Fully Stochastic Reconstruction for Inverse Radiative Transport

Simon R. Arridge[1]

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

Waves and Imaging in Complex Media
Paris
June 9-13 2025

Outline

Outline

1. Introduction
2. Modelling of Photon Propagation
3. Deterministic Reconstruction Methods in Inverse Transport
4. Fully Stochastic Reconstruction (FSR) for RTE problems
   - Example in 1D
   - Radiance Monte Carlo
   - Examples in 2D
5. Conclusions and Outlook

Main collaborators for the work presented in this talk

UCL	P. Beard, M. Betcke, B. Cox, A. Hauptmann, R. Huchli, N. Huynh, Z. Kereta, T. Leung, C. Macdonald, S. Powell, B. Treeby
UNottingham	S. Powell
UeF	N. Hanninen, J.Kangasniemi, A. Leino, A. Pulkkinen, M. Suhonen, T. Tarvainen
UOulu	A. Hauptmann
CWI	F. Lucka

- Photon propagation can be described by a range of models from fully transportive to fully diffusive
- Deterministic models are complex to solve and exhibit *bias* due to modelling error
- Stochastic models are accurate and unbiased but exhibit high variance
- In this talk we leverage three recent developments that allow the use of stochastic models in solving non-linear inverse problems for the first time

1. Adjoint Monte Carlo that allows fast computation of unbiased estimates of the backprojection operator in optical tomography
2. Radiance Monte Carlo that efficiently computes the distribution of photon directions in space
3. Stochastic Gradient Methods that allow the estimation of inverse problem solutions from inaccurate (but unbiased) estimates of functional gradients.

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

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

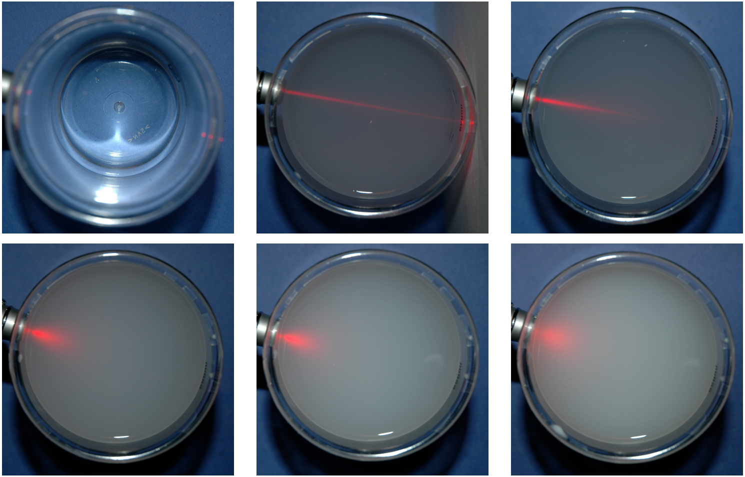
$\mu_{\mathrm{tr}} = \mu_{\mathrm{s}} + \mu_{\mathrm{a}}$ is the attenuation coefficent
$\mathcal{S}$ is the scattering operator, (local, non propagating).
Method of successive approximation (Sobolev 1963) :

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

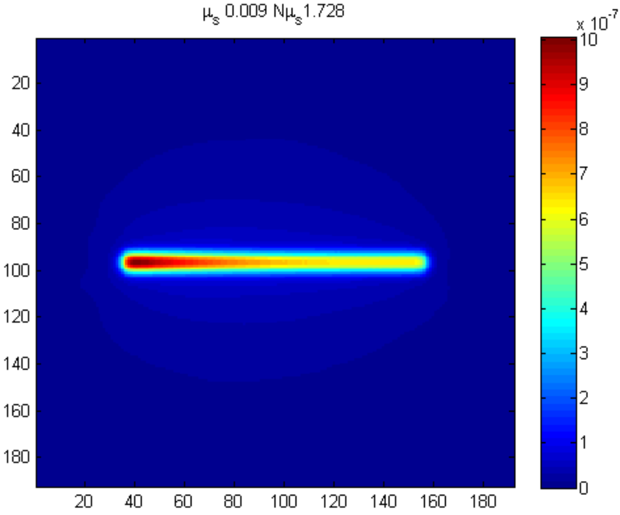
The first term may be found from the Ray Transform, giving an alternative equation for the *collided flux*

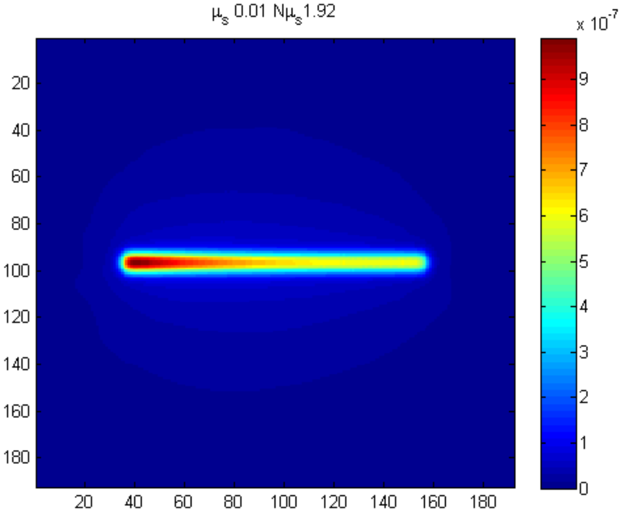
$$\left[\mathcal{T}_{\mu_{\mathrm{tr}}} - \mu_{\mathrm{s}}\mathcal{S}\right]\phi_{\mathrm{collided}} = \mu_{\mathrm{s}}\mathcal{S}\;\underbrace{\mathcal{T}_{\mu_{\mathrm{tr}}}^{-1} q}_{\text{uncollided}} \tag{6}$$

Light scattering



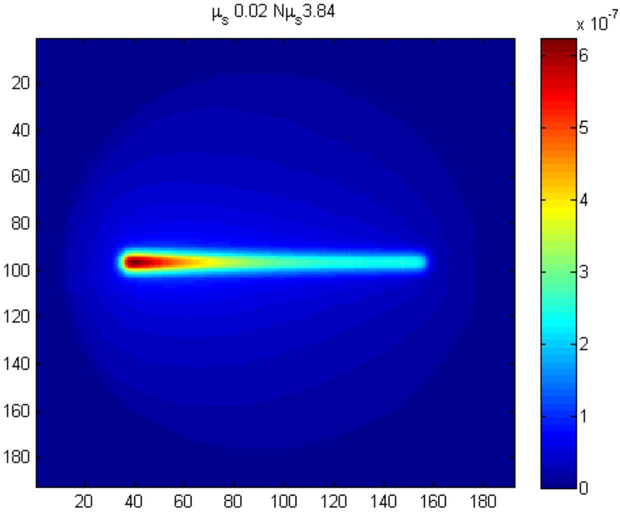
Experiment by Nina Schotland (age 8)

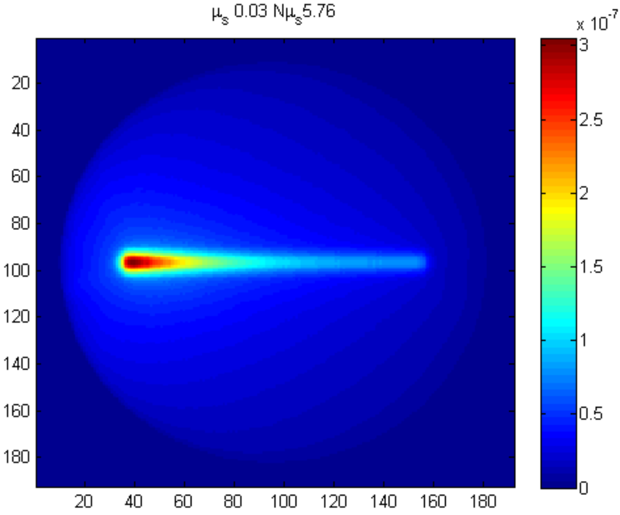

μs 0.009 Nμs 1.728
x 10^-7
20
40
60
80
100
120
140
160
180
0
1
2
3
4
5
6
7
8
9
10



μ_s 0.01 $N\mu_s$ 1.92
x 10^{-7}
20
40
60
80
100
120
140
160
180
9
8
7
6
5
4
3
2
1
0

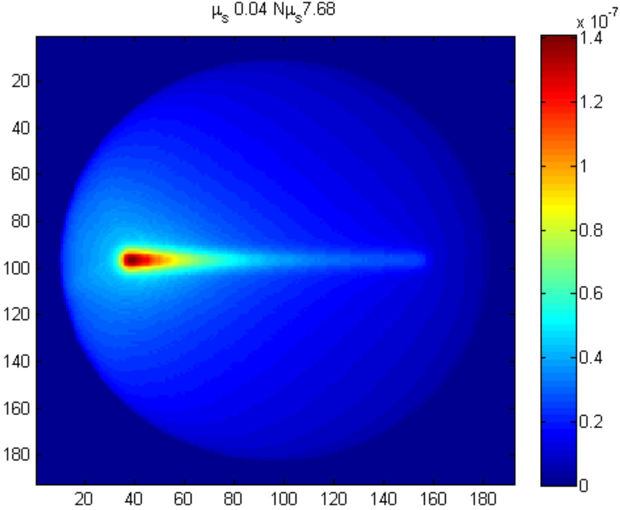
Modelling in Optical Tomography
RTE solutions

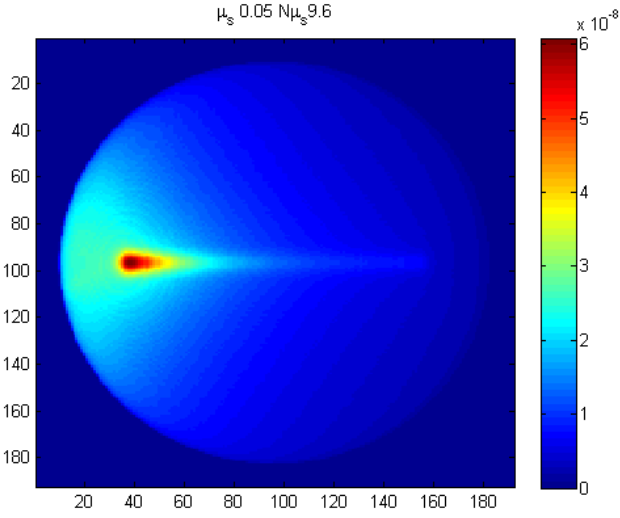



μ_s 0.03 $N\mu_s$5.76
x 10^{-7}
3
2.5
2
1.5
1
0.5
0
20
40
60
80
100
120
140
160
180

Modelling in Optical Tomography

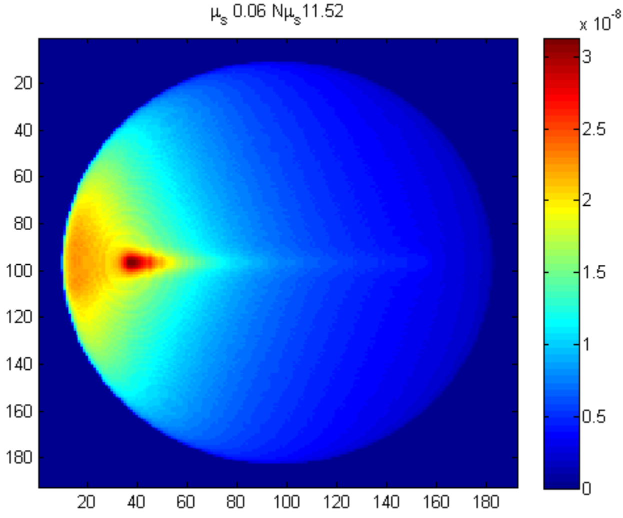
RTE solutions



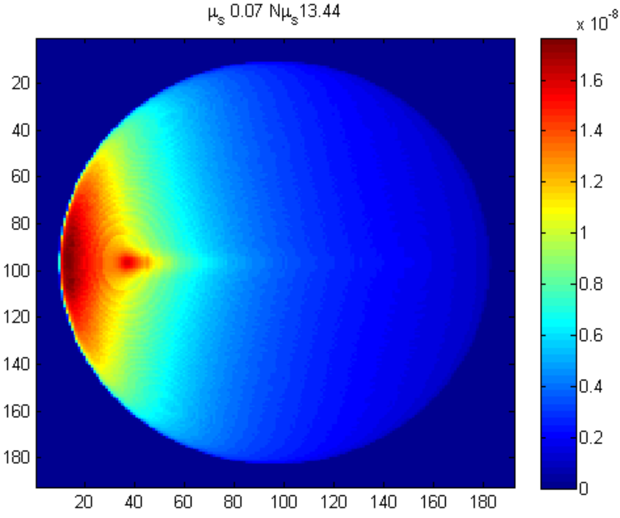


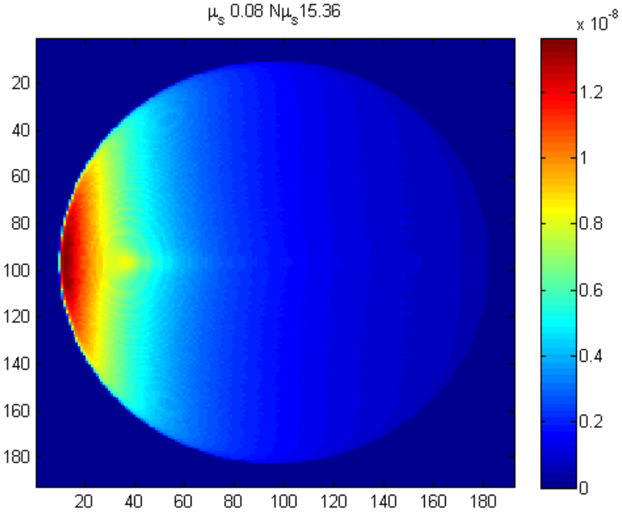
μ_s 0.05 Nμ_s9.6
x 10^{-8}
20
40
60
80
100
120
140
160
180
0
1
2
3
4
5
6

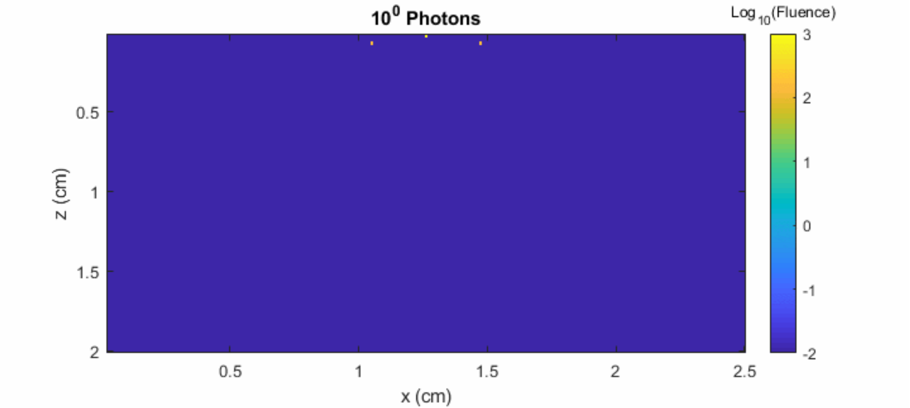
Modelling in Optical Tomography
RTE solutions



Modelling in Optical Tomography
RTE solutions




μs 0.08 Nμs 15.36
x 10^-8
20
40
60
80
100
120
140
160
180
1.2
1
0.8
0.6
0.4
0.2
0



10^0 Photons
Log 10(Fluence)
3
2
1
0
-1
-2
0.5
1
1.5
2
z (cm)
0.5
1
1.5
2
2.5
x (cm)

A standard reciprocity theorem for the Boltzmann equation is given in Case and Zweifel, section 2.7:

$$G^{\phi}(\boldsymbol{r}, \hat{\boldsymbol{s}}; \boldsymbol{r}_0, \hat{\boldsymbol{s}}_0) = G^{\phi}(r_0, -\hat{\boldsymbol{s}}_0; \boldsymbol{r} - \hat{\boldsymbol{s}})$$

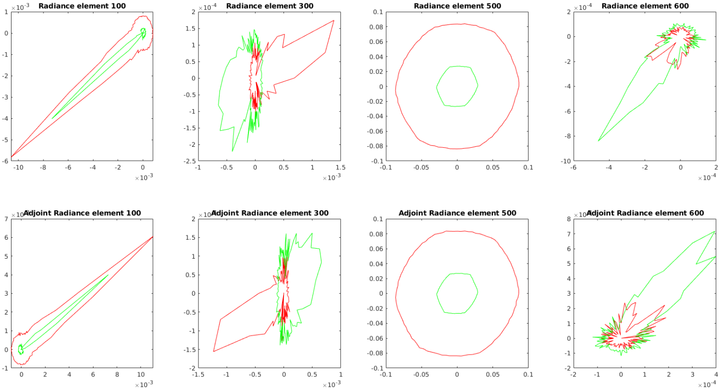
which states that the angular density (radiance) at $\boldsymbol{r}$ in direction $\hat{\boldsymbol{s}}$ due to a source at $\boldsymbol{r}_0$ in direction $\hat{\boldsymbol{s}}_0$ is the same as the angular density (radiance) at $\boldsymbol{r}_0$ in direction $-\hat{\boldsymbol{s}}_0$ due to a source at $\boldsymbol{r}$ in direction $-\hat{\boldsymbol{s}}$. By integrating over $-\hat{\boldsymbol{s}}_0$ we obtain

$$G^{\phi}(\boldsymbol{r}, \hat{\boldsymbol{s}}; q(\boldsymbol{r}_0)) = G^{\Phi}(r_0; \boldsymbol{r} - \hat{\boldsymbol{s}})$$

which states that the angular density (radiance) at $\boldsymbol{r}$ in direction $\hat{\boldsymbol{s}}$ due to a point isotropic source at $\boldsymbol{r}_0$ is the same as the photon density (fluence) at $\boldsymbol{r}_0$ due to a source at $\boldsymbol{r}$ in direction $-\hat{\boldsymbol{s}}$.
These basic relations give rise to the adjoint formulation of the sensitivity relations for the Boltzmann equation, and thus detemine how to set up our key method : *(Adjoint) Radiance Monte Carlo*

Modelling of Photon Propagation

Radiance Monte Carlo : Direct and adjoint



Outline

We define a *forward model* as the composition

$$y = f(x) = \mathcal{M} \circ \mathcal{A}(x) \quad = \mathcal{M} \circ K(x, \phi) = \mathcal{M} \circ K\left(x, \mathcal{L}^{-1}(x) q\right)$$

where $x = \begin{pmatrix} \mu_{\rm a} \\ \mu_{\rm s} \end{pmatrix}$ and $K(x, \phi)$ is a functional of radiance ϕ and parameters x
The inverse problem is to recover x from (sufficient) measurements y.
We treat this as minimisation of a cost function

$$F \quad = \quad \frac{1}{2} \int_\Omega (y^{\rm obs} - y)^2 {\rm d}r \quad = \quad \frac{1}{2} \left\langle y^{\rm obs} - y, y^{\rm obs} - y \right\rangle_{L^2(\Omega)} . \tag{7}$$

then the Fréchet derivative of F is

$$DF = -\left\langle y^{\rm obs} - y, Df \begin{pmatrix} \mu_{\rm a}^\delta \\ \mu_{\rm s}^\delta \end{pmatrix} \right\rangle_{L^2(\Omega)} , \tag{8}$$

where $\mu_{\rm a}^\delta, \mu_{\rm s}^\delta$ are small changes in absorption and scattering.

Writing the Fréchet derivative of K as

$$DK = \frac{\partial K}{\partial x} + \frac{\partial K}{\partial \phi}\frac{\partial \phi}{\partial x}, \tag{9}$$

we arrive at

$$DF = -\left\langle \frac{\partial K}{\partial x}\mathcal{M}^*(y^{\mathrm{obs}} - y), x^\delta \right\rangle_{L^2(\Omega)} - \left\langle \frac{\partial K}{\partial \phi}\mathcal{M}^*(y^{\mathrm{obs}} - y), \phi^\delta \right\rangle_{L^2(\Omega) n \times S^{n-1})}. \tag{10}$$

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

$$\mathcal{L}^*\phi^* = \frac{\partial K}{\partial \phi}\mathcal{M}^*(y^{\mathrm{obs}} - y) \tag{11}$$

where the right hand side describes the “adjoint source’.’ We then substitute the above into eq. 10 to give

$$\begin{aligned} DF &= -\left\langle \frac{\partial K}{\partial x}\mathcal{M}^*(y^{\mathrm{obs}} - y), x^\delta \right\rangle_{L^2(\Omega)} - \left\langle \mathcal{L}^*\phi^*, \phi^\delta \right\rangle_{L^2(\Omega\times S^{n-1})} && (12) \\ &= -\left\langle \frac{\partial K}{\partial x}\mathcal{M}^*(y^{\mathrm{obs}} - y), x^\delta \right\rangle_{L^2(\Omega)} - \left\langle \phi^*, \mathcal{L}\phi^\delta \right\rangle_{L^2(\Omega\times S^{n-1})}. && (13) \end{aligned}$$

Consider a change to eq. 4 where $\mu_{\rm a} \rightarrow \mu_{\rm a} + \mu_{\rm a}^\delta$, $\mu_{\rm s} \rightarrow \mu_{\rm s} + \mu_{\rm s}^\delta$, for the same source q, which results in a change in radiance $\phi \rightarrow \phi + \phi^\delta$. This implies

$$
\begin{aligned}
\left(\mathcal{T}_{\mu_{\rm a}+\mu_{\rm a}^\delta,\mu_{\rm s}+\mu_{\rm s}^\delta} - \mathcal{S}_{\mu_{\rm s}+\mu_{\rm s}^\delta}\right)\left(\phi + \phi^\delta\right) &= \left(\mathcal{T}_{\mu_{\rm a},\mu_{\rm s}} - \mathcal{S}_{\mu_{\rm s}}\right)\phi \\
\Rightarrow \quad \left(\mathcal{T}_{\mu_{\rm a},\mu_{\rm s}} - \mathcal{S}_{\mu_{\rm s}}\right)\phi^\delta &= -(\mu_{\rm a}^\delta + \mu_{\rm s}^\delta + \mathcal{S}_{\mu_{\rm s}^\delta})\phi \qquad (14) \\
\mathcal{L}_{\mu_{\rm a},\mu_{\rm s}}\phi^\delta &= -\underbrace{\left(\mu_{\rm a}^\delta + \mu_{\rm s}^\delta + \mathcal{S}_{\mu_{\rm s}^\delta}\right)}_{\mathcal{L}^\delta_{\mu_{\rm a}^\delta,\mu_{\rm s}^\delta}}\phi\,. \qquad (15)
\end{aligned}
$$

We have

$$
DF = -\left\langle \frac{\partial K}{\partial x}\mathcal{M}^*(y^{\rm obs} - y), x^\delta \right\rangle_{L^2(\Omega)} - \left\langle \phi^*\phi, \mu_{\rm a}^\delta \right\rangle_{L^2(\Omega\times S^{n-1})}\,, \qquad (16)
$$

allowing us to define the gradients

$$
\begin{aligned}
\frac{\partial F}{\partial \mu_{\rm a}} &= -\frac{\partial K}{\partial \mu_{\rm a}}\mathcal{M}^*(y^{\rm obs} - y) - \int_{S^{n-1}} \phi^*(\hat{\mathbf{s}})\phi(\hat{\mathbf{s}})\,{\rm d}\hat{\mathbf{s}} \qquad (17) \\
\frac{\partial F}{\partial \mu_{\rm s}} &= -\frac{\partial K}{\partial \mu_{\rm s}}\mathcal{M}^*(y^{\rm obs} - y) - \int_{S^{n-1}} \phi^*(\hat{\mathbf{s}})\left[\mathrm{I} - \int_{S^{n-1}} \Theta(\hat{\mathbf{s}},\hat{\mathbf{s}}')\phi((\hat{\mathbf{s}}')\,{\rm d}\hat{\mathbf{s}}'\right]{\rm d}\hat{\mathbf{s}} \qquad (18)
\end{aligned}
$$

Outline

- Solving inverse radiative transport problems requires an accurate forward model

- Solving inverse radiative transport problems requires an accurate forward model
- Deterministic RTE is computationally expensive and needs adapting to problem domain

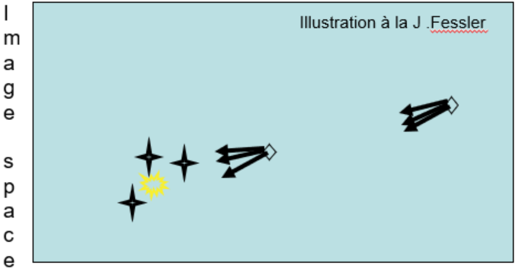
- Solving inverse radiative transport problems requires an accurate forward model
- Deterministic RTE is computationally expensive and needs adapting to problem domain
- Stochastic Monte Carlo modelling is "arbitrarily accurate" *in expectation* but not an operator between Banach spaces

- Solving inverse radiative transport problems 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 an 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]

- Solving inverse radiative transport problems 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 an 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].
- Related to MC^3 for posterior samlping using inaccurate forward models [Bal. Langmore, Marzouk, 2013].

FSR : Introduction

Subsets : Why Does it Work ?



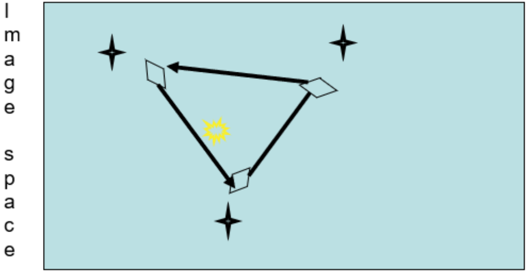
Stars: MAP/ML solutions for each subset of data.
Yellow: MAP/ML solution for all data.
Diamonds: images within limit-cycle.

Figure courtesy Kris Thielemans

The black crosses are the ML solutions when using each subset of the data (i.e. each subset will have a slightly different ML solution).
The arrows indicate an update (3 subsets). They point "towards" the ML solution of the subset.
As long as we are "far" away from the ML solutions, it doesn't matter which subset we use.
However, once we start to get "close", then each update will be somewhat different, and actually no longer point towards the true ML solution.

FSR : Introduction

Subsets : Close to convergence



Stars:	MAP/ML solutions for each subset of data.
Yellow:	MAP/ML solution for all data.
Diamonds:	images within limit-cycle.

After a lot of iterations, subsets will usually cycle between different images. The ML solutions for each subset will usually not be identical. Therefore, each update will change the image "towards" the ML-solution for the subset that it uses.

A few different strategies can be used to solve this problem:

(i) reduce the number of subsets after a few iterations (i.e. increase the size of data considered in each subset).

(ii) use "relaxation" (i.e. reduce the update step-size), [Block-RAMLA].

- *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 $\lim_{n\to\infty} F(x_n) = 0$.

- *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.

- 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

$$\textbf{norm test} \quad V^2_{\mathrm{tot}}(x_n) := \frac{\mathbb{E}\left[\left| \epsilon_{S_n}(x_n) \right|^2 \right]}{\left| \nabla F(x_n) \right|^2} \leq \gamma^2_{\mathrm{tot}} \,, \quad \gamma_{\mathrm{tot}} > 0. \tag{19}$$

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

- 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

$$\textbf{norm test} \quad V^2_{\text{tot}}(x_n) := \frac{\mathbb{E}\left[|\epsilon_{S_n}(x_n)|^2\right]}{|\nabla F(x_n)|^2} \leq \gamma^2_{\text{tot}}\,, \quad \gamma_{\text{tot}} > 0. \tag{19}$$

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,

$$\textbf{inner product test} \quad V^2_{\|}(x_n) := \frac{\mathbb{E}\left[\langle \epsilon_{S_n}(x_n), \nabla F(x_n)\rangle^2\right]}{|\nabla F(x_n)|^4} \leq \gamma^2_{\|}\,, \quad \gamma_{\|} > 0. \tag{20}$$

- 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,

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

- 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 $\kappa(n) = V_{\|}^2(x_n)/\gamma_{\|}^2$.

- 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.

- 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

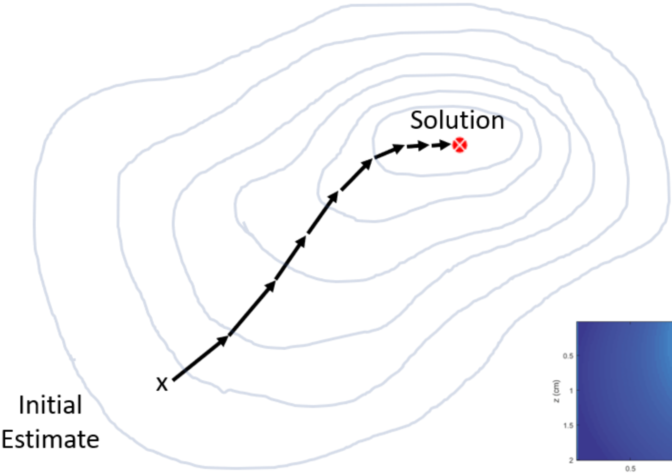
- If we are bounding the error in the sampled gradient, e.g. by increasing the sample size, then fixed step SGD may converge so long as the following is satisfied for all n [Bollapragada, Byrd, Nocedal 2018].

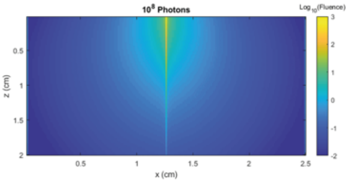
$$\alpha_n \leq \frac{1}{(1+\gamma_{\mathrm{tot}}^2)L} \quad ,$$

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

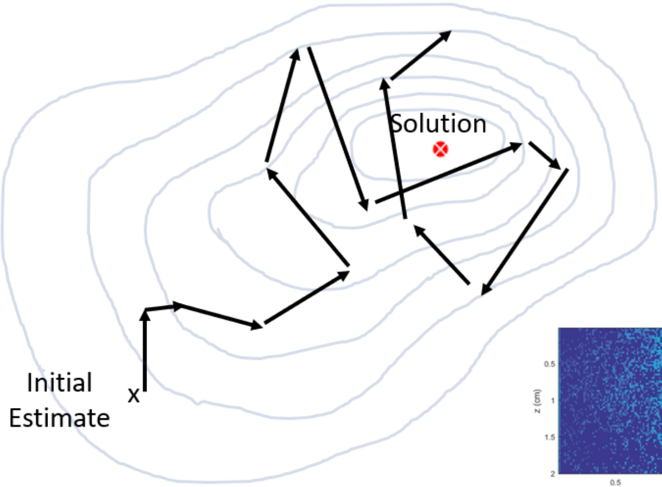
- As intuition indicates, when the sample size (e.g number of simulated photons) increases towards the maximum number of samples $|S_n| \to |S_{\max}|$ ($|S_{\max}| = \infty$ in the case of Monte Carlo RTE simulations), the expected error in the sampled gradient approaches zero, $|\epsilon_{S_n}| \to 0$, as do the measures of variance in the sampled gradients ($V_{\mathrm{tot}}^2 \to 0$, $V_{\parallel}^2 \to 0$), as defined in eq. 19 and eq. 20. 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].

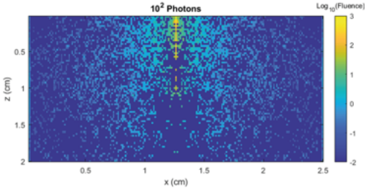
FSR : Introduction
Gradient Descent





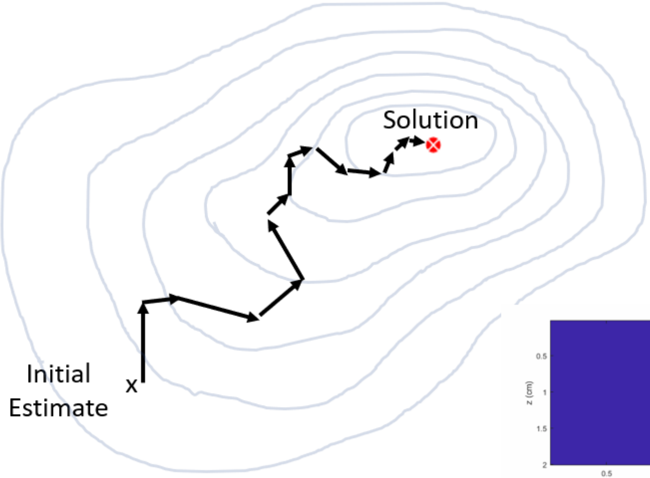
FSR : Introduction
Stochastic Gradient Descent

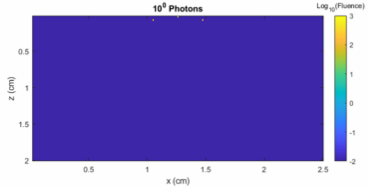




FSR : Introduction

Adaptive Stochastic Gradient Descent



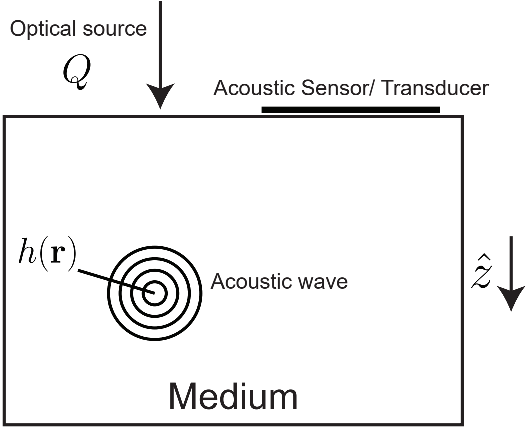


Algorithm 1 Inversion using Monte Carlo sampled gradients with adaptive sample size

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 $\mathcal{L}_{\mathrm{MC}}^{-1} Q \mapsto \phi,\ \Phi$, using $|S_n|/2$ photons
2. Construct internal adjoint source $Q_{\mathrm{adj}} = \mu_{\mathrm{a}}(h^{\mathrm{obs}} - h)$
3. Compute $\mathcal{L}_{\mathrm{MC}}^{-1*} Q_{\mathrm{adj}} \mapsto \phi^*,\ \Phi^*$, 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 $|S_{n+1}| = |S_n|$.

FSR : Results

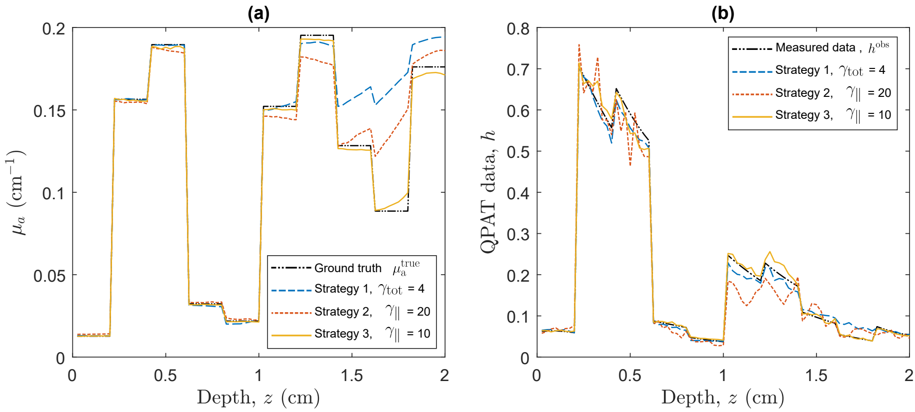


Figure: QPAT inversion: (a) - Ground truth absorption distribution, μ_a^{true}, and estimated absorption distribution μ_a at the point where the photon budget is expended, using each of the three strategies with the stated values of γ_{tot} or $\gamma_{\|}$. (b) - Associated measured data from ground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

FSR : Results

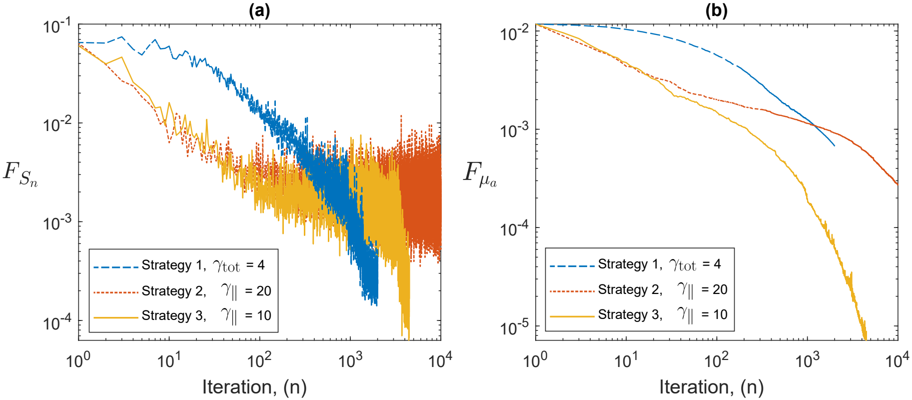


Figure: QPAT inversion: (a) - Sampled cost function, F_{S_n}, as a function of iteration, n. (b) - Error in absorption estimate, F_{μ_a}, as a function of iteration, n.

FSR : Results

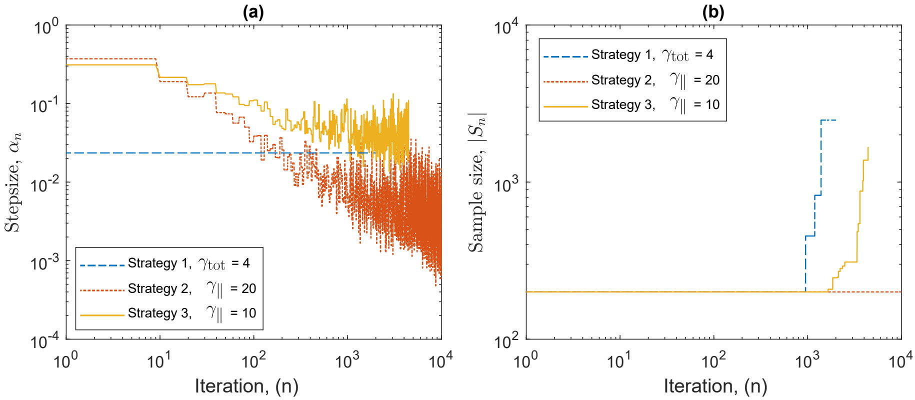
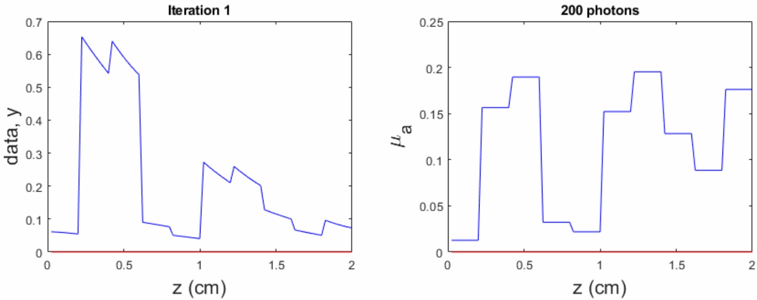


Figure: QPAT inversion: (a) - Step sizes, α_n, as a function of iteration, n. (b) - Adaptive sample size, $|S_n|$, as a function of iteration.

FSR : Results

Test QPrecon movies

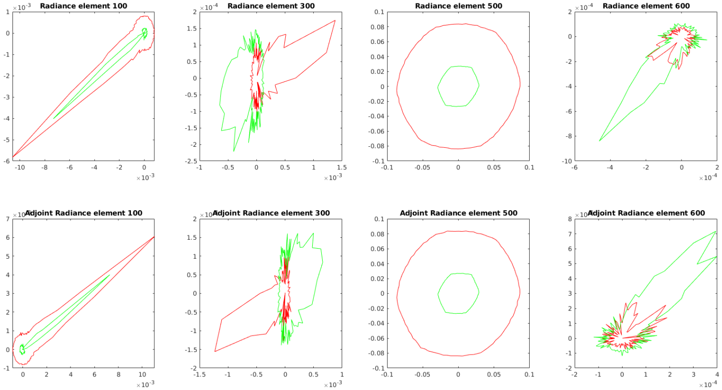


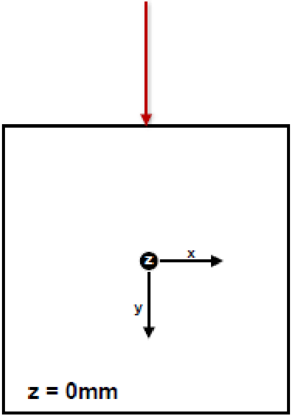
- Radiance Monte Carlo (RMC) is an extension to conventional MC to allow storage and calculation of radiance.
- Storage of phton weights and directions at all points in space is prohibitive in terms of storage
- solution : store spherical harmonic cofficients $c_{j,m,n} = \langle Y_{n,m}, \phi \rangle_{S^2}$
- Integrals on unit shere "naturally" estimated by MC samples.
- 2D : use Fourier coefficients

Adjoint Issues :
Solution to adjoint RTE is reversal of first order dervative $\Rightarrow$ store -1^n for radial coefficients

Radiance Monte Carlo

Radiance Monte Carlo 2D example

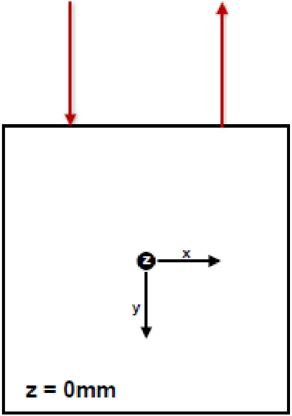




$\mu_a = 0.01\ \mathrm{mm}^{-1}$

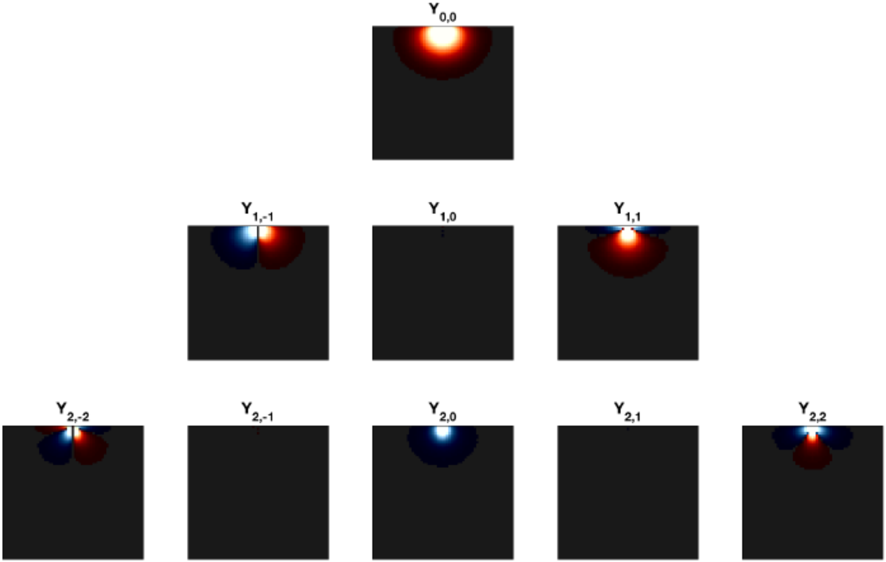
$\mu_s = 5.00\ \mathrm{mm}^{-1}$

$g = 0.8$



Radiance Monte Carlo

Radiance Monte Carlo 3D example



Radiance Monte Carlo

Radiance Monte Carlo 3D example

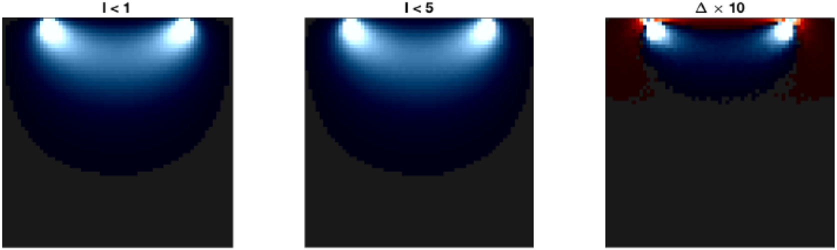
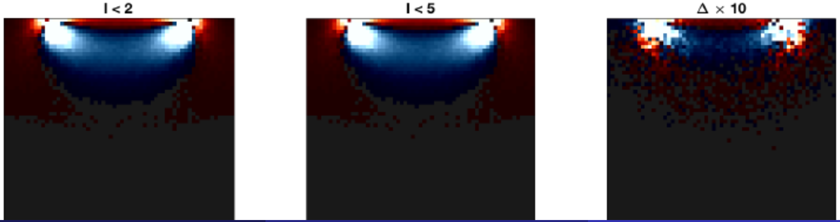
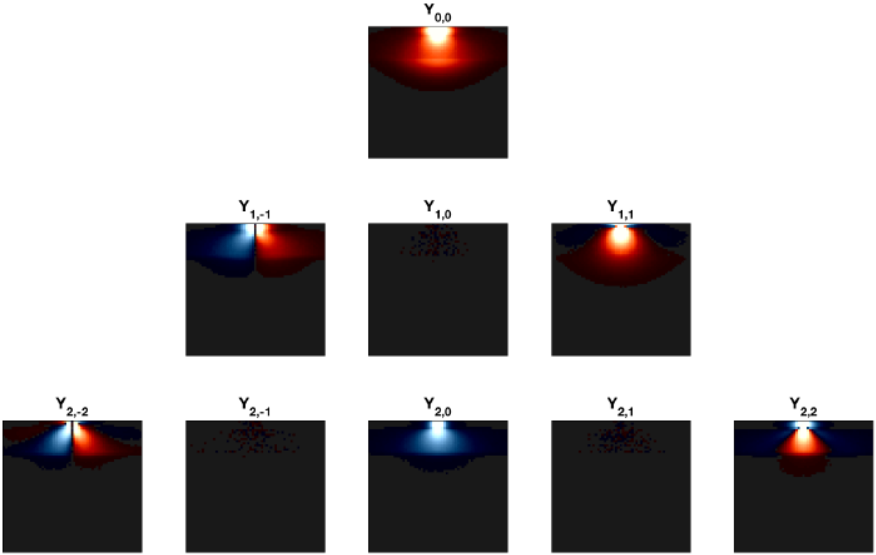


Figure 3. Sensitivity functions for the absorption coefficient in an homogenous cube, built using spherical harmonic functions of degree $\ell \leq 1$ (left) and $\ell \leq 4$ (centre), and the difference between the two results (right). The scale of the difference plot is multiplied by ten relative to the sensitivity functions



Radiance Monte Carlo

Radiance Monte Carlo 3D example - with void



Radiance Monte Carlo

Radiance Monte Carlo 3D example - with void

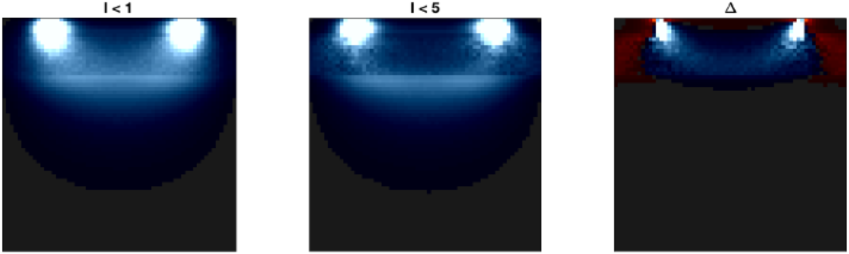
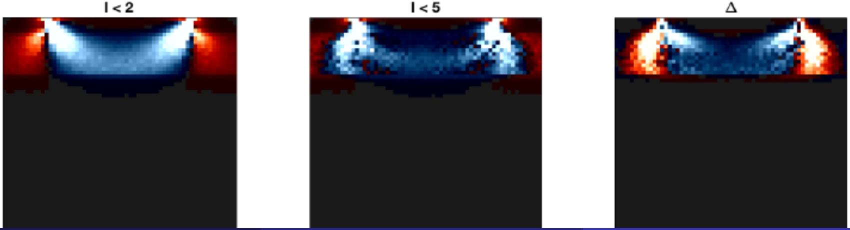
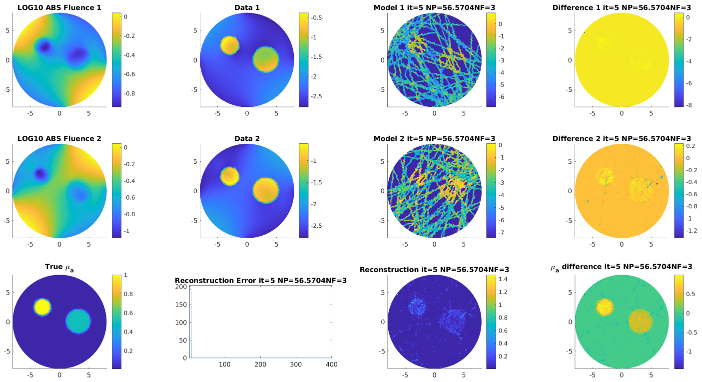


Figure 7. Sensitivity functions for the absorption coefficient in a cube with a void region, built using spherical harmonic functions of degree $\ell \leq 1$ (left) and $\ell \leq 4$ (centre), and the difference between the two results (right). The colour scale is identical for all plots.

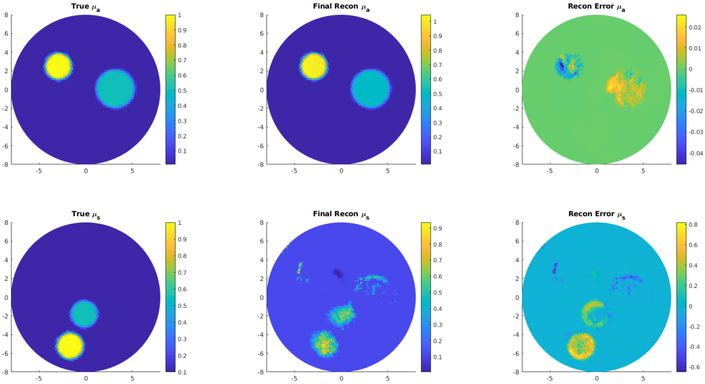


FSR : Results (2D) : QPAT



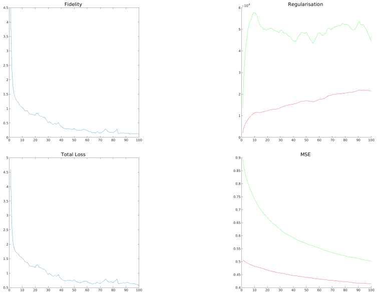
FSR : Results (2D)

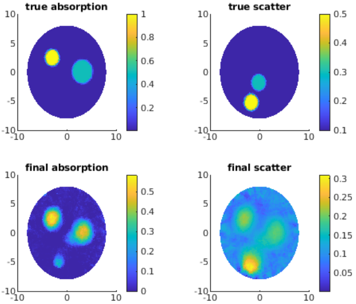
QPAT joint recon



FSR : Results (2D)

DOT joint recon

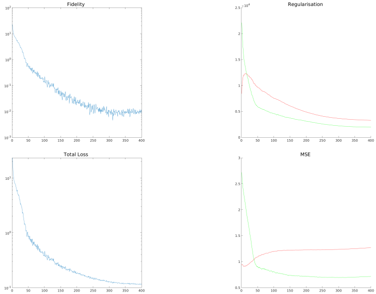


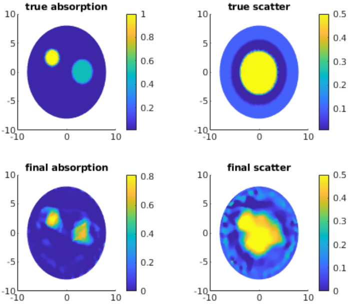


N Photons : 10^4, frequency $200\mathrm{MHz}$, NFourier $3 \to 13$.

FSR : Results (2D)

DOT joint recon : ring void case





N Photons : $5.10^3 \rightarrow 1.10^6$, frequency 1000$\mathrm{MHz}$, NFourier 3.

Outline

- Inverse Problems based on Radiative Transport Equation
- Several methods for modelling
- Stochastic models of light propagation combined with stochastic optimisation : Fully Stochastic Inversion

- 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
- Posterior sampling using MC^3.