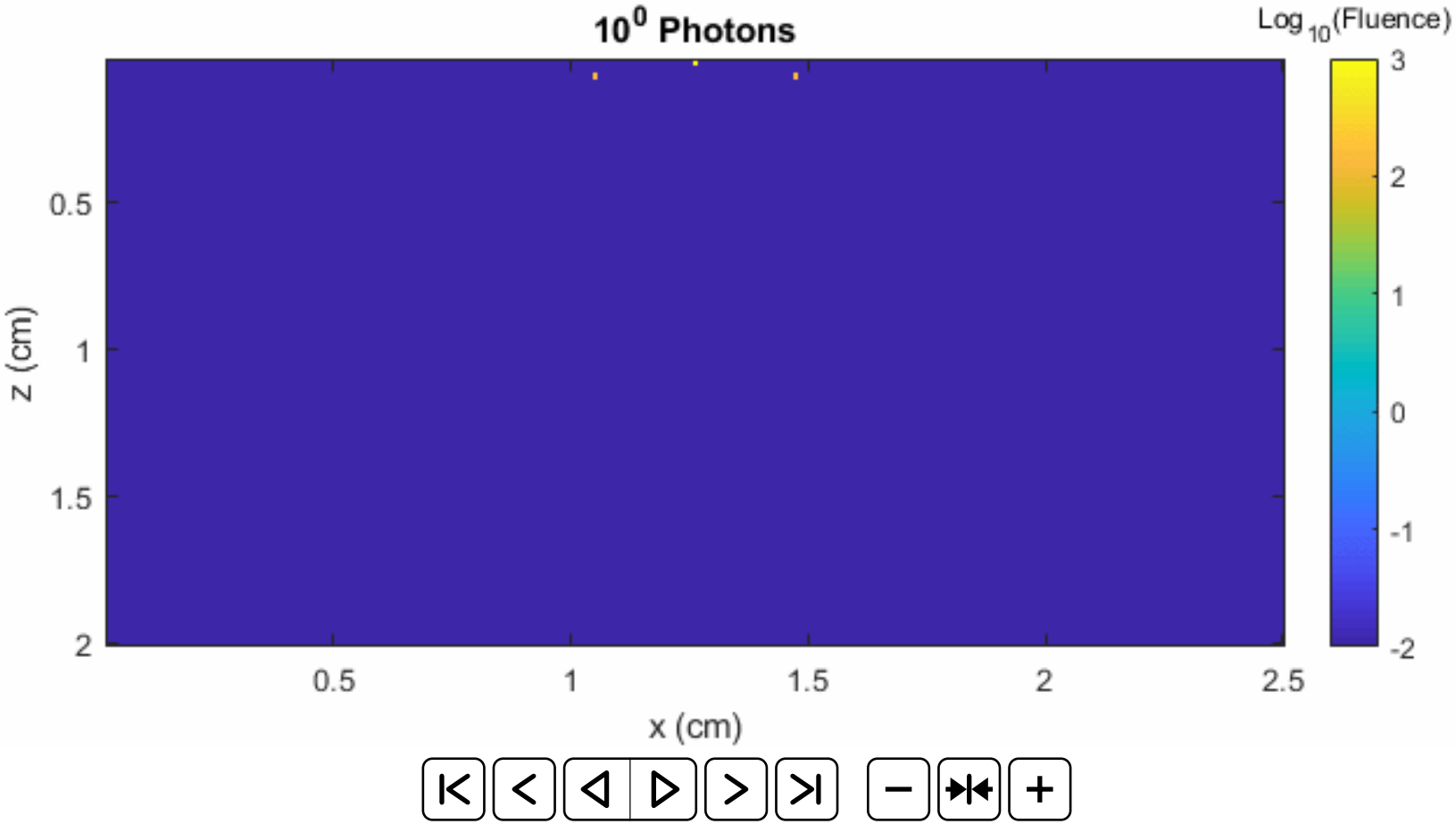
Modelling in Optical Tomography

RTE solutions



Modelling of Light

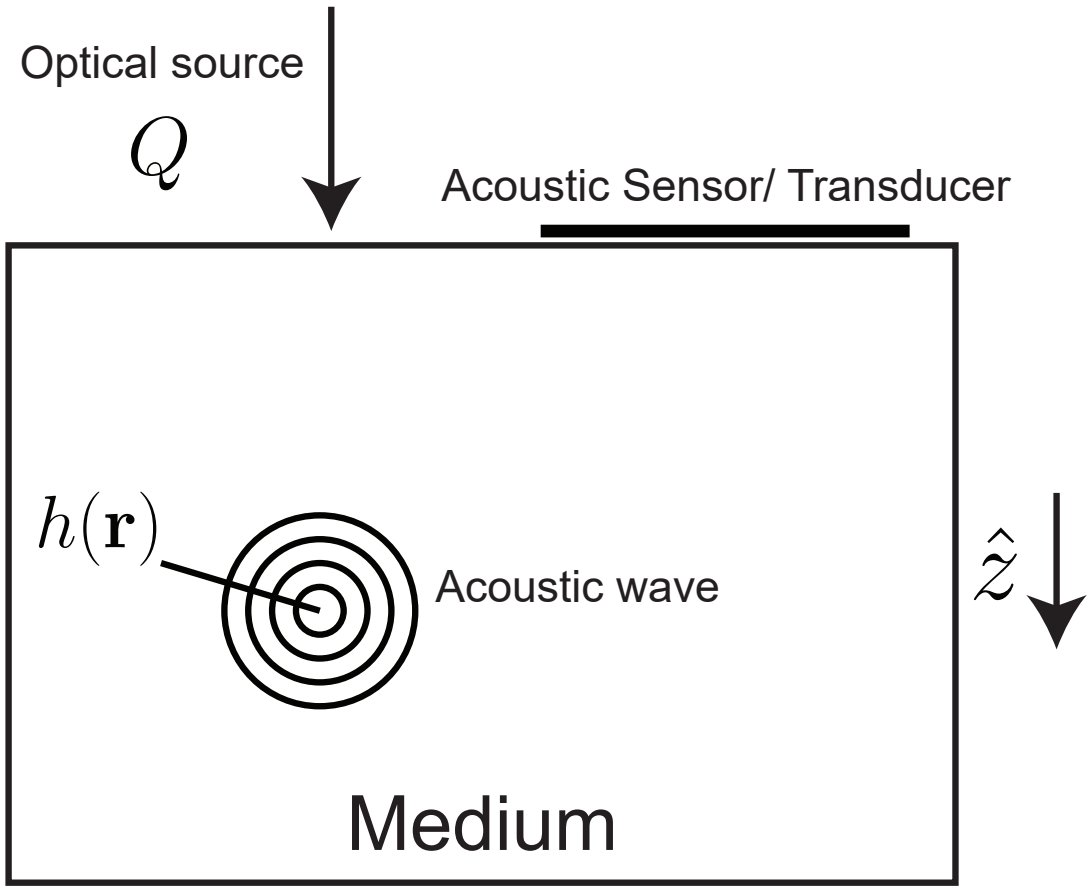
The Monte Carlo Approach



Outline

Deterministic Reconstruction based on RTE

Setup for Quantitative Photoacoustic Tomography



Deterministic Reconstruction based on RTE

QPAT case

In QPAT the internal pressure distribution is related to the spatial distribution of absorbed optical energy, h, where (omitting the Grüneisen parameter)

$$h(\boldsymbol{r}) = \mu_{\mathrm{a}}(\boldsymbol{r})\Phi(\boldsymbol{r}), \tag{7}$$

and where Φ is the optical fluence given by the angular integral of the radiance,

$$\Phi(\boldsymbol{r}) = \int_{\mathcal{S}^2} \phi(\boldsymbol{r}, \hat{\boldsymbol{s}})\, d\hat{\boldsymbol{s}}\,. \tag{8}$$

Assuming that we can recover the absorbed optical energy (through PAT)), h, the QPAT problem is to find the distribution of $\mu_{\mathrm{a}}(\boldsymbol{r})$ within the medium [Cox 2012, Saratoon2013].
Restating our cost function in terms of the QPAT data function, h, we have

$$F^{\mathrm{QPAT}} \quad = \quad \frac{1}{2}\int_{\Omega}(h^{\mathrm{obs}} - h)^2 \mathrm{d}r \quad = \quad \frac{1}{2}\left\langle h^{\mathrm{obs}} - h, h^{\mathrm{obs}} - h\right\rangle_{L^2(\Omega)}\,. \tag{9}$$

then the Fréchet derivative of F^{QPAT} is

$$DF^{\mathrm{QPAT}} = -\left\langle h^{\mathrm{obs}} - h, Dh\mu_{\mathrm{a}}^{\delta}\right\rangle_{L^2(\Omega)}\,, \tag{10}$$

where $\mu_{\mathrm{a}}^{\delta}$ is a small change in absorption.

Deterministic Reconstruction based on RTE

QPAT case

Writing the Fréchet derivative of h as

$$Dh = \Phi + \mu_a \cdot D\Phi \ , \tag{11}$$

and defining $\Phi^\delta = D\Phi\mu_a^\delta$, we arrive at

$$DF^{\text{QPAT}} = -\left\langle \Phi(h^{\text{obs}} - h), \mu_a^\delta \right\rangle_{L^2(\Omega)} - \left\langle \mu_a(h^{\text{obs}} - h), \Phi^\delta \right\rangle_{L^2(\Omega)} \ . \tag{12}$$

Next, we define the adjoint radiance, ϕ^*, as the solution to

$$\mathcal{L}^*\phi^* = \mu_a(h^{\text{obs}} - h) \tag{13}$$

where the right hand side describes the "adjoint source" which is isotropic in $\hat{\vec{s}}$. We then substitute the above into eq. 12 to give

$$DF^{\text{QPAT}} = -\left\langle \Phi(h^{\text{obs}} - h), \mu_a^\delta \right\rangle_{L^2(\Omega)} - \left\langle \mathcal{L}^*\phi^*, \phi^\delta \right\rangle_{L^2(\Omega \times S^{n-1})} \tag{14}$$

where we exploited the fact that the right hand side of eq. 13 does not depend on direction. Using the definition of the adjoint operator, and the fact that the change in radiance is zero on the boundary $\partial\Omega$ yields

$$DF^{\text{QPAT}} = -\left\langle \Phi(h^{\text{obs}} - h), \mu_a^\delta \right\rangle_{L^2(\Omega)} - \left\langle \phi^*, \mathcal{L}\phi^\delta \right\rangle_{L^2(\Omega \times S^{n-1})} \ . \tag{15}$$

Deterministic Reconstruction based on RTE

QPAT case

Consider a change to eq. 4 where $\mu_a \rightarrow \mu_a + \mu_a^\delta, \mu_s \rightarrow \mu_s + \mu_s^\delta$, for the same source Q, which results in a change in radiance $\phi \rightarrow \phi + \phi^\delta$. This implies

$$\begin{aligned}
\left(\mathcal{T}_{\mu_a+\mu_a^\delta,\mu_s+\mu_s^\delta} - \mathcal{S}_{\mu_s+\mu_s^\delta}\right)\left(\phi+\phi^\delta\right) &= \left(\mathcal{T}_{\mu_a,\mu_s} - \mathcal{S}_{\mu_s}\right)\phi \\
\Rightarrow \quad \left(\mathcal{T}_{\mu_a,\mu_s} - \mathcal{S}_{\mu_s}\right)\phi^\delta &= -(\mu_a^\delta + \mu_s^\delta + \mathcal{S}_{\mu_s^\delta})\phi \qquad (16) \\
\mathcal{L}_{\mu_a,\mu_s}\phi^\delta &= -\underbrace{(\mu_a^\delta + \mu_s^\delta + \mathcal{S}_{\mu_s^\delta})}_{\mathcal{L}^\delta_{\mu_a^\delta,\mu_s^\delta}}\phi\,. \qquad (17)
\end{aligned}$$

Neglecting scattering, this gives

$$DF^{\mathrm{QPAT}} = -\left\langle \Phi(h^{\mathrm{obs}} - h), \mu_a^\delta \right\rangle_{L^2(\Omega)} + \left\langle \phi^*\phi, \mu_a^\delta \right\rangle_{L^2(\Omega\times S^{n-1})}\,, \qquad (18)$$

allowing us to define the (absorption) gradient

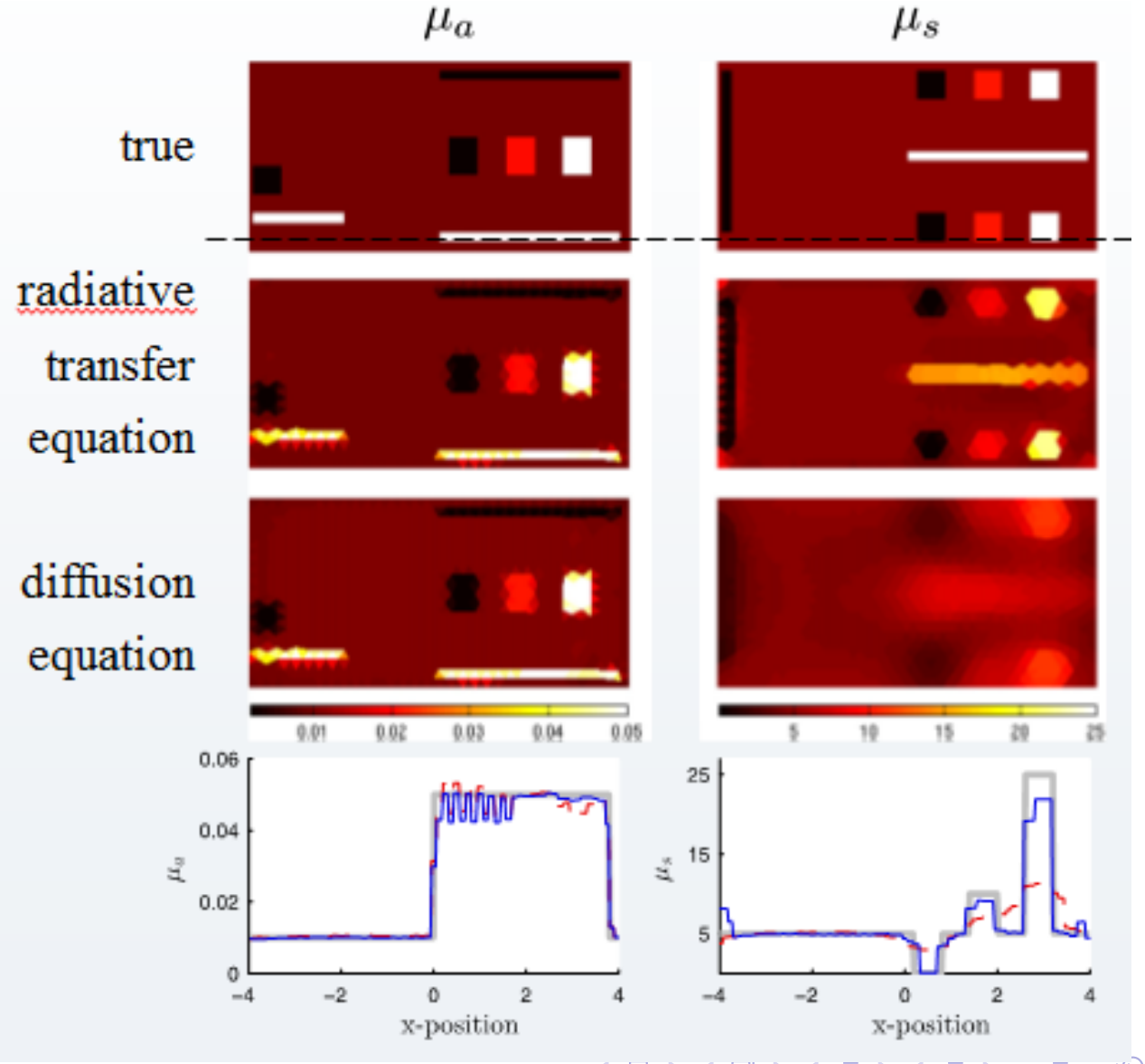
$$\frac{\partial F^{\mathrm{QPAT}}}{\partial \mu_a} = \nabla F^{\mathrm{QPAT}} = -\Phi(h^{\mathrm{obs}} - h) + \int_{S^{n-1}} \phi^*\phi \, \mathrm{d}\hat{\boldsymbol{s}} \qquad (19)$$

- Fixed-point iteration, known scattering (Yao, Sun, Jiang 2009)
- Use separated unscattered, singly-scattered and multiply scattered components (Bal, Jollivet, Jugnon 2010)
- Gauss-Newton inversions with TV regularization (Cox, Tarvainen, A., 2011; Tarvainen, Cox, Kaipio, A. 2012)
- Gradient-based inversions with Tikhonov reg. (Saratoon, Tarvainen, Cox, A., 2013)

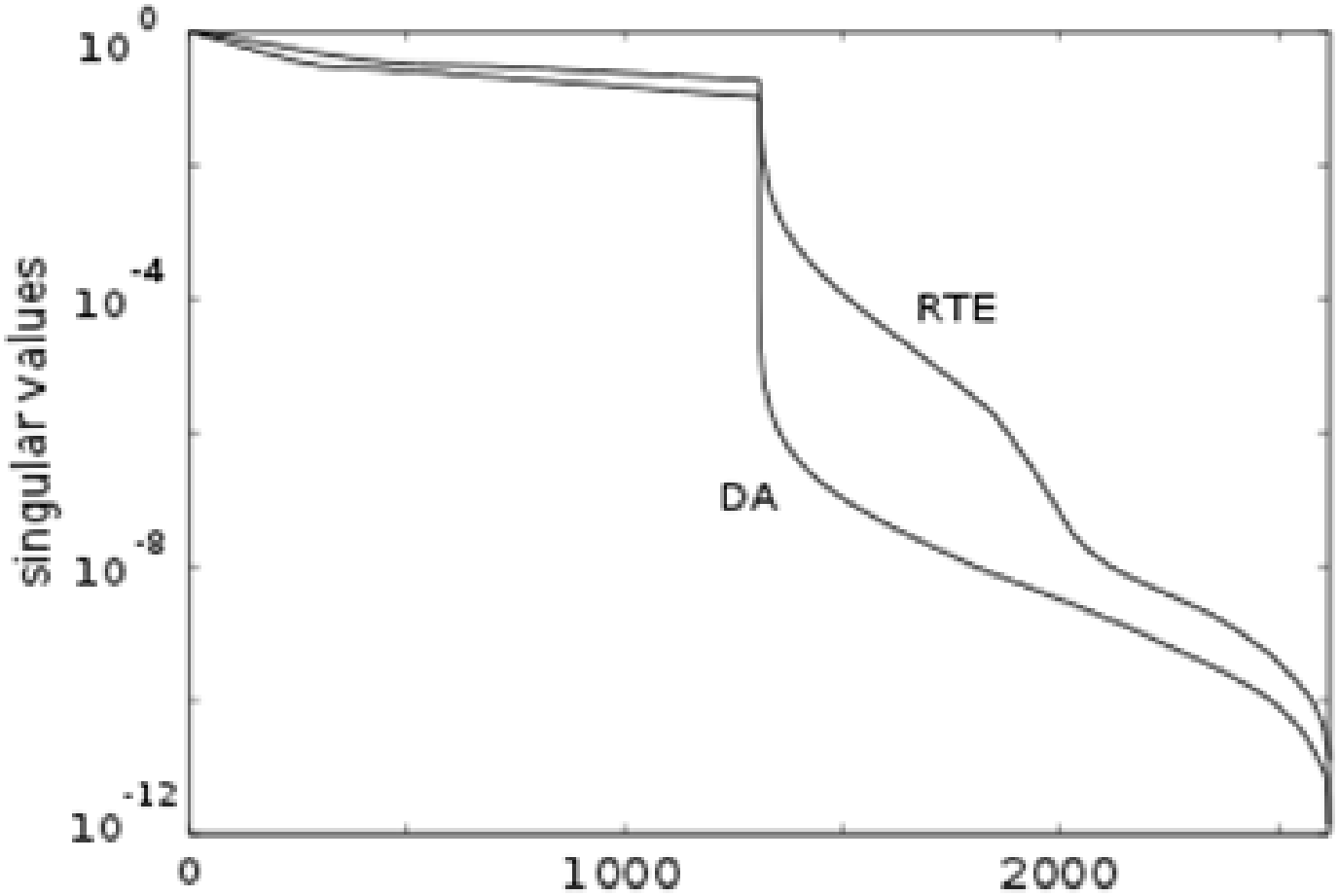
Quantitative PhotoAcoustic Tomography

RTE-based Inversions (Gauss-Newton)

Using 4 images from 4 illumination directions (Tarvainen, Cox., Kaipio, A. 2012)



SVD of Hessian reveals different information content

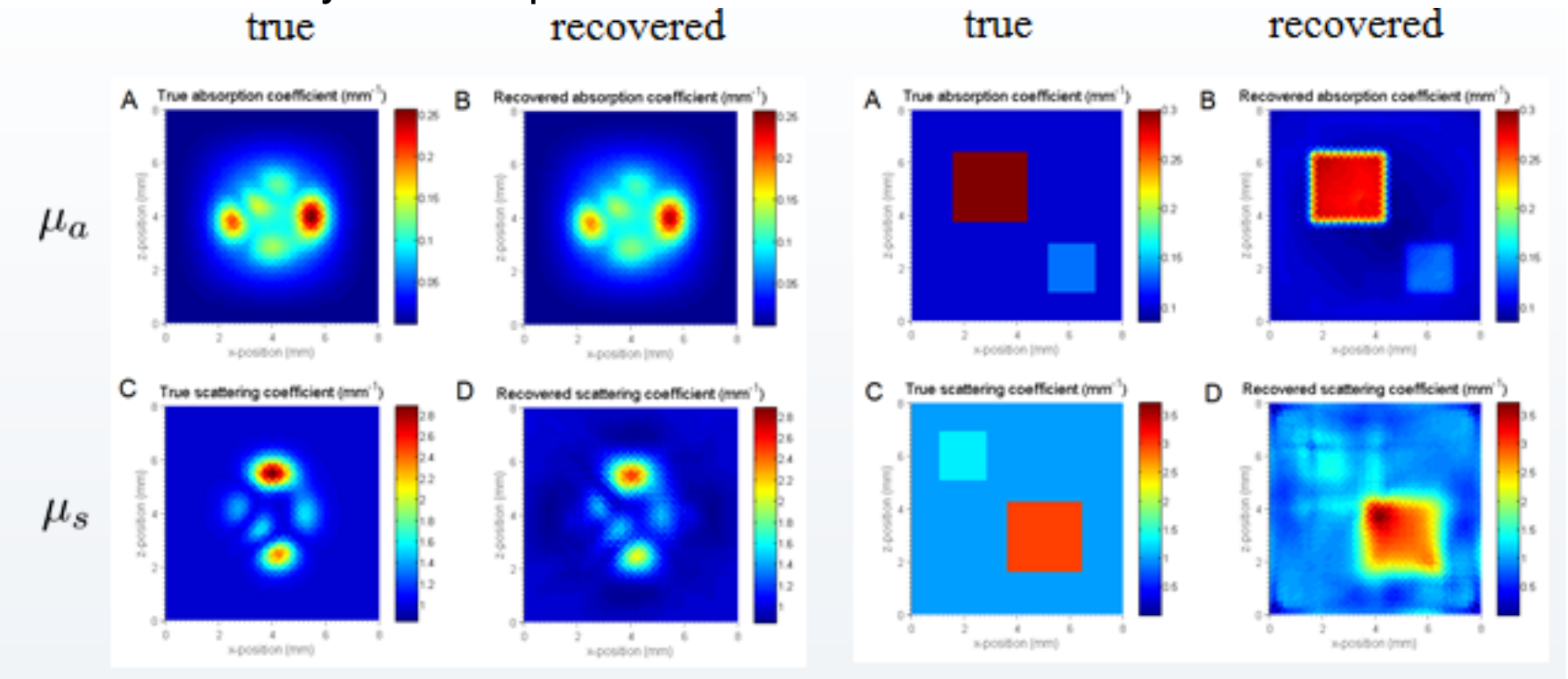


Quantitative PhotoAcoustic Tomography

Matrix Free method

Explicit construction of Jacobians is too expensive $\Rightarrow$ use matrix free method based on adjoint fields

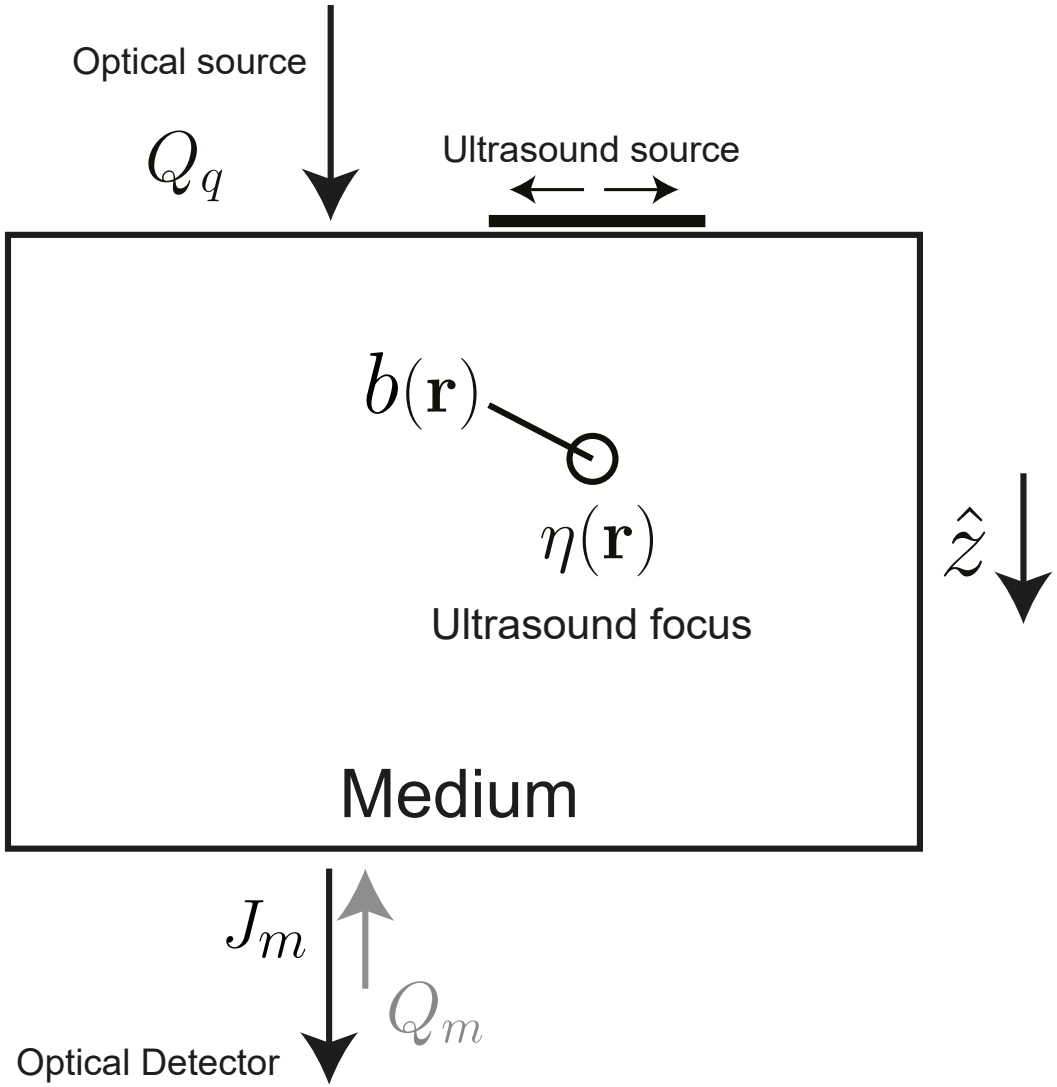
Limited memory BFGS optimisation



Using 4 images from 4 illumination directions, Tikhonov regularisation (Saratoon, Tarvainen, Cox, A., 2013)

Deterministic Reconstruction based on RTE

Setup for Ultrasound-Modulated Optical Tomography in the transmission geometry



Deterministic Reconstruction based on RTE

UMOT case

In UMOT we have an optical light source Q_q incident on a medium, as well as an optical detector J_m. In addition, an ultrasound source is incident on the medium, where the focus $\eta(\boldsymbol{r})$ is scanned through the sample [5]. Assuming for simplicity an ideal (delta-function) ultrasound focus, the data of interest in this case is found to be of the form[6]

$$b(r) = \eta(\boldsymbol{r})\Phi_q(\boldsymbol{r})\Phi_m(\boldsymbol{r}), \tag{20}$$

where Φ_q is the fluence resulting from the optical source Q_q, and Φ_m is the resulting fluence from a virtual source Q_m which is reciprocal to the detector J_m From this point we proceed in similar fashion as for QPAT, where now our data fitting error is given by

$$F^{\mathrm{UMOT}} = \frac{1}{2}\int_{\Omega}(b^{\mathrm{obs}} - b)^2 \mathrm{d}r = \frac{1}{2}\left\langle b^{\mathrm{obs}} - b, b^{\mathrm{obs}} - b\right\rangle_{L^2(\Omega)} \tag{21}$$

[5] Ammari 2014, Chung 2017]

[6] [Powell ,A, Leung, 2015]

Deterministic Reconstruction based on RTE

UMOT case

To get the Fréchet derivative

$$DF^{\mathrm{UMOT}} = -\left\langle b^{\mathrm{obs}} - b, Db\mu_{\mathrm{a}}^{\delta} \right\rangle_{L^2(\Omega)} \tag{22}$$

In this case the Fréchet derivative of b becomes

$$Db = \eta\Phi_q \cdot D\Phi_m + \eta\Phi_m \cdot D\Phi_q \tag{23}$$

leading to

$$DF^{\mathrm{UMOT}} = -\left\langle \eta\Phi_q(b^{\mathrm{obs}} - b), \Phi_m^{\delta} \right\rangle_{L^2(\Omega)} - \left\langle \eta\Phi_m(h^{\mathrm{obs}} - h), \Phi_q^{\delta} \right\rangle_{L^2(\Omega)} . \tag{24}$$

Here we need to define two adjoint radiances, $\phi^{*,1}$, $\phi^{*,2}$, as the solution to

$$\mathcal{L}^*\phi^{*,1} = \eta\Phi_q(b^{\mathrm{obs}} - b) \tag{25}$$

$$\mathcal{L}^*\phi^{*,2} = \eta\Phi_m(b^{\mathrm{obs}} - b) \tag{26}$$

and substituting into eq. 24 to give

$$DF^{\mathrm{UMOT}} = -\left\langle \mathcal{L}^*\phi^{*,1}, \phi_m^{\delta} \right\rangle_{L^2(\Omega\times S^{n-1})} - \left\langle \mathcal{L}^*\phi^{*,2}, \phi_q^{\delta} \right\rangle_{L^2(\Omega\times S^{n-1})} \tag{27}$$

Deterministic Reconstruction based on RTE

UMOT case

by the same arguments as for QPAT we get

$$DF^{\mathrm{UMOT}} = -\left\langle \phi^{*,1}, \mathcal{L}\phi_m^{\delta} \right\rangle_{L^2(\Omega\times S^{n-1})} - \left\langle \phi^{*,2}, \mathcal{L}\phi_q^{\delta} \right\rangle_{L^2(\Omega\times S^{n-1})} . \tag{28}$$

Again using the perturbation expression eq. 17 we have

$$DF^{\mathrm{UMOT}} = \left\langle \phi^{*,1}\phi_m, \mu_{\mathrm{a}}^{\delta} \right\rangle_{L^2(\Omega\times S^{n-1})} + \left\langle \phi^{*,2}\phi_q, \mu_{\mathrm{a}}^{\delta} \right\rangle_{L^2(\Omega\times S^{n-1})} . \tag{29}$$

allowing us to define the (absorption) gradient as

$$\frac{\partial F^{\mathrm{UMOT}}}{\partial \mu_{\mathrm{a}}} = \nabla F^{\mathrm{UMOT}} = \int_{S^{n-1}} \left(\phi^{*,1}\phi_m + \phi^{*,2}\phi_q\right) \mathrm{d}\hat{\boldsymbol{s}} \tag{30}$$

Deterministic Reconstruction UMOT

UMOT fields

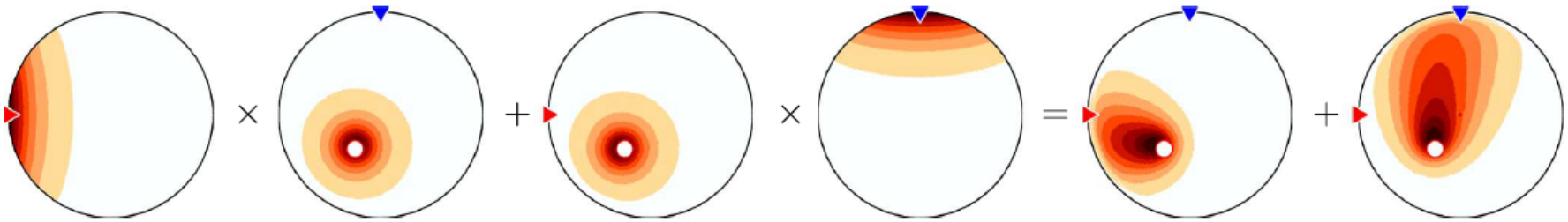


Figure: Pictorial description of eq.30. From left to right: ϕ_q, $\phi^{*,2}$, $\phi^{*,1}$, ϕ_m, $\phi_1 \times \phi^{*,2}$, $\phi_m \times \phi^{*,1}$. Red inwards arrows indicates real source positions, blue inwards arrows indicate adjoint source positions. White dots indicates ultrasound field focal point. The colour scale varies between plots.

Deterministic Reconstruction UMOT

UMOT sensitivities

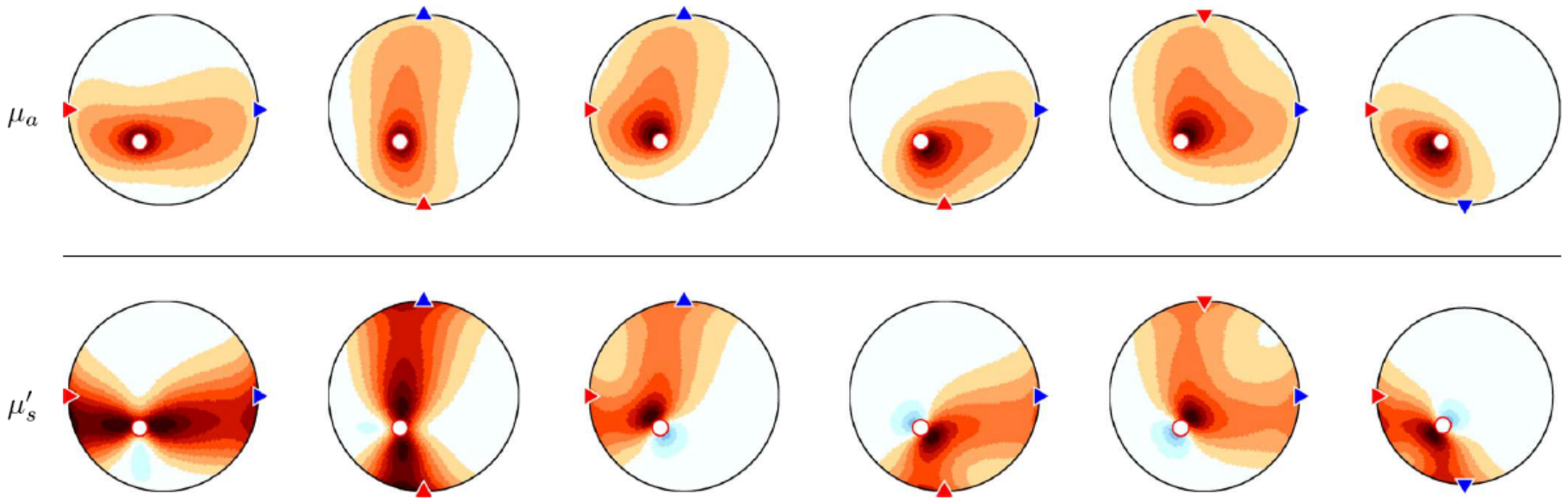


Figure: A set of CMDFs in μ_a (top) and μ_s (bottom) for all optical source-detector pairs one (left) to six (right). Red arrows indicate source locations, blue arrows indicate detector locations, white dots indicate the focal point of the acoustic field. Red regions indicate that increases in perturbations cause reductions in the measured data, blue regions indicate the converse. The colour scale vary between plots.

Deterministic Reconstruction UMOT

UMOT reconstruction (2D)

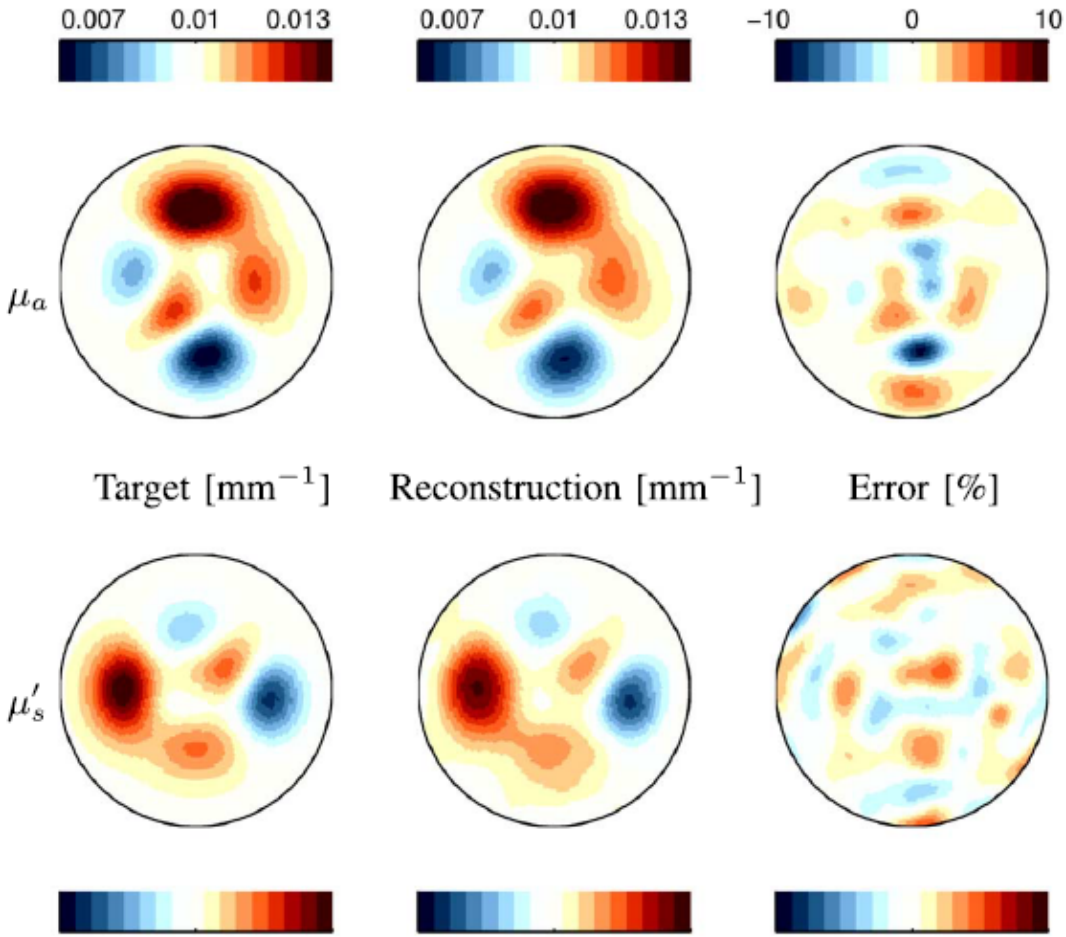


Figure: Target (left), reconstruction (middle) and percentage error (right) images of μ_a (top) and μ_s (bottom) for two-dimensional reconstruction.

Outline

Fully Stochastic Reconstruction (FSR)

FSR : Introduction

- Solving quantitative coupled optical imaging requires an accurate forward model

Fully Stochastic Reconstruction (FSR)

FSR : Introduction

- Solving quantitative coupled optical imaging requires an accurate forward model
- Deterministic RTE is computationally expensive and needs adapting to problem domain

Fully Stochastic Reconstruction (FSR)

FSR : Introduction

- Solving quantitative coupled optical imaging requires an accurate forward model
- Deterministic RTE is computationally expensive and needs adapting to problem domain
- Stochastic Monte Carlo modelling is "arbitrarily accurate" *in expectation* but not a operator between Banach spaces

Fully Stochastic Reconstruction (FSR)

FSR : Introduction

- Solving quantitative coupled optical imaging requires an accurate forward model
- Deterministic RTE is computationally expensive and needs adapting to problem domain
- Stochastic Monte Carlo modelling is "arbitrarily accurate" *in expectation* but not a operator between Banach spaces
- One possible approach is to run forward and adjoint MC to "sufficient" accuracy and use as a proxy for deterministic RTE [Hochuli, Powell, A, Cox 2016]

Fully Stochastic Reconstruction (FSR)

FSR : Introduction

- Solving quantitative coupled optical imaging requires an accurate forward model
- Deterministic RTE is computationally expensive and needs adapting to problem domain
- Stochastic Monte Carlo modelling is "arbitrarily accurate" *in expectation* but not a operator between Banach spaces
- One possible approach is to run forward and adjoint MC to "sufficient" accuracy and use as a proxy for deterministic RTE [Hochuli, Powell, A, Cox 2016]
- New idea : use few-photons MC as approximate (inaccurate) model and leverage methods from *Adaptive Stochastic Gradient Descent* (ASGD) [Bollapragada, Byrd, Nocedal 2018].

- *Deterministic model* : gradient descent method ("Batch Gradient Descent"), with step size α_n ("training rate")
$$x_n = x_{n-1} - \alpha_n \nabla F(x_{n-1}) \ ,$$
Converges if $\lim_{n\to\infty} F(x_n) = 0$.

FSR : Introduction

Stochastic Gradient Descent

- *Deterministic model* : gradient descent method ("Batch Gradient Descent"), with step size α_n ("training rate")

$$x_n = x_{n-1} - \alpha_n \nabla F(x_{n-1}) \ ,$$

  Converges if $\lim_{n\to\infty} F(x_n) = 0$.

- *Stochastic setting* : true cost F and gradient ∇F not directly available $\Rightarrow$ use (unbaised) *estimates* of the cost function and gradient

$$\mathbb{E}[F_{S_n}(x_n)] = F(x_n) \ , \qquad \mathbb{E}[\nabla F_{S_n}(x_n)] = \nabla F(x_n) \ ,$$

  Here S_n denotes the n^{th} "sample" used in the computation.

- *Deterministic model* : gradient descent method ("Batch Gradient Descent"), with step size α_n ("training rate")
$$x_n = x_{n-1} - \alpha_n \nabla F(x_{n-1}) \ ,$$
Converges if $\lim_{n\to\infty} F(x_n) = 0$.
- *Stochastic setting* : true cost F and gradient ∇F not directly available $\Rightarrow$ use (unbaised) *estimates* of the cost function and gradient
$$\mathbb{E}[F_{S_n}(x_n)] = F(x_n) \ , \qquad \mathbb{E}[\nabla F_{S_n}(x_n)] = \nabla F(x_n) \ ,$$
Here S_n denotes the n^{th} "sample" used in the computation.
- In Monte Carlo modelling of radiative transport, the sample refers to the set of virtual photons (and their associated random number seeds) that are initiated in the simulation to represent an optical source, which are subsequently used to estimate $F(x_n)$ and $\nabla F(x_n)$.

- The stochastic version of Gradient Descent (SGD) thus attempts to minimize a *sampled* objective function, F_{S_n}, by updating the previous iterate with a scaled *sampled* gradient,

$$x_n = x_{n-1} - \alpha_n \nabla F_{S_n}(x_{n-1}) \; .$$

- The stochastic version of Gradient Descent (SGD) thus attempts to minimize a *sampled* objective function, F_{S_n}, by updating the previous iterate with a scaled *sampled* gradient,

$$x_n = x_{n-1} - \alpha_n \nabla F_{S_n}(x_{n-1}) \ .$$

- If α_n is fixed for all n, eventually there will come a point where the next update of the estimate (with the term $\alpha_n \nabla F_{S_n}(x_{n-1})$) will reliably "undo" the work of the prior step, which will effectively halt the descent. The point at which this occurs depends on the variance of ∇F_{S_n}. We can see this by re-writing the sampled stochastic gradient estimate as,

$$\nabla F_{S_n}(x_n) = \nabla F(x_n) + \epsilon_{S_n}(x_n) \ ,$$

where ϵ is a random vector with $\mathbb{E}[\epsilon_{S_n}(x_n)] = 0$ for all n.

FSR : Introduction

Stochastic Gradient Descent

- To prevent iteration gradient steps becoming comparable to a random walk, we may:
  - i) reduce the step size at each iteration such that we can avoid "backtracking" in the descent, or
  - ii) gradually improve the accuracy of our sampled gradient such that the variance of the sampled gradient remains below some threshold value compared to the norm of the true gradient ∇F.
- Second point tries to ensure the inequality

$$\textbf{norm test} \quad V^2_{\text{tot}}(x_n) := \frac{\mathbb{E}\left[|\epsilon_{S_n}(x_n)|^2\right]}{|\nabla F(x_n)|^2} \leq \gamma^2_{\text{tot}} \,, \quad \gamma_{\text{tot}} > 0.$$

where γ_{tot} is a positive coefficient describing the acceptable threshold.

FSR : Introduction

Stochastic Gradient Descent

- To prevent iteration gradient steps becoming comparable to a random walk, we may:
  - i) reduce the step size at each iteration such that we can avoid "backtracking" in the descent, or
  - ii) gradually improve the accuracy of our sampled gradient such that the variance of the sampled gradient remains below some threshold value compared to the norm of the true gradient ∇F.
- Second point tries to ensure the inequality

$$\textbf{norm test} \quad V^2_{\text{tot}}(x_n) := \frac{\mathbb{E}\left[|\epsilon_{S_n}(x_n)|^2\right]}{|\nabla F(x_n)|^2} \leq \gamma^2_{\text{tot}}\,, \quad \gamma_{\text{tot}} > 0.$$

  where γ_{tot} is a positive coefficient describing the acceptable threshold.
- Alternatively restrict the component of variance in the sampled gradient parallel to the true gradient ∇F,

$$\textbf{inner product test} \quad V^2_{\|}(x_n) := \frac{\mathbb{E}\left[\langle \epsilon_{S_n}(x_n), \nabla F(x_n)\rangle^2\right]}{|\nabla F(x_n)|^4} \leq \gamma^2_{\|}\,, \quad \gamma_{\|} > 0.$$

FSR : Introduction

Adaptive Sample Size

- Increasing the sample size in the event where the inner product and/or norm tests fail can be done in a number of ways. A simple method is to scale the current sample size by some factor $\kappa(n)$, to increase the number of photons used in the next iteration,

$$|S_{n+1}| = \kappa(n)\,|S_n|$$

FSR : Introduction

Adaptive Sample Size

- Increasing the sample size in the event where the inner product and/or norm tests fail can be done in a number of ways. A simple method is to scale the current sample size by some factor $\kappa(n)$, to increase the number of photons used in the next iteration,

$$|S_{n+1}| = \kappa(n)\,|S_n|$$

- One option for $\kappa(n)$ is to use the same factor by which the variance exceeds our imposed limit at a given point in the descent. For instance, upon failure of the inner product test for a chosen value of $\gamma_{\|}$, we can increase the sample size on the next iteration using $\kappa(n) = V_{\|}^2(x_n)/\gamma_{\|}^2$.

FSR : Introduction

Adaptive Sample Size

- Increasing the sample size in the event where the inner product and/or norm tests fail can be done in a number of ways. A simple method is to scale the current sample size by some factor $\kappa(n)$, to increase the number of photons used in the next iteration,

$$|S_{n+1}| = \kappa(n)\,|S_n|$$

- One option for $\kappa(n)$ is to use the same factor by which the variance exceeds our imposed limit at a given point in the descent. For instance, upon failure of the inner product test for a chosen value of $\gamma_{\|}$, we can increase the sample size on the next iteration using $\kappa(n) = V_{\|}^2(x_n)/\gamma_{\|}^2$.
- However, we also investigate other forms of $\kappa(n)$ in the results, which better cope with statistical variations that can lead to over-estimating the required sample size increase.

FSR : Introduction

Adaptive Step Size

- If we are bounding the error in the sampled gradient, e.g. by increasing the sample size, then fixed step SGD may converge so long as the following is satisfied for all n [Bollapragada, Byrd, Nocedal 2018].

$$\alpha_n \leq \frac{1}{(1+\gamma_{\text{tot}}^2)L} \quad ,$$

where L is the Lipschitz constant for F. This has to be estimated for a stochastic forward model such as Monte Carlo

FSR : Introduction

Adaptive Step Size

- If we are bounding the error in the sampled gradient, e.g. by increasing the sample size, then fixed step SGD may converge so long as the following is satisfied for all n [Bollapragada, Byrd, Nocedal 2018].

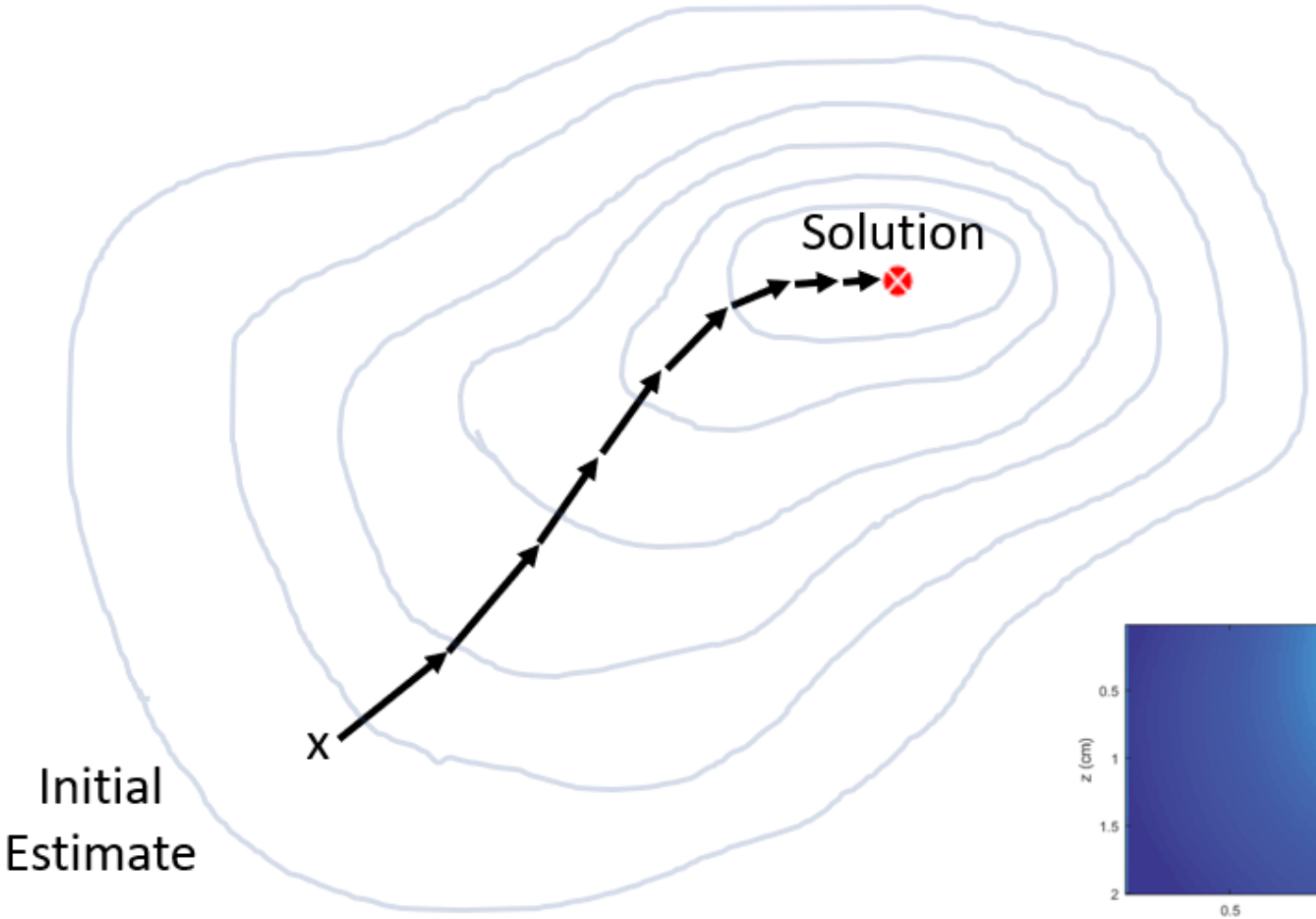
$$\alpha_n \leq \frac{1}{(1+\gamma_{\text{tot}}^2)L} \quad ,$$

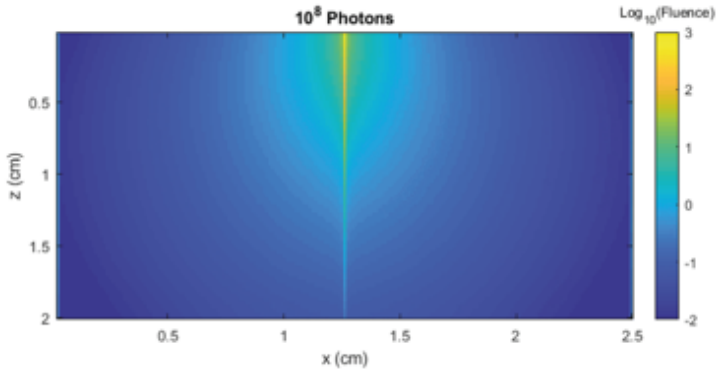
where L is the Lipschitz constant for F. This has to be estimated for a stochastic forward model such as Monte Carlo

- As intuition indicates, when the sample size (e.g number of simulated photons) increases towards the maximum number of samples $|S_n| \rightarrow |S_{\text{max}}|$ ($|S_{\text{max}}| = \infty$ in the case of Monte Carlo RTE simulations), the expected error in the sampled gradient approaches zero, $|\epsilon_{S_n}| \rightarrow 0$, as do the measures of variance in the sampled gradients ($V_{\text{tot}}^2 \rightarrow 0$, $V_{\|}^2 \rightarrow 0$), as defined in eq. 39 and eq. 39. In other words, as the stochasticity in the problem reduces to zero, we approach the classical step size of the deterministic problem given by $\alpha = \frac{1}{L}$ [Nesterov 2013].

FSR : Introduction

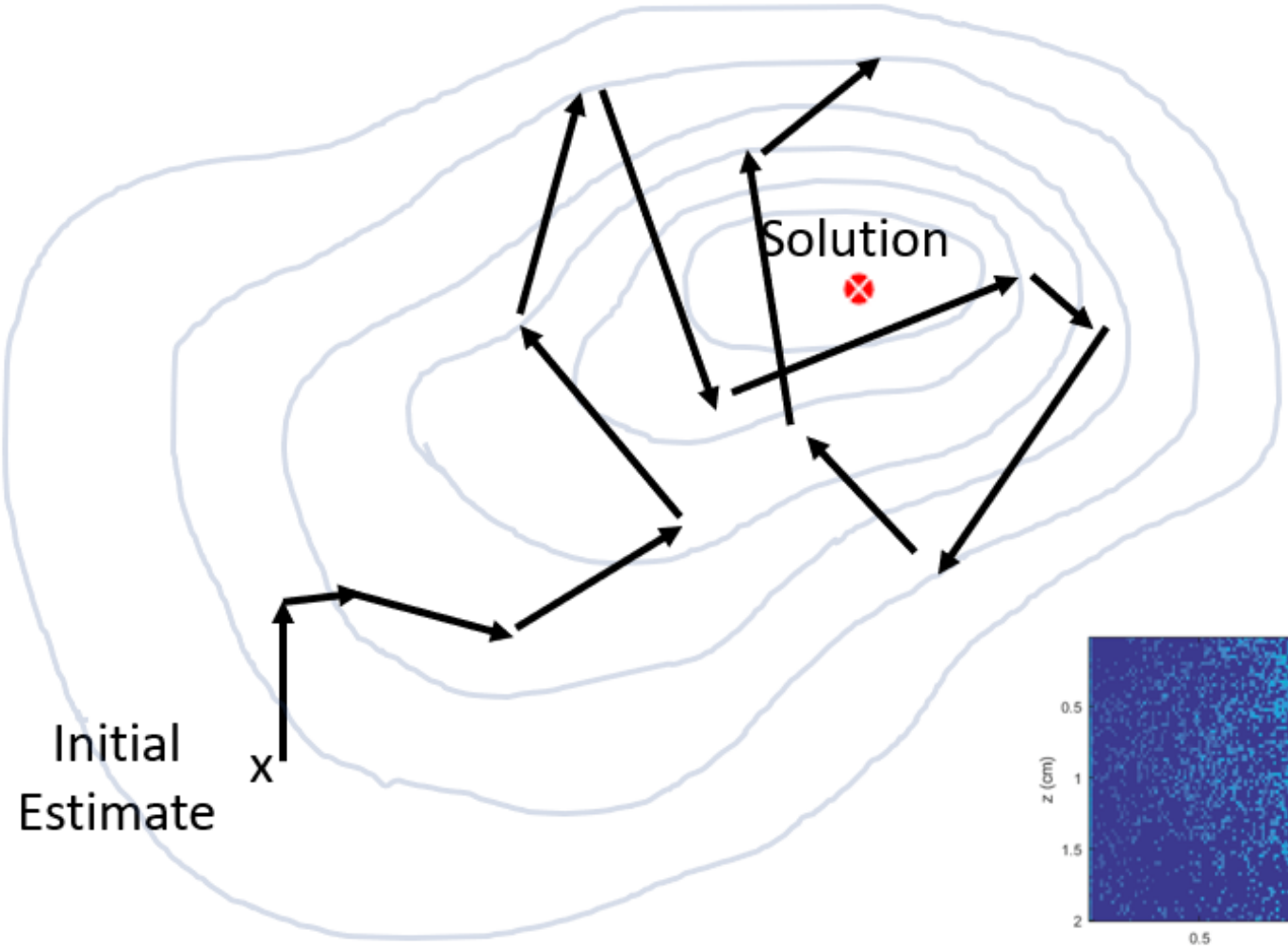
Gradient Descent

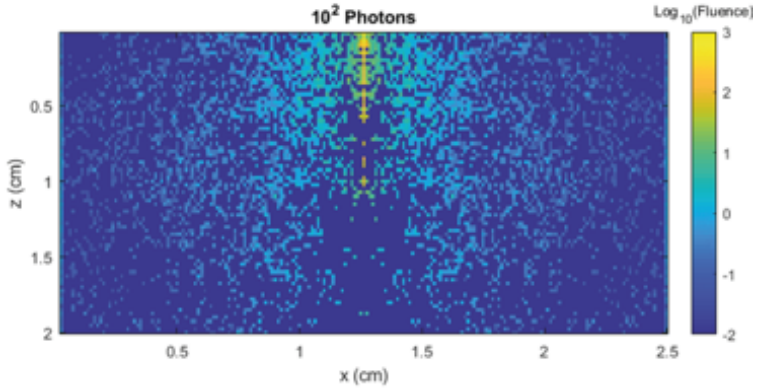




FSR : Introduction

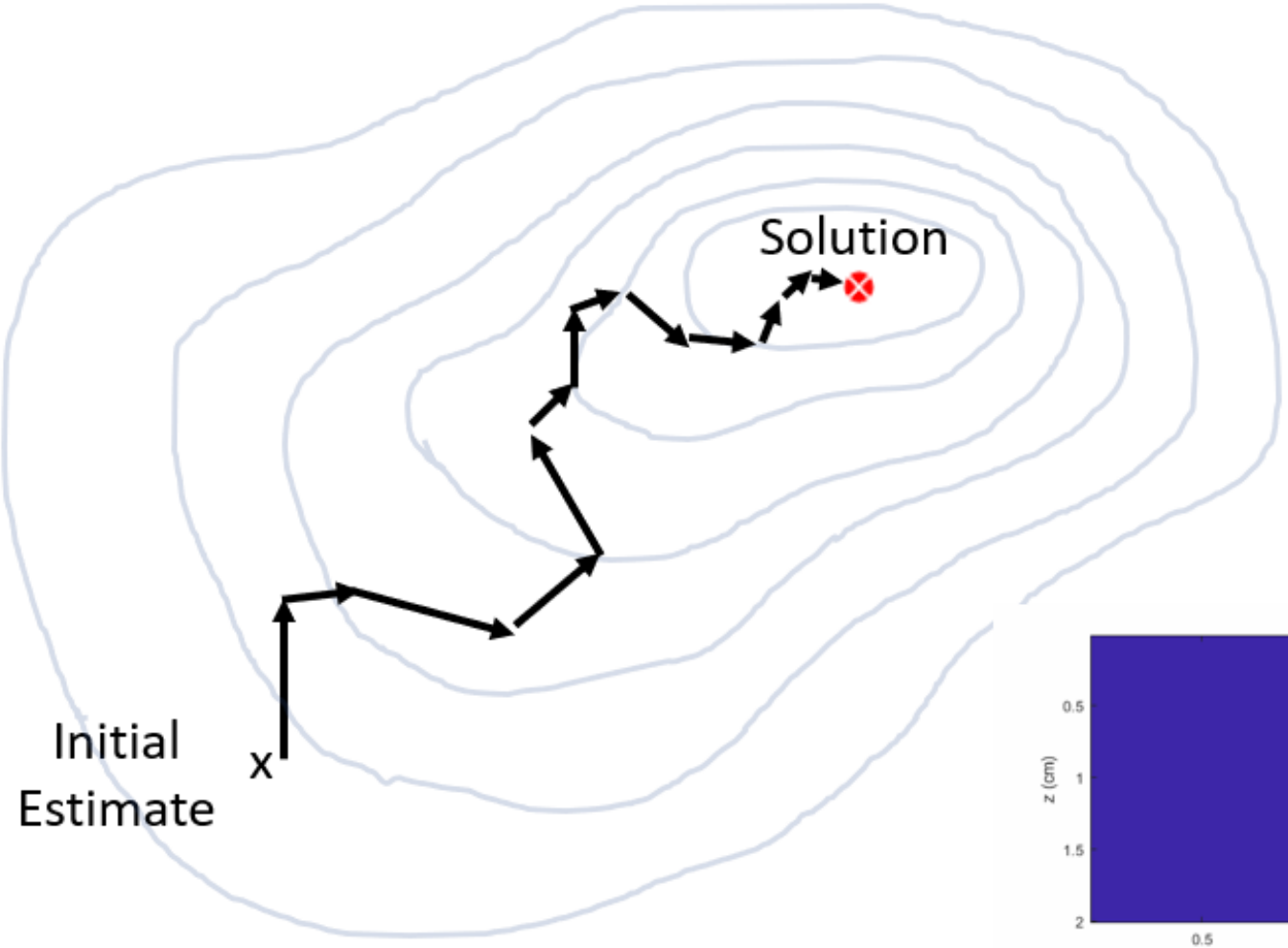
Stochastic Gradient Descent

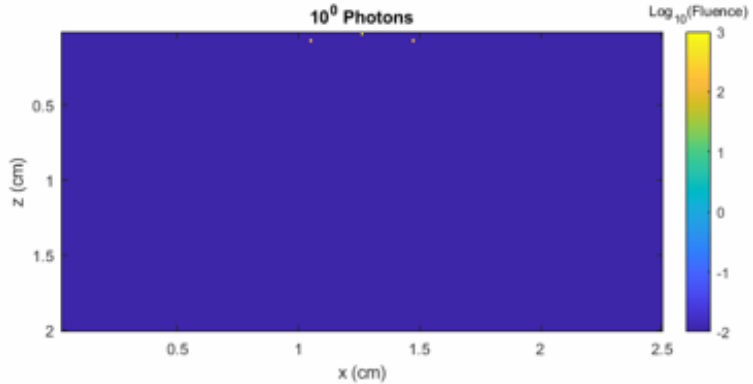




FSR : Introduction

Adaptive Stochastic Gradient Descent





FSR : Testing

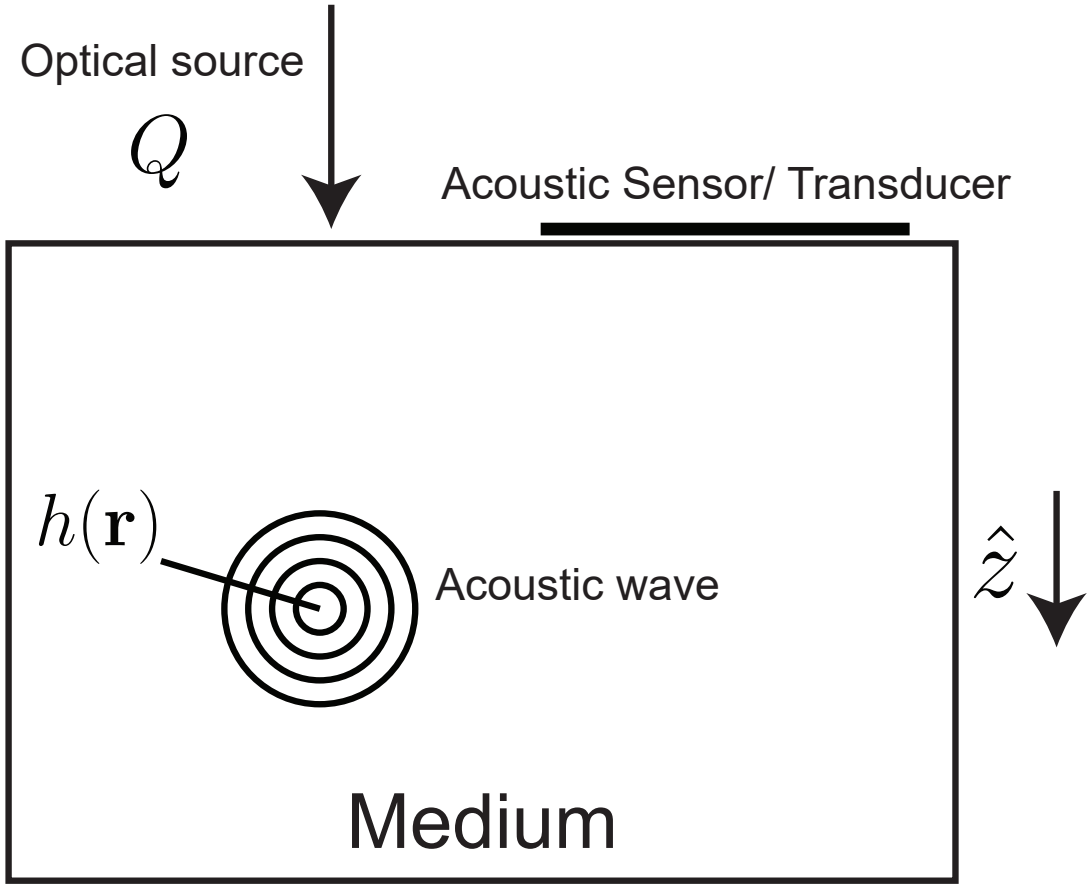
Inverse MC algorithm

Algorithm 1 Inversion using Monte Carlo sampled gradients with adaptive sample size

Choose initial photon sample size $|S_1|$, and desired value of $\gamma_{\|}$, or γ_{tot}
while $\sum_{i=1}^{n} |S_i| < N_{ph}$ **do**
  if run test? **then**
    compute sampled gradient, ∇F_{S_n}, and true gradient, ∇F
    check norm test (or) inner product test is satisfied
    if test fail **then**
      increase sample size on next iteration $|S_{n+1}| = \kappa(n)\,|S_n|$
    else
      set $|S_{n+1}| = |S_n|$
    end if
  else
    compute sampled gradient only ∇F_{S_n}
    set $|S_{n+1}| = |S_n|$
  end if
  update $x_{n+1} = x_n - \alpha_n \nabla F_{S_n}$
end while

FSR : Testing

Setup for Quantitative Photoacoustic Tomography



FSR : Testing

QPAT algorithm

Algorithm 2 Monte Carlo sampled QPAT gradient.

1. Compute $\mathcal{L}_{\mathrm{MC}}^{-1} Q \mapsto \phi$, Φ, using $|S_n|/2$ photons
2. Construct internal adjoint source $Q_{\mathrm{adj}} = \mu_{\mathrm{a}}(h^{\mathrm{obs}} - h)$
3. Compute $\mathcal{L}_{\mathrm{MC}}^{-1*} Q_{\mathrm{adj}} \mapsto \phi^*$, Φ^*, using $|S_n|/2$ photons
4. Use Eq. (26) to compute gradient ∇F_{S_n}

FSR : Testing

Strategy	Step Size, α_n	Sample Size, $\lvert S_{n+1}\rvert = \kappa(n)\lvert S_n\rvert$
1	$\frac{1}{(1+\gamma_{\text{tot}}^2)L}$	$\lvert S_{n+1}\rvert = \frac{V_{\text{tot}}^2}{\gamma_{\text{tot}}^2}\lvert S_n\rvert$
2	$\frac{1}{(1+V_{\text{tot}}^2)L}$	$\lvert S_{n+1}\rvert = \frac{V_{\parallel}^2}{\gamma_{\parallel}^2}\lvert S_n\rvert$
3	$\frac{1}{(1+V_{\text{tot}})L}$	$\lvert S_{n+1}\rvert = \frac{V_{\parallel}}{\gamma_{\parallel}}\lvert S_n\rvert$

Table: Table showing the different inversion strategies used. Strategy 1 has a constant step size, with adaptive sample size. Strategies 2 & 3 both have adaptive step sizes, and adaptive sample sizes. Note that in accordance with Algorithm 1 the sample size is only increased upon a failure of the relevant test. If the test passes, then $\lvert S_{n+1}\rvert = \lvert S_n\rvert$.

FSR : Results

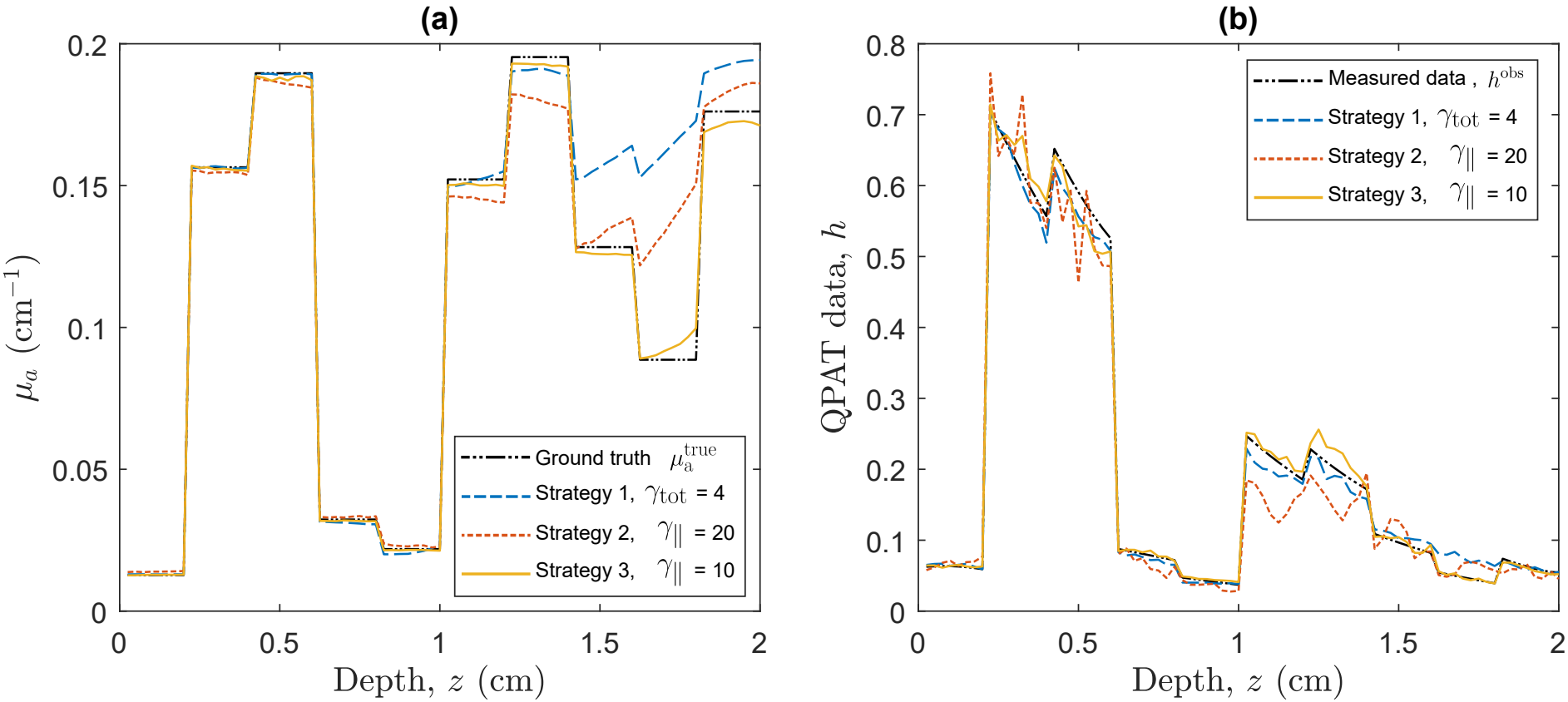


Figure: QPAT inversion: (a) - Ground truth absorption distribution, $\mu_{\text{a}}^{\text{true}}$, and estimated absorption distribution μ_{a} at the point where the photon budget is expended, using each of the three strategies with the stated values of γ_{tot} or $\gamma_{\|}$. (b) - Associated measured data from ground truth medium, and simulated forward data at the end of the inversion using each strategy.

FSR : Results

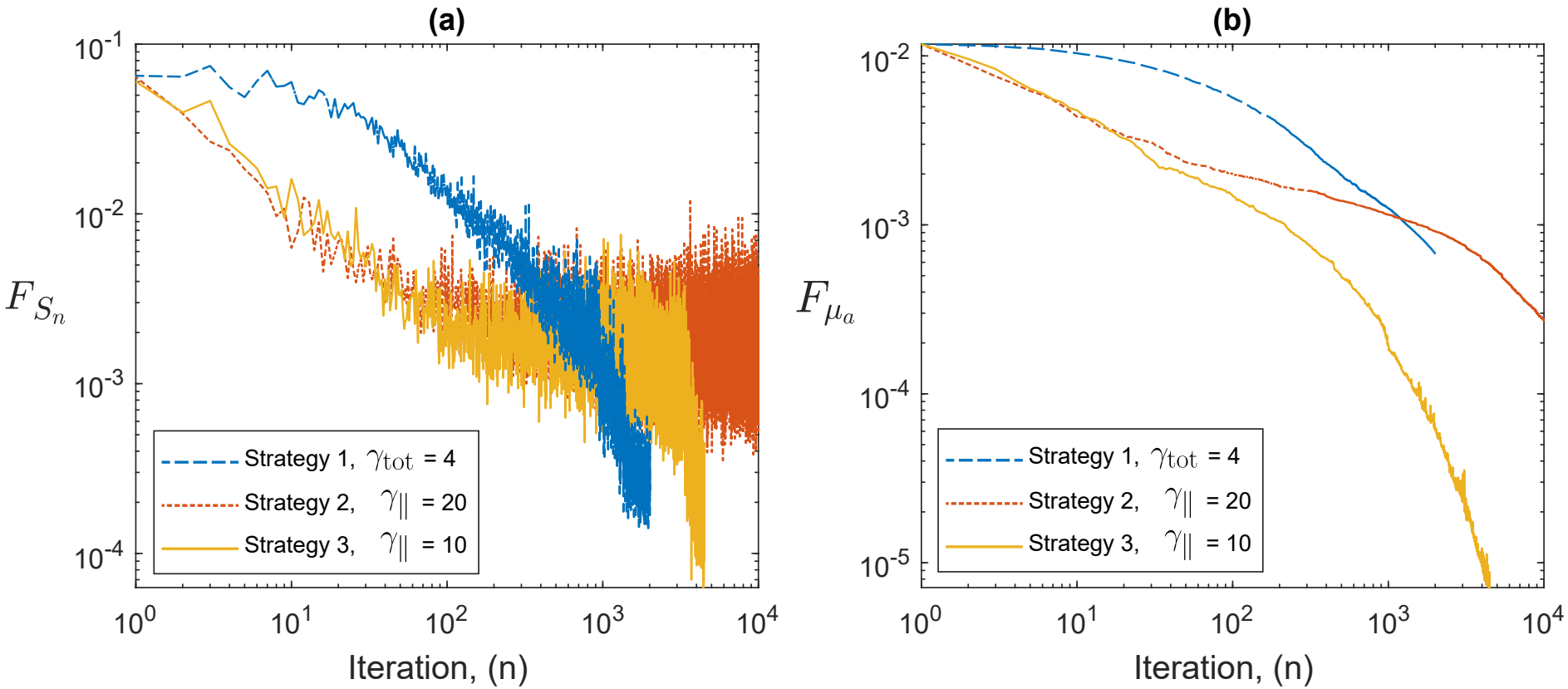


Figure: QPAT inversion: (a) - Sampled cost function, F_{S_n}, as a function of iteration, *n*. (b) - Error in absorption estimate, F_{μ_a}, as a function of iteration, *n*.

FSR : Results

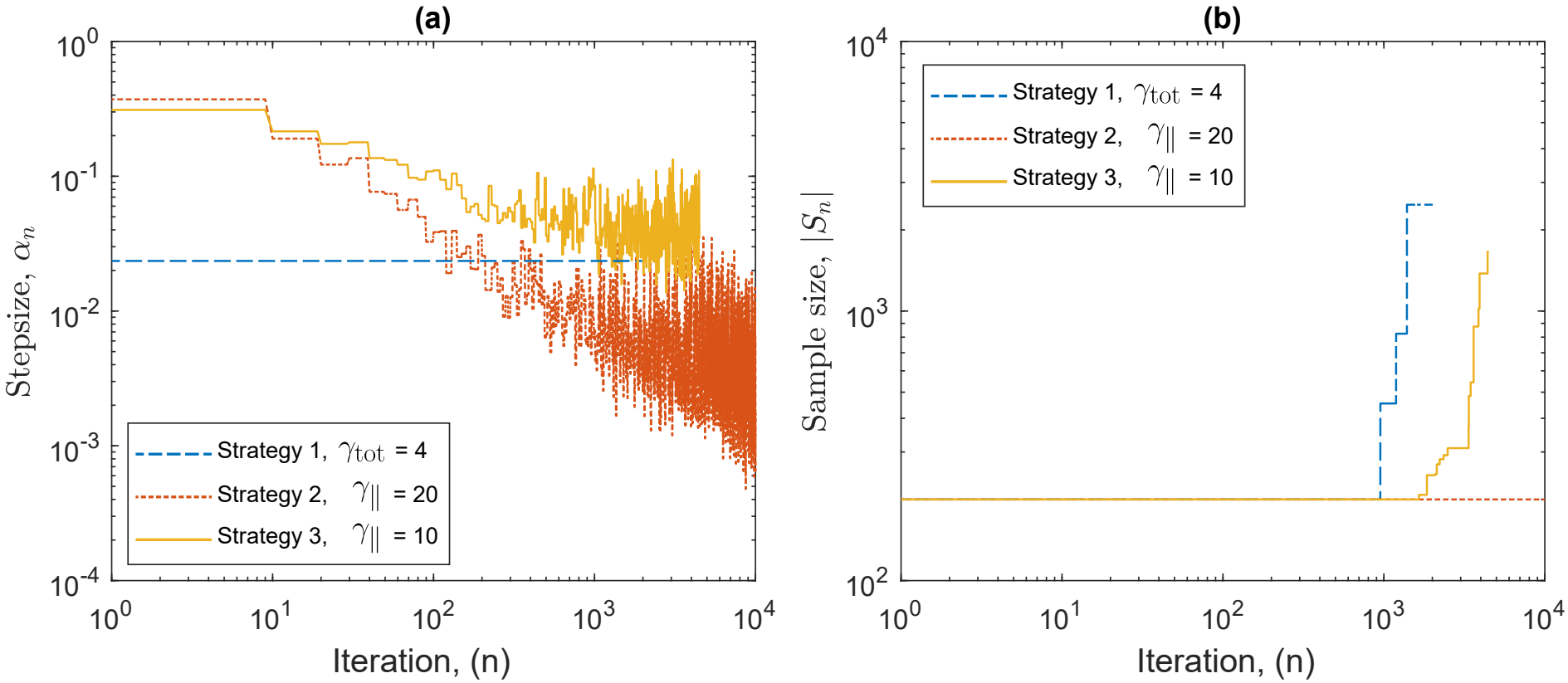
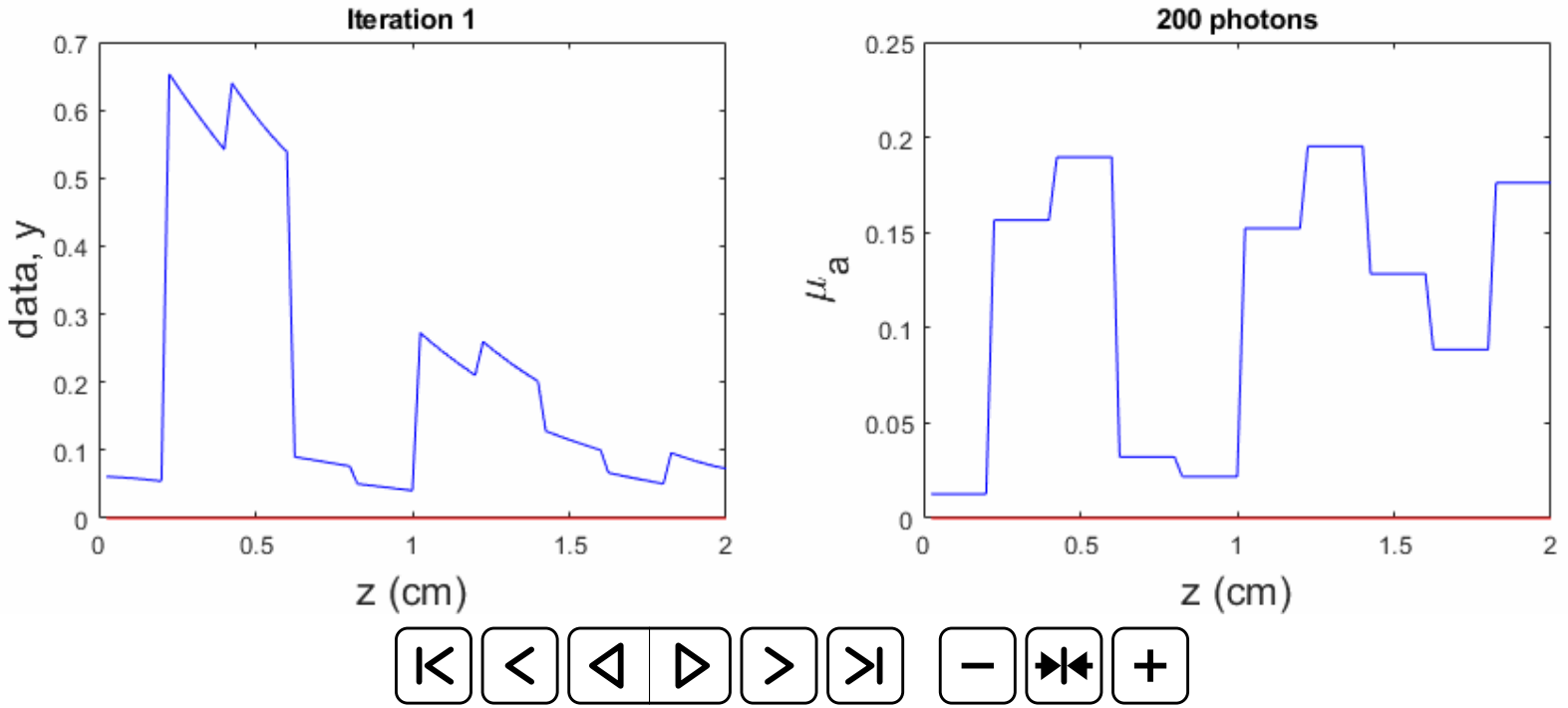


Figure: QPAT inversion: (a) - Step sizes, α_n, as a function of iteration, n. (b) - Adaptive sample size, $|S_n|$, as a function of iteration.

FSR : Results

Test QPrecon movies



FSR : Results

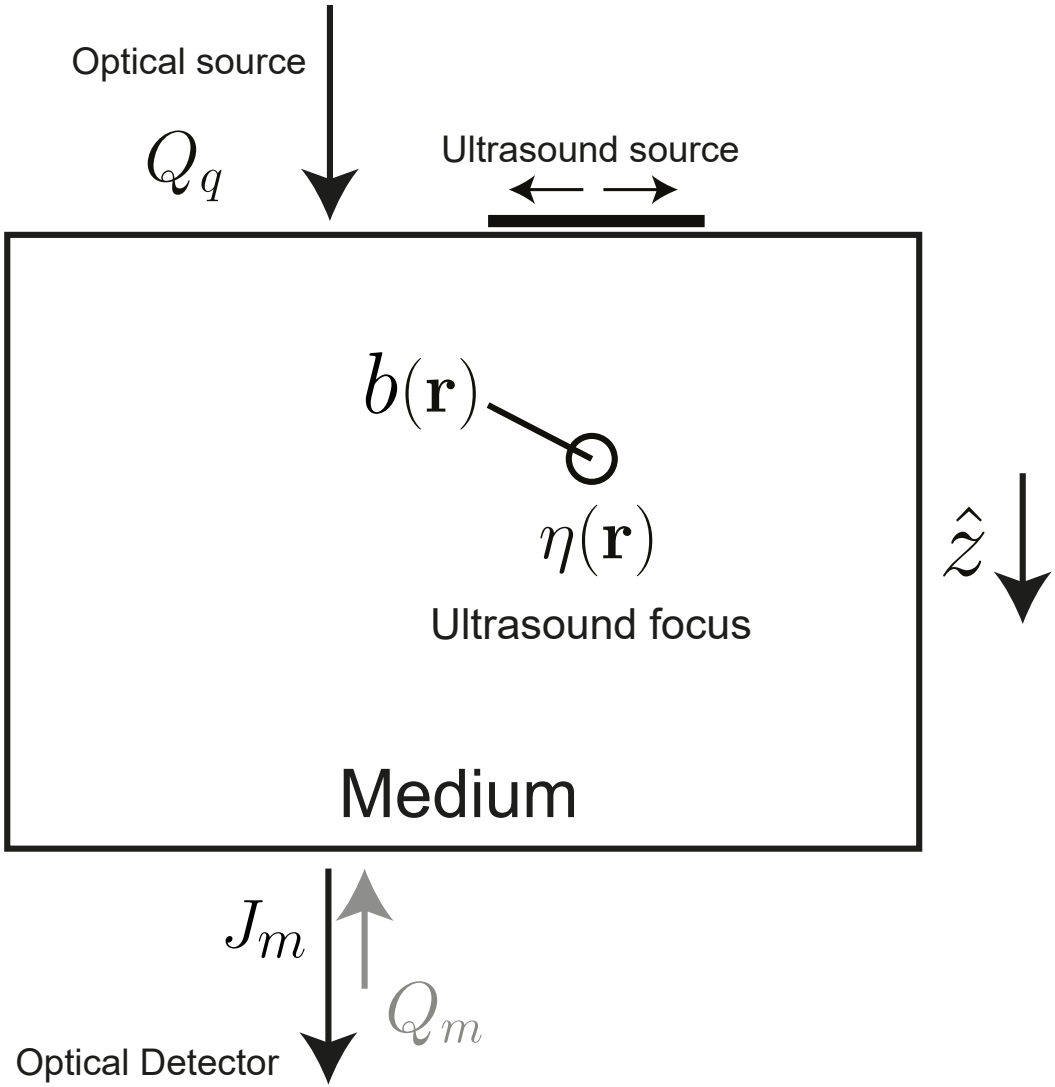
FSR : QPAT Results

		Starting μ_a (cm^{-1})		
		0.01	0.2	1.0
Medium properties	$g = 0.9$ $\mu_s = 40$ cm^{-1}	$\gamma_{\|\|} = 20$, 10,000 iterations $F_{S_n} = 2.35 \times 10^{-3}$ $F_{\mu_a} = 3.26 \times 10^{-5}$	$\gamma_{\|\|} = 10$, 4476 iterations $F_{S_n} = 4.20 \times 10^{-4}$ $F_{\mu_a} = 7.65 \times 10^{-6}$	$\gamma_{\|\|} = 10$, 6533 iterations $F_{S_n} = 3.38 \times 10^{-4}$ $F_{\mu_a} = 1.01 \times 10^{-5}$
	$g = 0.9$ $\mu_s = 4$ cm^{-1}	$\gamma_{\|\|} = 5$, 3819 iterations $F_{S_n} = 5.30 \times 10^{-4}$ $F_{\mu_a} = 2.3 \times 10^{-7}$	$\gamma_{\|\|} = 5$, 2579 iterations $F_{S_n} = 9.31 \times 10^{-5}$ $F_{\mu_a} = 3.56 \times 10^{-7}$	$\gamma_{\|\|} = 5$, 2834 iterations $F_{S_n} = 1.06 \times 10^{-4}$ $F_{\mu_a} = 2.19 \times 10^{-7}$
	$g = 0$ $\mu_s = 4$ cm^{-1}	$\gamma_{\|\|} = 20$, 10,000 iterations $F_{S_n} = 4.01 \times 10^{-3}$ $F_{\mu_a} = 8.36 \times 10^{-5}$	$\gamma_{\|\|} = 5$, 2056 iterations $F_{S_n} = 1.39 \times 10^{-4}$ $F_{\mu_a} = 3.44 \times 10^{-5}$	$\gamma_{\|\|} = 10$, 10,000 iterations $F_{S_n} = 2.92 \times 10^{-3}$ $F_{\mu_a} = 8.72 \times 10^{-5}$

Figure: Final outcomes of QPAT inversions with various medium optical properties and starting values of μ_a. Values of F_{S_n} and F_{μ_a} are the final values at the end of each inversion after the stated number of iterations. In each case Strategy 3 was employed, with a starting sample size of $|S_1| = 200$ photons per iteration, and a total photon budget of $N_{\text{ph}} = 2 \times 10^6$ photons. Slab thickness is 2cm in all cases, with the same ground truth μ_a^{true} distribution as shown in Fig. 5(a).

FSR : Testing

Setup for Ultrasound-Modulated Optical Tomography in the transmission geometry



FSR : Testing

UMOT algorithm

Algorithm 3 Monte Carlo sampled UMOT gradient.

1. Compute $\mathcal{L}_{\mathrm{MC}}^{-1} Q_q \mapsto \phi_q$, Φ_q, and $\mathcal{L}_{\mathrm{MC}}^{-1} Q_m \mapsto \phi_m$, Φ_m, each using $|S_n|/4$ photons
2. Construct internal adjoint sources $Q_{\mathrm{adj}}^1 = \eta \Phi_q (b^{\mathrm{obs}} - b)$ and $Q_{\mathrm{adj}}^2 = \eta \Phi_m (b^{\mathrm{obs}} - b)$
3. Compute $\mathcal{L}_{\mathrm{MC}}^{-1*} Q_{\mathrm{adj}}^1 \mapsto \phi^{*,1}$, $\Phi^{*,1}$, and $\mathcal{L}_{\mathrm{MC}}^{-1*} Q_{\mathrm{adj}}^1 \mapsto \phi^{*,2}$, $\Phi^{*,2}$, each using $|S_n|/4$ photons
4. Use Eq. (37) to compute ∇F_{S_n}

FSR : Testing

FSR : UMOT Results

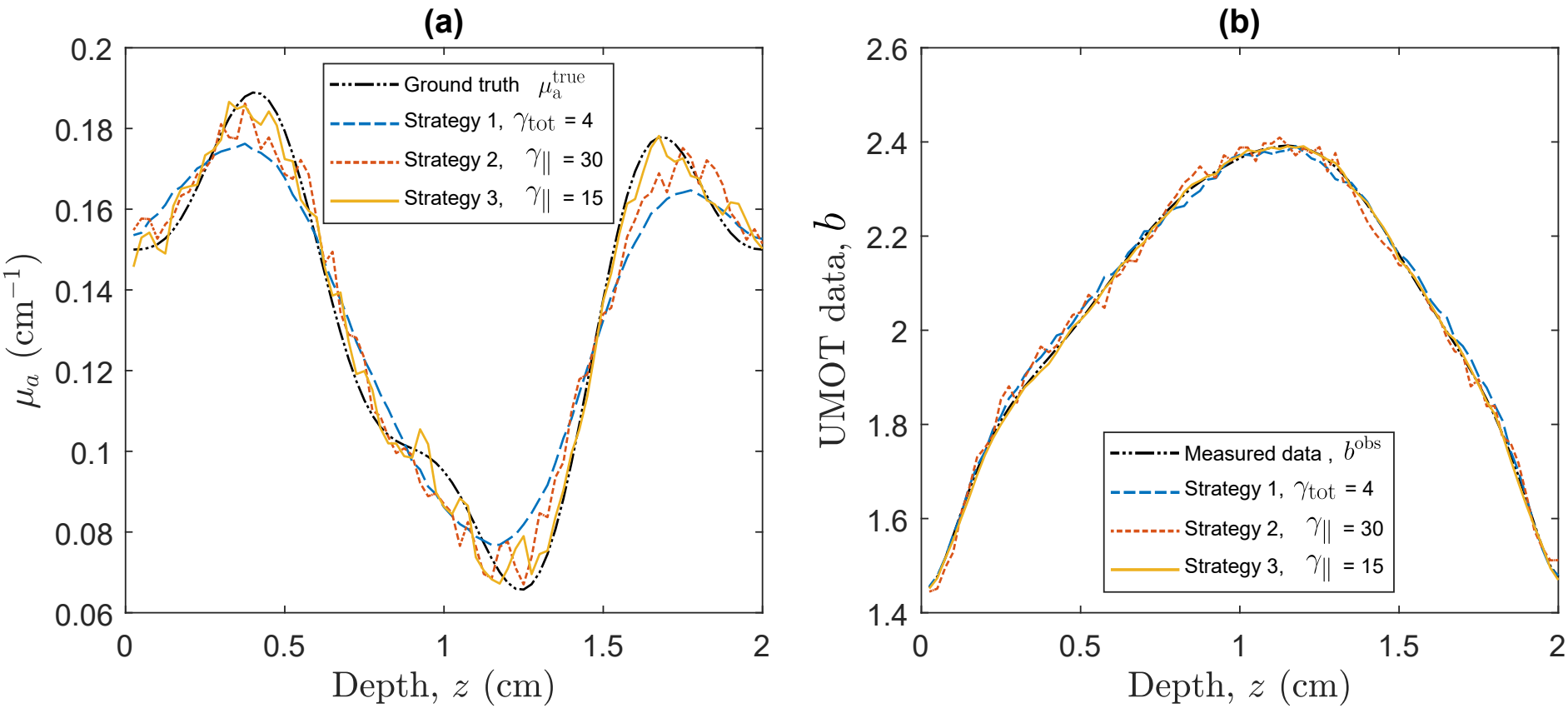


Figure: UMOT inversion: (a) - Ground truth absorption distribution, μ_a^{true}, and recovered absorption distribution μ_a using each of the three strategies with the stated values of γ_{tot} or $\gamma_{||}$. (b) - Associated measured data from ground truth medium, and simulated forward data at the end of the inversion using each strategy.

FSR : Testing

FSR : UMOT Results

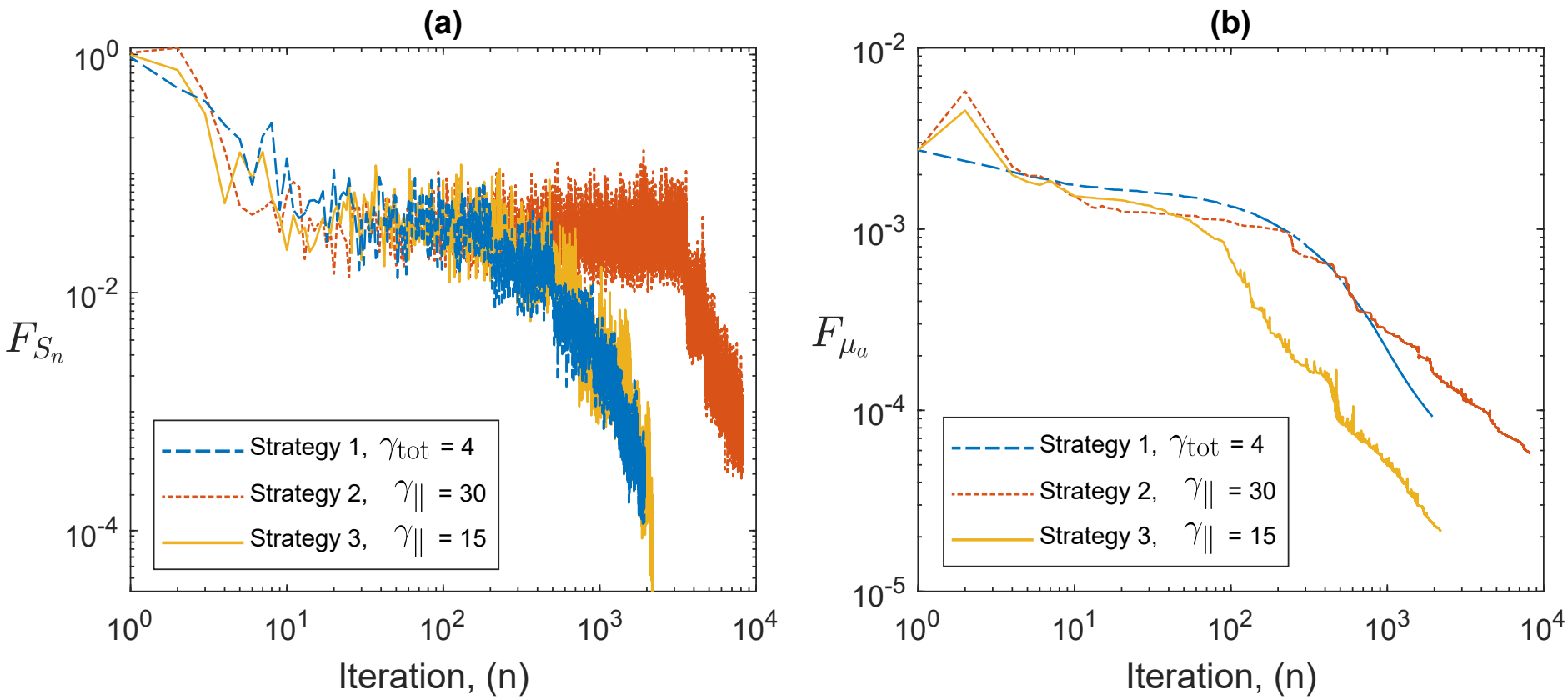


Figure: UMOT inversion: (a) - Sampled cost function, F_{S_n}, as a function of iteration, n. (b) - Error in absorption estimate, F_{μ_a}, as a function of iteration, n.

FSR : Testing

FSR : UMOT Results

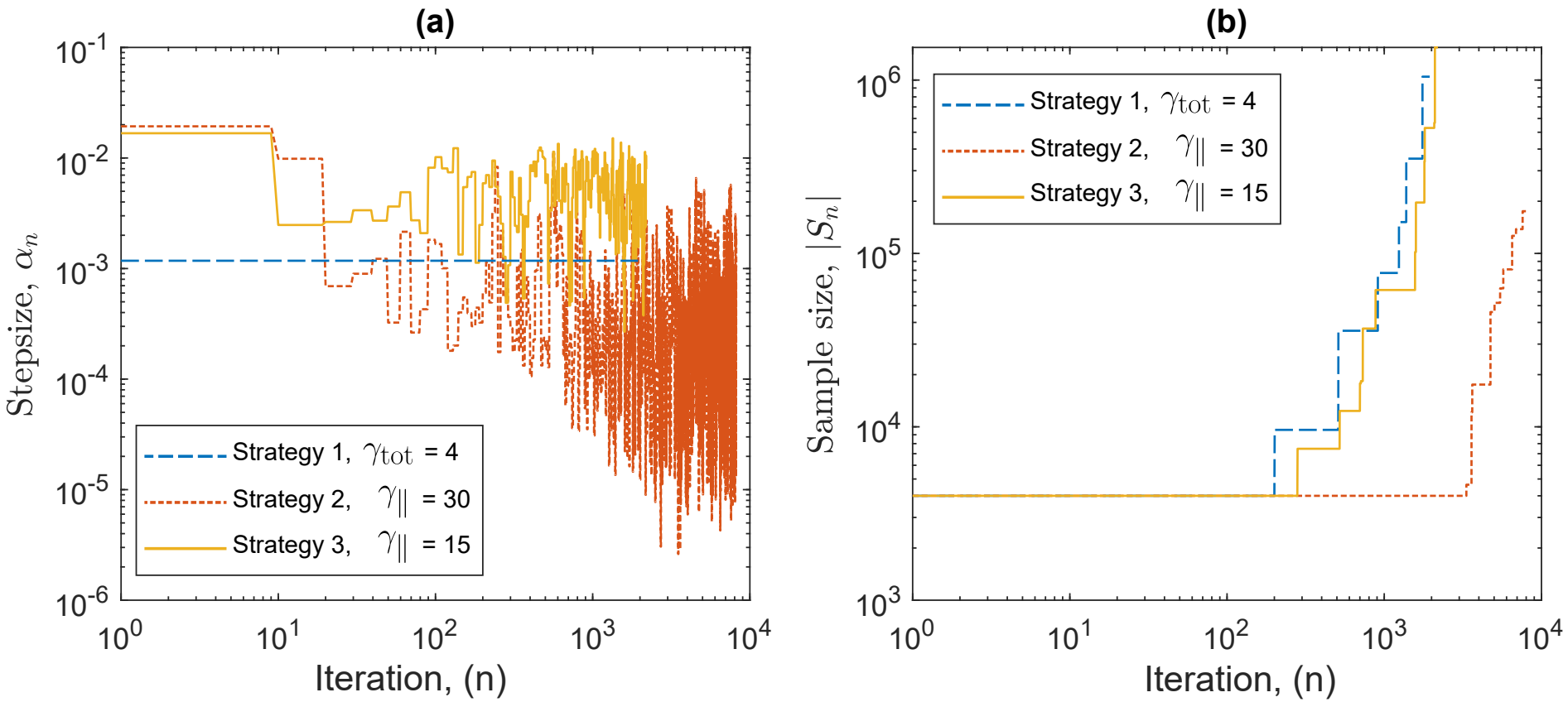


Figure: UMOT inversion: (a) - Step sizes, α_n, as a function of iteration, *n*. (b) - Adaptive sample size, $|S_n|$, as a function of iteration.

FSR : Testing

FSR : UMOT Results

		Starting μ_a (cm^{-1})		
		0.01	0.1	1.0
Medium properties	$g = 0.9$ $\mu_s = 40\ cm^{-1}$	$\gamma_{\|\|} = 15$, 11,412 iterations $F_{S_n} = 1.88 \times 10^{-3}$ $F_{\mu_a} = 4.23 \times 10^{-5}$	$\gamma_{\|\|} = 10$, 3495 iterations $F_{S_n} = 2.50 \times 10^{-4}$ $F_{\mu_a} = 2.20 \times 10^{-5}$	$\gamma_{\|\|} = 10$, 7823 iterations $F_{S_n} = 5.69 \times 10^{-4}$ $F_{\mu_a} = 9.37 \times 10^{-5}$
	$g = 0.9$ $\mu_s = 4\ cm^{-1}$	$\gamma_{\|\|} = 15$, 19,017 iterations $F_{S_n} = 4.71 \times 10^{-3}$ $F_{\mu_a} = 8.19 \times 10^{-5}$	$\gamma_{\|\|} = 15$, 15,813 iterations $F_{S_n} = 5.48 \times 10^{-3}$ $F_{\mu_a} = 2.07 \times 10^{-5}$	$\gamma_{\|\|} = 15$, 14,249 iterations $F_{S_n} = 4.08 \times 10^{-3}$ $F_{\mu_a} = 3.36 \times 10^{-5}$
	$g = 0$ $\mu_s = 4\ cm^{-1}$	$\gamma_{\|\|} = 15$, 11,594 iterations $F_{S_n} = 1.65 \times 10^{-3}$ $F_{\mu_a} = 7.13 \times 10^{-5}$	$\gamma_{\|\|} = 15$, 7304 iterations $F_{S_n} = 5.87 \times 10^{-4}$ $F_{\mu_a} = 4.17 \times 10^{-5}$	$\gamma_{\|\|} = 15$, 13,981 iterations $F_{S_n} = 1.05 \times 10^{-3}$ $F_{\mu_a} = 7.98 \times 10^{-5}$

Figure: Final outcomes of UMOT inversions with various medium optical properties and starting values of μ_a. Values of F_{S_n} and F_{μ_a} are the final values at the end of each inversion after the stated number of iterations. In each case Strategy 3 was employed, with a starting sample size of $|S_1| = 4000$ photons per iteration, and a total photon budget of $N_{ph} = 4 \times 10^8$ photons. Slab thickness is 2cm in all cases, with the same ground truth μ_a^{true} distribution as shown in Fig. 9(a).

Outline

Conclusions and Outlook

Conclusions

- Imaging from Coupled Optics and Acoustics : PAT, UMOT
- Several methods for modelling
- Combine approximate models with learning
- Stochastic models of light propagation combined with stochastic optimisation : Fully Stochastic Inversion
- Topics not mentioned : "one step" reconstruction, multispectral aspects

Conclusions and Outlook

Outlook

- FSR examples shown were "noise free". Should include explicit regularisation and/or early stopping
- Other noise models for photon counting e.g. Kullback-Leibler, Wasserstein Distance
- Further adaptive subsampling scheme (SAG, SAGA, SARAH etc)
- Estimation of Lipschitz coefficient for stochastic forward models
- Preconditioning methods