Inverse Problems **25** (2009) 123010 (59pp) doi:10.1088/0266-5611/25/12/123010

TOPICAL REVIEW

Optical tomography: forward and inverse problems

Simon R Arridge[1] **and John C Schotland**[2]

[1] Department of Computer Science, University College London, Gower Street, London WC1E 6BT, UK
[2] Department of Bioengineering and Graduate Group in Applied Mathematics and Computational Science, University of Pennsylvania, Philadelphia, PA 19104, USA

E-mail: s.arridge@cs.ucl.ac.uk and schotland@seas.upenn.edu

Received 13 July 2009
Published 3 December 2009
Online at stacks.iop.org/IP/25/123010

Abstract
This is a review of recent mathematical and computational advances in optical tomography. We discuss the physical foundations of forward models for light propagation on microscopic, mesoscopic and macroscopic scales. We also consider direct and numerical approaches to the inverse problems that arise at each of these scales. Finally, we outline future directions and open problems in the field.

1. Introduction

Optical tomography is a biomedical imaging modality that uses scattered light as a probe of structural variations in the optical properties of tissue. In a typical experiment, a highly scattering medium is illuminated by a narrow collimated beam and the light that propagates through the medium is collected by an array of detectors. There are many variants of this basic scenario. For instance, the source may be pulsed or time harmonic, coherent or incoherent, and the illumination may be spatially structured or multispectral. Likewise, the detector may be time or frequency resolved, polarization or phase sensitive, located in the near- or far-field and so on. The inverse problem that is considered is to reconstruct the optical properties of the medium from boundary measurements. The mathematical formulation of the corresponding forward problem is dictated primarily by *spatial scale*, ranging from the Maxwell equations at the microscale, to the radiative transport equation at the mesoscale, and to diffusion theory at the macroscale. In addition, experimental time scales vary from the femtosecond on which light pulses are generated, through the nanosecond on which diffuse waves propagate, to the millisecond scale on which biological activation takes place and still longer for pathophysiologic changes. In this review, we discuss the mathematical structure and computational approaches to the forward and inverse problems that arise at each of these scales.

The state of progress on the inverse problem of optical tomography was reviewed a decade ago in this journal [1]. Since then, the field has grown out of all proportion. It is almost impossible to comprehensively summarize the numerous new measurement systems, applications, algorithms, physical approximations and theoretical results. Nevertheless, in this review we attempt to give an overview of those advances that have proven to be most enduring, and to enumerate those topics which remain unresolved and which represent the most fruitful areas for forthcoming research.

We have structured the review as a tutorial, emphasizing aspects of wave propagation in random media, the mathematical structure of related inverse problems, and computational tool for image reconstruction. The selection of topics is to some extent personal and reflects the tastes and biases of the authors. We do not attempt to summarize the considerable body of work that concerns instrumentation or biomedical applications of optical tomography. For some reviews of these aspects we refer the reader to several articles [2–8]. Nevertheless, it is important to note that there have been significant advances on the experimental front that have motivated and impacted on the corresponding theoretical developments. These include the following:

- The introduction of noncontact imaging systems wherein a scanned beam and a lens-coupled CCD is employed to replace the illumination and detection fiber-optics of conventional optical tomography experiments [9–12]. Such systems generate data sets that are orders of magnitude larger than those acquired with fiber-based systems, leading to significant computational challenges for image reconstruction.
- The development of continuously tunable high-power pulsed light sources has led to increased attention to the multispectral aspects of optical tomography.
- The substantial growth in the availability of targeted fluorescent probes for molecular imaging has stimulated the development of computationally efficient algorithms for fluorescence optical tomography.

In addition, from a theoretical point of view, the rapid progress in the analysis of optical phenomena at smaller scales has allowed the field to move from simply considering diffuse transport as the model of light propagation to include more general radiative transport, coherence and polarization effects.

The review is organized as follows. In section 2 we introduce some general concepts and an abstract framework within which it is possible to describe the class of inverse problems that arise in optical tomography. In section 3 we present the essential physical ideas needed to discuss the propagation of light in a random medium. We also sketch the scattering theory of diffuse waves within radiative transport theory. In section 4 we define the various imaging modalities that we will study using the framework established in section 2. In section 5 we introduce numerical methods for computing Green's functions for the RTE and diffusion equation in the context of forward modeling. In section 6 we describe direct reconstruction methods and associated fast algorithms for several inverse problems in optical tomography. Such methods are well suited to reconstructing images from the large data sets that are available in noncontact optical tomography systems. In section 7 we focus on numerical inversion methods using the tools of optimization theory and Bayesian statistics. In section 8 we introduce the basic concepts of shape-based methods for image reconstruction. In section 9 we survey some remaining topics including reconstruction of anisotropic and time-varying optical parameters.

There are a number of topics that we do not discuss. These include thermoacoustic and photoacoustic tomography, acousto-optic imaging, coherent imaging and nanoscale optical

tomography. Lack of space or expertise, rather than interest, prohibit the authors from treating these important subjects.

2. A general framework

In this section we introduce some general concepts and notation for the forward and inverse problems occurring in optical tomography. Since the various models are described in terms of different physical quantities (space, time or frequency, direction, wavelength), we attempt some abstract definitions in terms of a *generalized coordinate* that we denote by ξ.

2.1. Forward problems

We consider a domain Ω with boundary $\partial\Omega$. To allow for generalized coordinates we define the generalized domains Ξ with boundary $\partial\Xi$. Experiments are performed by applying inward directed photon currents $J^-(\xi_s), \xi_s \in \partial\Xi$ and measuring outgoing photon currents $J^+(\xi_m), \xi_m \in \partial\Xi$. In the 'non-contact' application, these functions are applied and/or measured by free-space propagation from $\partial\Xi$ to an exterior surface Σ_{ext}. The propagation of light inside the domain is governed by a differential operator $\mathcal{L}(x)$ parameterized by functions $x(\boldsymbol{r}), \boldsymbol{r} \in \Omega$. Let X denote the function space of these parameters. In most cases the inverse problem is a parameter estimation problem for $x \in \mathrm{X}$.

Let $\mathbb{S}$ represent a (possibly infinite) set of source functions on $\partial\Xi$ or Σ_{ext}. Let Q represent the space spanned by the functions in $\mathbb{S}$. Then for a particular applied source current $J_i^- \in \mathbb{S}$ the measurable current is found by solving

$$\mathcal{L}(x)U_i = 0 \tag{2.1}$$

$$\mathcal{B}^- U_i = J_i^- \tag{2.2}$$

$$J_i^+ = \mathcal{B}^+ U_i \tag{2.3}$$

where $\mathcal{B}^-, \mathcal{B}^+$ are the boundary conditions obeyed by $\mathcal{L}$. We assume that in all cases considered herein, these boundary conditions are specific enough to ensure uniqueness of U_i. The measurable currents belong to a space Z.

In the case where J_i^- is a δ-function located at ξ_s the solution U_i defines *Green's function* $G(\xi, \xi_s)$, and we define the linear Green's operator

$$U_i = \mathcal{G}(x)J_i^- = \int_{\partial\Xi} G(\xi, \xi_s)J_i^-(\xi_s)\,\mathrm{d}\xi_s, \tag{2.4}$$

where $\mathrm{d}\xi_s$ is the surface measure on $\partial\Xi$. Note that in (2.4) we have made the dependence of $\mathcal{G}(x)$ on x explicit, whereas in Green's function G it is only implicit.

Definitions (2.1)–(2.3) allow the definition of the mappings

$$\text{Transfer function} : \Lambda : \mathrm{Q} \to \mathrm{Z}, \qquad J_i^+ = \Lambda(x)J_i^- = \mathcal{B}^+\mathcal{G}(x)J_i^- \tag{2.5}$$

$$\text{Forward Map} : F : \mathrm{X} \to \mathrm{Z}, \qquad J_i^+ = F_i(x). \tag{2.6}$$

The (nonlinear) forward map is specified for a particular source pattern. Let $\mathbb{F}_{\mathbb{S}} = \{F_i, J_i^- \in \mathbb{S}\}$ be the (possibly infinite) set of all forward mappings for the set of source currents $\mathbb{S}$.

Measured data are obtained as the result of a measurement operator

$$\mathcal{M} : \mathrm{Z} \to \mathrm{Y}, \tag{2.7}$$

where Y is a (possibly infinite) vector space. Combining the forward map and measurement operator defines a forward operator

$$\mathcal{F} = \mathcal{M}F : \mathrm{X} \to \mathrm{Y}. \tag{2.8}$$

We introduce the concept of an 'aperture function' as an arbitrary element $w_j \in \mathrm{Z}$ of the measurement space, which physically can be used to model the finite dimension and angular acceptance of a detector. Let $\mathbb{M}$ represent a (possibly infinite) set of aperture functions supported on $\partial\Xi$ or Σ_{ext} and spanning Z. Then a particular datum is obtained by applying a given aperture $w_j \in \mathbb{M}$ to the measurable quantity

$$y_{j,i} = \mathcal{M}_j J_i^+ = \left\langle w_j, J_i^+ \right\rangle_{\partial\Xi} = \langle w_j, F_i(x)\rangle_{\partial\Xi} = \left\langle w_j, \mathcal{B}^+\mathcal{G}(x)J_i^- \right\rangle_{\partial\Xi}. \tag{2.9}$$

For the particular case of a δ-function source we have

$$y_{j,\delta(\xi_{\mathrm{s}})} = \langle w_j, \mathcal{B}^+ G(\xi_{\mathrm{m}}, \xi_{\mathrm{s}})\rangle_{\partial\Xi}. \tag{2.10}$$

Similarly, if the measurements are simply point samples, we can consider

$$y_{\delta(\xi_{\mathrm{m}}),\delta(\xi_{\mathrm{s}})} = \mathcal{B}^+ G(\xi_{\mathrm{m}}, \xi_{\mathrm{s}}). \tag{2.11}$$

2.2. Linearization

The key tool for the study of the inverse problem is the linearization of the forward map F. Consider a perturbation in the parameters $x(\boldsymbol{r}) \to x(\boldsymbol{r}) + \eta(\boldsymbol{r})$. The Fréchet derivative of the forward mapping is the linear mapping defined by

$$F_i'(x)\eta = F_i(x+\eta) - F_i(x) + o(\|\eta\|^2). \tag{2.12}$$

When the Fréchet derivative is projected onto an aperture function we define the change in measurement (for a particular perturbation η)

$$y'_{j,i} = \left\langle w_j, F_i'(x)\eta \right\rangle_{\partial\Xi} = \left\langle F_i'^*(x)w_j, \eta \right\rangle_{\Xi} \tag{2.13}$$

which serves as the definition of the *adjoint* Fréchet derivative. The adjoint Fréchet derivative will involve the solution of the adjoint problem

$$\mathcal{L}^*(x)U_j^* = 0 \tag{2.14}$$

$$\mathcal{B}^{+*}U_j^* = w_j. \tag{2.15}$$

It is convenient to define another notation for the Fréchet derivative as a general linear integral operator

$$F_i'\eta = \mathcal{K}_i\eta = \int_{\Xi} K_i(\xi_{\mathrm{m}}, \xi)\eta(\boldsymbol{r})\,\mathrm{d}\xi, \tag{2.16}$$

with kernel K_i and $\mathrm{d}\xi$ the volume integral measure on Ξ.

From the Fréchet derivative of the forward mapping we are lead to a definition of the Fréchet derivative of Green's operator

$$\begin{aligned}\mathcal{K}_i\eta = \mathcal{B}^+\mathcal{G}'\left[J_i^- \otimes \eta\right] &= \int_{\partial\Xi}\int_{\Xi} \mathcal{B}^+ G'(\xi_{\mathrm{m}}, \xi, \xi_{\mathrm{s}})\eta(\xi)J^-(\xi_{\mathrm{s}})\,\mathrm{d}\xi\,\mathrm{d}\xi_{\mathrm{s}} \\ &= \langle\langle \mathcal{B}^+ G', \eta\rangle_{\Xi}, J^-\rangle_{\partial\Xi}.\end{aligned} \tag{2.17}$$

The operator $\mathcal{G}'$ has kernel $G'(\xi_{\mathrm{m}}, \xi, \xi_{\mathrm{s}})$ whose form depends on the particular problem considered. We may define in general terms a potential operator

$$\mathcal{V}(\eta) = \mathcal{L}(x+\eta) - \mathcal{L}(x), \tag{2.18}$$

which leads to

$$\mathcal{G}' = -\mathcal{G}\mathcal{V}\mathcal{G}, \qquad \text{and} \qquad \mathcal{G}'^{*} = -\mathcal{G}^{*}\mathcal{V}\mathcal{G}. \tag{2.19}$$

These definitions will be made more precise in section 6.

For the particular case of δ-function source and sampling measurements $w_j = \delta(\xi_{\mathrm{m}})$ the change in measurement given by (2.13) is simply the inner product of G' with the perturbation

$$y'_{\delta(\xi_{\mathrm{m}}),\delta(\xi_{\mathrm{s}})} = \langle \mathcal{B}^{+} G'(\xi_{\mathrm{m}}, \xi, \xi_{\mathrm{s}}), \eta(\boldsymbol{r}) \rangle_{\Xi} = \langle \rho, \eta \rangle_{\Omega}\,, \tag{2.20}$$

which defines the *sensitivity function* or *measurement density function* $\rho(\boldsymbol{r})$. More generally, the photon measurement density functions are the projection of the derivative Green's function onto a given source and measurement aperture

$$\rho_{j,i}(\boldsymbol{r}) = \int_{\partial\Xi} w_j(\xi_{\mathrm{m}}) \int_{\partial\Xi} J_i^{-}(\xi_{\mathrm{s}}) G'(\xi_{\mathrm{m}}, \xi, \xi_{\mathrm{s}})\, \mathrm{d}\xi_{\mathrm{s}}\, \mathrm{d}\xi_{\mathrm{m}}. \tag{2.21}$$

For the case of finite source and aperture functions, the forward mapping $\mathbb{F}_{\mathbb{S}}$ and all its Fréchet derivatives are continuous-to-discrete, and the adjoint mappings are discrete-to-continuous; such mappings necessarily contain an infinite dimensional null-space. When considering an infinite set of source and aperture functions, the forward mapping does not contain a null-space but is bounded, leading to an inverse problem that is unbounded and ill-posed.

For computational purposes, the parameters of the inverse problem are considered in a basis representation

$$x(\boldsymbol{r}) = \sum_k x_k b_k(\boldsymbol{r}). \tag{2.22}$$

This allows the definition of a discrete-to-continuous linear mapping

$$F_i' \eta = \mathcal{K}_i \eta = \sum_k \eta_k \int_{\Xi} K_i(\xi_{\mathrm{m}}, \xi) b_k(\boldsymbol{r})\, \mathrm{d}\xi. \tag{2.23}$$

Considering a finite set of measurement functions of dimension n_i and a finite set of aperture functions of dimension n_j leads to a vector of measurements of size $n_i \times n_j$. Taking the basis representation (2.22) leads to a discrete matrix representation of the linearized problem

$$\delta \boldsymbol{y} = \mathsf{A}\eta, \tag{2.24}$$

where A has matrix elements

$$\int_{\Omega} \rho_{j,i}(\boldsymbol{r}) b_k(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r} = \int_{\partial\Xi} w_j(\xi_{\mathrm{m}}) \int_{\partial\Xi} J_i^{-}(\xi_{\mathrm{s}}) \int_{\Xi} G'(\xi_{\mathrm{m}}, \xi, \xi_{\mathrm{s}}) b_k(\boldsymbol{r})\, \mathrm{d}\xi\, \mathrm{d}\xi_{\mathrm{s}}\, \mathrm{d}\xi_{\mathrm{m}}. \tag{2.25}$$

This matrix is typically referred to as the *Jacobian*, *weight matrix* or *sensitivity matrix* for the linearized inverse problem.

The adjoint operation

$$\eta = \mathsf{A}^{\mathrm{T}} \delta \boldsymbol{y} \tag{2.26}$$

can be considered as the discretization of the adjoint operator

$$\eta = \mathcal{F}_i'^{*} \delta y = \mathcal{K}_i^{*} \mathcal{M}^{*} \delta y = \int_{\partial\Xi} K_i(\xi_{\mathrm{m}}, \xi) \mathcal{M}^{*} \delta y(\boldsymbol{r}_{\mathrm{m}})\, \mathrm{d}\xi_{\mathrm{m}} \tag{2.27}$$

$$= \mathcal{G}'^{*} \mathcal{B}^{+*} \left[J_i^{-} \otimes \delta y \right] = \left\langle \langle G'^{*}, \mathcal{B}^{+*} \delta y \rangle_{\partial\Xi}, J_i^{-} \right\rangle_{\partial\Xi}. \tag{2.28}$$

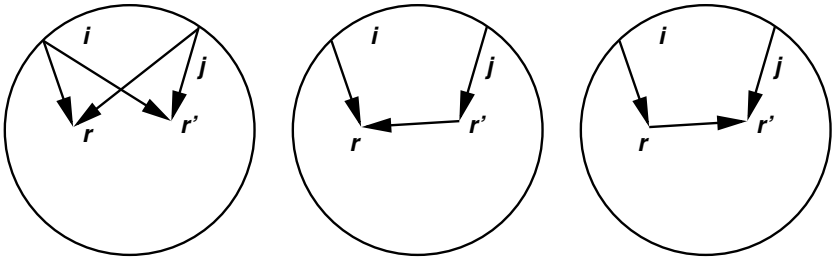

Figure 1. Second-order derivative operators. On the left is the kernel of $\mathcal{K}^*\mathcal{K}$. In the centre the term $\mathcal{G}^*\mathcal{V}\mathcal{G}\mathcal{V}\mathcal{G}$. On the right the term $\mathcal{G}^*\mathcal{V}\mathcal{G}^*\mathcal{V}\mathcal{G}$.

2.3. Inverse problems

Our goal is to invert the forward map F and recover the parameters x. To proceed, we note that the adjoint Fréchet derivative operator defined in (2.27) acts on an element in data space to give an element in parameter space, representing a change in x that reduces the norm of the residual difference between the data and the forward operator acting on x. In terms of optimization theory this direction is the steepest descent update direction for the inverse problem in the sense that it gives the direction in which this residual norm decreases locally most rapidly. This operator is therefore the basis for gradient-based optimization methods. We explore this idea in more detail in section 7.1.

Solution methods for inverse problems typically perform better by making use of second derivative information. The second Fréchet derivative is given by

$$F_i''(x)(\eta_1, \eta_2) = F_i(x + \eta_1 + \eta_2) - F_i'(x)(\eta_1 + \eta_2) - F_i(x) + o(\|\eta\|^3), \tag{2.29}$$

with corresponding second derivative Green's operator $\mathcal{G}''$ and kernel G'', given by

$$\mathcal{G}'' = \mathcal{G}\mathcal{V}\mathcal{G}\mathcal{V}\mathcal{G}, \tag{2.30}$$

and in general we can define the nonlinear mapping F as

$$F_i(x + \eta) = \mathcal{B}^+[\mathcal{G}(x) - \mathcal{G}(x)\mathcal{V}(\eta)\mathcal{G}(x) + \cdots + (-1)^n\mathcal{G}(x)\,(\mathcal{V}(\eta)\mathcal{G}(x))^n + \cdots]J_i^- \tag{2.31}$$

which is the Born series. We discuss this series in more detail in (3.9) and (3.86) in section 3.

For optimization methods the second Fréchet derivative gives rise to a Hessian operator

$$\mathcal{H} = \mathcal{K}^*\mathcal{K} - \mathcal{F}''^*\delta y, \tag{2.32}$$

where the adjoint operator $\mathcal{F}''^*$ is composed of two terms

$$\mathcal{G}''^* = \mathcal{G}^*\mathcal{V}\mathcal{G}\mathcal{V}\mathcal{G} + \mathcal{G}^*\mathcal{V}\mathcal{G}^*\mathcal{V}\mathcal{G}, \tag{2.33}$$

The second derivative Green's kernels are represented in figure 1.

Taking the basis representation as in (2.24) leads to a discrete Hessian representation

$$\mathsf{H} = \mathsf{A}^{\mathrm{T}}\mathsf{A} - \mathsf{F}''^{\mathrm{T}}\delta\boldsymbol{y}, \tag{2.34}$$

which forms the basis of the Newton update scheme

$$\mathsf{H}\eta = \mathsf{A}^{\mathrm{T}}\delta\boldsymbol{y}. \tag{2.35}$$

This is discussed in further detail in section 7.3.

3. Waves, transport and diffusion

The mathematical description of light propagation in random media changes according to the length scale of interest [13]. We begin with the Maxwell equations, which are valid on microscopic scales. The mesoscale, which is characterized by the scattering length, is described by the radiative transport equation (RTE). Finally, we discuss the macroscale, which is described by the diffusion approximation to the RTE.

3.1. Electromagnetic waves

The Maxwell equations together with the Lorentz force law govern the classical description of all electromagnetic phenomena. For simplicity, we restrict our attention to nonmagnetic media and to monochromatic fields. We employ the convention that physical quantities are given by the real part of the corresponding complex quantities. In the absence of sources, the electric field $\boldsymbol{E}$ in an inhomogeneous medium with a position-dependent permittivity ε satisfies the time-independent wave equation

$$\nabla \times \nabla \times \boldsymbol{E} - k_0^2 \varepsilon \boldsymbol{E} = 0, \tag{3.1}$$

where k_0 is the free-space wavenumber. In addition, the divergence condition

$$\nabla \cdot \varepsilon \boldsymbol{E} = 0 \tag{3.2}$$

must be obeyed. If ε varies slowly on the scale of the wavelength, then $\boldsymbol{E}$ satisfies the scalar wave equation

$$\nabla^2 U + k_0^2 \varepsilon U = 0, \tag{3.3}$$

where U is any of the components of $\boldsymbol{E}$. Note that the components of $\boldsymbol{E}$ are still coupled according to (3.2). If we restrict our attention to (3.3) we will say that we are working within the scalar theory of electromagnetic fields.

The conservation of energy is governed by the relation

$$\nabla \cdot \boldsymbol{J} + \frac{4\pi k_0}{c} \operatorname{Im}(\varepsilon) \bar{I} = 0, \tag{3.4}$$

where the energy current density J is defined by

$$\boldsymbol{J} = \frac{1}{2\mathrm{i}k_0}(U^* \nabla U - U \nabla U^*) \tag{3.5}$$

and the intensity $\bar{I}$ is given by

$$\bar{I} = \frac{c}{4\pi} |U|^2. \tag{3.6}$$

Note that $\boldsymbol{J}$ plays the role of the Poynting vector in the scalar theory.

The solution to (3.3), which obeys the Sommerfeld radiation condition, is given by the Lippmann–Schwinger equation

$$U(\boldsymbol{r}) = U_{\mathrm{inc}}(\boldsymbol{r}) + k_0^2 \int \mathrm{d}\boldsymbol{r}'\, G(\boldsymbol{r}, \boldsymbol{r}') U(\boldsymbol{r}') \eta(\boldsymbol{r}'), \tag{3.7}$$

where $\eta = (\varepsilon - 1)/4\pi$ is the dielectric susceptibility. Here U_{inc} is the incident field, which obeys (3.3) with $\eta = 0$ and G is Green's function which, in free space, is given by

$$G(\boldsymbol{r}, \boldsymbol{r}') = \frac{\mathrm{e}^{\mathrm{i}k_0|\boldsymbol{r}-\boldsymbol{r}'|}}{|\boldsymbol{r} - \boldsymbol{r}'|}. \tag{3.8}$$

If we iterate (3.7) starting from $U = U_{\text{inc}}$, we obtain an infinite series for U of the form

$$U(\boldsymbol{r}) = U_{\text{inc}}(\boldsymbol{r}) + k_0^2 \int \mathrm{d}\boldsymbol{r}'\, G(\boldsymbol{r}, \boldsymbol{r}')\eta(\boldsymbol{r}')U_{\text{inc}}(\boldsymbol{r}') + k_0^4 \int \mathrm{d}\boldsymbol{r}'\, \mathrm{d}\boldsymbol{r}''\, G(\boldsymbol{r}, \boldsymbol{r}')\eta(\boldsymbol{r}')G(\boldsymbol{r}', \boldsymbol{r}'')\eta(\boldsymbol{r}'')U_{\text{inc}}(\boldsymbol{r}'') + \cdots. \tag{3.9}$$

Equation (3.9) is known as the Born series (see (2.31)) and each of its terms corresponds to successively higher orders of scattering of the incident field. If only the first term in the series is retained, this is known as the Born approximation.

3.2. Radiative transport

In radiative transport theory the propagation of light through a material medium is formulated in terms of a conservation law that accounts for gains and losses of photons due to scattering and absorption [14, 15]. The fundamental quantity of interest is the specific intensity $I(\boldsymbol{r}, \hat{\mathbf{s}})$, defined as the intensity at the position $\boldsymbol{r}$ in the direction $\hat{\mathbf{s}}$. The specific intensity obeys the radiative transport equation (RTE):

$$\hat{\mathbf{s}} \cdot \nabla I + (\mu_a + \mu_s)I = \mu_s \int p(\hat{\mathbf{s}}', \hat{\mathbf{s}})I(\boldsymbol{r}, \hat{\mathbf{s}}')\, \mathrm{d}\hat{\mathbf{s}}', \qquad \boldsymbol{r} \in \Omega, \tag{3.10}$$

which we have written in its time-independent form. Here μ_a and μ_s are the absorption and scattering coefficients and p is the phase function. The specific intensity also satisfies the half-range boundary condition

$$I(\boldsymbol{r}, \hat{\mathbf{s}}) = J^-(\boldsymbol{r}, \hat{\mathbf{s}}), \qquad \hat{\mathbf{s}} \cdot \hat{\nu} < 0, \qquad \boldsymbol{r} \in \partial\Omega, \tag{3.11}$$

where $\hat{\nu}$ is the outward unit normal to $\partial\Omega$ and J^- is the incident specific intensity at the boundary. The above choice of boundary condition guarantees the uniqueness of solutions to the RTE [15]. The phase function p is symmetric with respect to interchange of its arguments and obeys the normalization condition

$$\int p(\hat{\mathbf{s}}, \hat{\mathbf{s}}')\, \mathrm{d}\hat{\mathbf{s}}' = 1, \tag{3.12}$$

for all $\hat{\mathbf{s}}$. We will often assume that $p(\hat{\mathbf{s}}, \hat{\mathbf{s}}')$ depends only upon the angle between $\hat{\mathbf{s}}$ and $\hat{\mathbf{s}}'$, which holds for scattering by spherically symmetric particles. Note that the choice $p = 1/(4\pi)$ corresponds to isotropic scattering. For scattering that is strongly peaked in the forward direction ($\hat{\mathbf{s}} \cdot \hat{\mathbf{s}}' \approx 1$), an asymptotic expansion of the right-hand side of (3.10) may be performed [16]. This leads to the Fokker–Planck form of the RTE

$$\hat{\mathbf{s}} \cdot \nabla I + (\mu_a + \mu_s)I = \mu_s L I, \qquad L = \tfrac{1}{2}(1 - g)\Delta_{\hat{\mathbf{s}}}, \tag{3.13}$$

where $\Delta_{\hat{\mathbf{s}}}$ is the Laplace–Beltrami operator on the two-dimensional unit sphere S^2 and $g \approx 1$ is the anisotropy of scattering, as defined by (3.72) in section 3.5. A rational approximation to the transport term in which the operator L is taken to be of the form

$$L = a\Delta_{\hat{\mathbf{s}}}(1 - b\Delta_{\hat{\mathbf{s}}})^{-1} \tag{3.14}$$

has also been proposed [16, 17]. Here the constants a, b can be computed from angular moments of the phase function. In the case of the Henyey–Greenstein phase function

$$p(\hat{\mathbf{s}} \cdot \hat{\mathbf{s}}') = \frac{1}{4\pi} \sum_{l=0}^{\infty} (2l + 1)g^l P_l(\hat{\mathbf{s}} \cdot \hat{\mathbf{s}}'), \tag{3.15}$$

it can be seen that

$$a = \frac{1}{2}(1 - g)(1 + 2b), \qquad b = \frac{2 - 3g + g^2}{6g(1 - g)}. \tag{3.16}$$

Note that L as given in (3.14) is a bounded operator on $L^2(S^2)$, which is not the case for the spherical Laplacian that appears in (3.13).

The total power P passing through a surface Σ is related to the specific intensity by

$$P = \int_\Sigma \mathrm{d}\boldsymbol{r} \int_{S^2} \mathrm{d}\hat{\mathbf{s}}\, I(\boldsymbol{r}, \hat{\mathbf{s}})\hat{\mathbf{s}} \cdot \hat{\nu}. \tag{3.17}$$

The energy density Φ is obtained by integrating out the angular dependence of the specific intensity:

$$\Phi(\boldsymbol{r}) = \frac{1}{c} \int I(\boldsymbol{r}, \hat{\mathbf{s}})\, \mathrm{d}\hat{\mathbf{s}}. \tag{3.18}$$

We note that the RTE allows for the addition of intensities. As a result, it cannot explain certain wavelike phenomena.

3.2.1. From waves to transport. The RTE can be derived by considering the high-frequency asymptotics of wave propagation in a random medium. We briefly recall the main ideas in the context of monochromatic scalar waves. The general theory for vector electromagnetic waves is presented in [18]. We assume that the random medium is statistically homogeneous and that the susceptibility η is a Gaussian random field such that

$$\langle \eta(\boldsymbol{r}) \rangle = 0, \qquad \langle \eta(\boldsymbol{r})\eta(\boldsymbol{r}') \rangle = C(|\boldsymbol{r} - \boldsymbol{r}'|), \tag{3.19}$$

where C is the two-point correlation function and $\langle \cdots \rangle$ denotes statistical averaging. Let L_0 denote the propagation distance of the wave. At high frequencies, L_0 is large compared to the wavelength and we introduce a small parameter $\epsilon = 1/(k_0 L_0) \ll 1$. We suppose that the fluctuations in η are weak so that C is of the order $O(\epsilon)$. We then rescale the spatial variable according to $\boldsymbol{r} \to \boldsymbol{r}/\epsilon$ and define the scaled field $U_\epsilon(\boldsymbol{r}) = U(\boldsymbol{r}/\epsilon)$, so that (3.3) becomes

$$\epsilon^2 \nabla^2 U_\epsilon(\boldsymbol{r}) + U_\epsilon(\boldsymbol{r}) = -4\pi\sqrt{\epsilon}\,\eta(\boldsymbol{r}/\epsilon)\, U_\epsilon(\boldsymbol{r}). \tag{3.20}$$

Here we have introduced a rescaling of η to be consistent with the assumption that the fluctuations are of strength $O(\epsilon)$.

Although (3.4) gives some indication of how the intensity of the field is distributed in space, it does not prescribe how the intensity propagates. To overcome this difficulty, we introduce the Wigner distribution $W_\epsilon(\boldsymbol{r}, \boldsymbol{k})$, which is a function of the position $\boldsymbol{r}$ and the wave vector $\boldsymbol{k}$:

$$W_\epsilon(\boldsymbol{r}, \boldsymbol{k}) = \int \mathrm{d}\boldsymbol{R}\, \mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot\boldsymbol{R}} U_\epsilon\left(\boldsymbol{r} - \frac{1}{2}\epsilon\boldsymbol{R}\right) U_\epsilon^*\left(\boldsymbol{r} + \frac{1}{2}\epsilon\boldsymbol{R}\right). \tag{3.21}$$

The Wigner distribution has several important properties. It is real valued and is related to the intensity and energy current density by the formulas

$$\bar{I} = \frac{c}{4\pi} \int \frac{\mathrm{d}\boldsymbol{k}}{(2\pi)^3} W_\epsilon(\boldsymbol{r}, \boldsymbol{k}), \qquad \boldsymbol{J} = \int \frac{\mathrm{d}\boldsymbol{k}}{(2\pi)^3} \boldsymbol{k} W_\epsilon(\boldsymbol{r}, \boldsymbol{k}). \tag{3.22}$$

Making use of (3.20), it can be seen that the Wigner distribution obeys the equation

$$\boldsymbol{k} \cdot \nabla_r W_\epsilon + \mathrm{i}\frac{2\pi}{\sqrt{\epsilon}} \int \mathrm{d}\boldsymbol{q}\, \mathrm{e}^{-\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{r}/\epsilon} \tilde{\eta}(\boldsymbol{q}) \left(W_\epsilon\left(\boldsymbol{r}, \boldsymbol{k} + \frac{1}{2}\boldsymbol{q}\right) - W_\epsilon(\boldsymbol{r}, \boldsymbol{k} - \frac{1}{2}\boldsymbol{q}) \right) = 0, \tag{3.23}$$

where we have assumed that η is real-valued and $\tilde{\eta}$ denotes the Fourier transform of η which is defined by

$$\tilde{\eta}(\boldsymbol{q}) = \int \mathrm{d}\boldsymbol{r}\, \mathrm{e}^{\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{r}} \eta(\boldsymbol{r}). \tag{3.24}$$

We now consider the asymptotics of the Wigner function in the homogenization limit $\epsilon \to 0$. This corresponds to the regime of high frequencies and weak fluctuations. We proceed by introducing a two-scale expansion for W_ϵ of the form

$$W_\epsilon(\boldsymbol{r}, \boldsymbol{r}', \boldsymbol{k}) = W_0(\boldsymbol{r}, \boldsymbol{k}) + \sqrt{\epsilon} W_1(\boldsymbol{r}, \boldsymbol{r}', \boldsymbol{k}) + \epsilon W_2(\boldsymbol{r}, \boldsymbol{r}', \boldsymbol{k}) + \cdots, \tag{3.25}$$

where $\boldsymbol{r}' = \boldsymbol{r}/\epsilon$ is a fast variable. By averaging over the fluctuations on the fast scale, it is possible to show that $\langle W_0 \rangle$, which we denote by W, obeys the equation

$$\boldsymbol{k} \cdot \nabla_{\boldsymbol{r}} W = \int \mathrm{d}\boldsymbol{k}' \tilde{C}(\boldsymbol{k} - \boldsymbol{k}') \delta(k^2 - k'^2)(W(\boldsymbol{r}, \boldsymbol{k}') - W(\boldsymbol{r}, \boldsymbol{k})). \tag{3.26}$$

Evidently, (3.26) has the form of a time-independent transport equation where W is proportional to the specific intensity I. The role of the delta function is to conserve momentum, making it possible to view W as a function of position and the direction $\hat{\mathbf{s}} = \boldsymbol{k}/|\boldsymbol{k}|$. We note that the phase function and scattering coefficient are related to statistical properties of the random medium, as reflected in the appearance of the correlation function C in (3.26).

The above derivation of the RTE can be modified when the speed of light is not constant. Following [19], the statistically averaged Wigner function can be seen to obey the equation

$$n\boldsymbol{k} \cdot \nabla_{\boldsymbol{r}} W - k_0 \nabla n \cdot (I_3 - \hat{\boldsymbol{k}} \otimes \hat{\boldsymbol{k}}) \nabla_{\hat{\boldsymbol{k}}} W = \int \mathrm{d}\boldsymbol{k}' \tilde{C}(\boldsymbol{k} - \boldsymbol{k}') \delta(k^2 - k'^2)(W(\boldsymbol{r}, \boldsymbol{k}') - W(\boldsymbol{r}, \boldsymbol{k})), \tag{3.27}$$

where n is the index of refraction and I_3 is the identity matrix. Using the relation $I = c(k_0/n)^3 W$ and the above result, we obtain the required generalization of the RTE.

Radiative transport theory can be generalized to account for energy transport for different states of polarization [18, 20]. The RTE is then formulated in terms of the Stokes vector $\mathbf{I} = (I_\parallel, I_\perp, U, V)$. Here the Stokes parameters, which characterize the polarization of the field, are defined as

$$I_\parallel = \langle E_\parallel E_\parallel^* \rangle, \qquad I_\perp = \langle E_\perp E_\perp^* \rangle, \tag{3.28}$$

$$U = \langle E_\parallel E_\perp^* + E_\parallel^* E_\perp \rangle, \qquad V = \mathrm{i}(\langle E_\parallel E_\perp^* - E_\parallel^* E_\perp \rangle), \tag{3.29}$$

where $E_\parallel$, $E_\perp$ denote two orthogonal components of the electric field in the plane perpendicular to the direction of propagation. Note that the total intensity $I = I_\parallel + I_\perp$. The Stokes vector satisfies the generalized RTE

$$\hat{\mathbf{s}} \cdot \nabla \mathbf{I} + (\mu_a + \mu_s)\mathbf{I} = \mu_s \int \mathsf{M}(\hat{\mathbf{s}}', \hat{\mathbf{s}}) \cdot \mathbf{I}(\boldsymbol{r}, \hat{\mathbf{s}}') \, \mathrm{d}\hat{\mathbf{s}}', \tag{3.30}$$

where M denotes the Mueller matrix, which accounts for mixing of different states of polarization upon scattering.

3.3. Collision expansion

The RTE (3.10), obeying the boundary condition (3.11), is equivalent to the integral equation

$$I(\boldsymbol{r}, \hat{\mathbf{s}}) = I_0(\boldsymbol{r}, \hat{\mathbf{s}}) + \int G_0(\boldsymbol{r}, \hat{\mathbf{s}}; \boldsymbol{r}', \hat{\mathbf{s}}') \mu_s(\boldsymbol{r}') p(\hat{\mathbf{s}}', \hat{\mathbf{s}}'') I(\boldsymbol{r}', \hat{\mathbf{s}}'') \, \mathrm{d}\boldsymbol{r}' \, \mathrm{d}\hat{\mathbf{s}}' \, \mathrm{d}\hat{\mathbf{s}}''. \tag{3.31}$$

Here I_0 is the unscattered (ballistic) specific intensity, which satisfies the equation

$$[\hat{\mathbf{s}} \cdot \nabla + \mu_a + \mu_s] I_0 = 0, \tag{3.32}$$

and G_0 is the ballistic Green's function

$$G_0(\boldsymbol{r}, \hat{\mathbf{s}}; \boldsymbol{r}', \hat{\mathbf{s}}') = g(\boldsymbol{r}, \boldsymbol{r}') \delta\left(\hat{\mathbf{s}}' - \frac{\boldsymbol{r} - \boldsymbol{r}'}{|\boldsymbol{r} - \boldsymbol{r}'|}\right) \delta(\hat{\mathbf{s}} - \hat{\mathbf{s}}'), \tag{3.33}$$

where

$$g(\boldsymbol{r},\boldsymbol{r}') = \frac{1}{|\boldsymbol{r}-\boldsymbol{r}'|^2}\exp\left[-\int_0^{|\boldsymbol{r}-\boldsymbol{r}'|}\mu_t\left(\boldsymbol{r}'+\ell\frac{\boldsymbol{r}-\boldsymbol{r}'}{|\boldsymbol{r}-\boldsymbol{r}'|}\right)\mathrm{d}\ell\right], \tag{3.34}$$

and the extinction coefficient $\mu_t = \mu_a + \mu_s$. Note that if a narrow collimated beam produced by a source of power S_0 is incident on the medium at the point $\boldsymbol{r}_0$ in the direction $\hat{\mathbf{s}}_0$, then $I_0(\boldsymbol{r},\hat{\mathbf{s}})$ is given by

$$I_0(\boldsymbol{r},\hat{\mathbf{s}}) = S_0 G_0(\boldsymbol{r},\hat{\mathbf{s}};\boldsymbol{r}_0,\hat{\mathbf{s}}_0). \tag{3.35}$$

To derive the collision expansion, we iterate (3.31) starting from $I^{(0)} = I_0$ and obtain

$$I(\boldsymbol{r},\hat{\mathbf{s}}) = I^{(0)}(\boldsymbol{r},\hat{\mathbf{s}}) + I^{(1)}(\mathbf{r},\hat{\mathbf{s}}) + I^{(2)}(\boldsymbol{r},\hat{\mathbf{s}}) + \cdots, \tag{3.36}$$

where each term of the series is given by

$$I^{(n)}(\boldsymbol{r},\hat{\mathbf{s}}) = \int \mathrm{d}\boldsymbol{r}'\,\mathrm{d}\hat{\mathbf{s}}'\,\mathrm{d}\hat{\mathbf{s}}''G_0(\boldsymbol{r},\hat{\mathbf{s}};\boldsymbol{r}',\hat{\mathbf{s}}')\mu_s(\boldsymbol{r}')p(\hat{\mathbf{s}}',\hat{\mathbf{s}}'')I^{(n-1)}(\boldsymbol{r}',\hat{\mathbf{s}}''), \tag{3.37}$$

with $n = 1, 2, \ldots$. The above series is the analogue of the Born series for the RTE, since each term accounts for successively higher orders of scattering.

It is instructive to examine the expression for $I^{(1)}$, which is the contribution to the specific intensity from single scattering:

$$I^{(1)}(\boldsymbol{r},\hat{\mathbf{s}}) = \int \mathrm{d}\boldsymbol{r}'\,\mathrm{d}\hat{\mathbf{s}}'\,\mathrm{d}\hat{\mathbf{s}}''G_0(\boldsymbol{r},\hat{\mathbf{s}};\boldsymbol{r}',\hat{\mathbf{s}}')\mu_s(\boldsymbol{r}')p(\hat{\mathbf{s}}',\hat{\mathbf{s}}'')I_0(\boldsymbol{r}',\hat{\mathbf{s}}''). \tag{3.38}$$

It follows from (3.35) that the change in intensity $\delta I = I - I^{(0)}$ measured by a point detector at $\boldsymbol{r}_2$ in the direction $\hat{\mathbf{s}}_2$, due to a unit-amplitude point source at $\boldsymbol{r}_1$ in the direction $\hat{\mathbf{s}}_1$ is given by

$$\begin{aligned}\delta I(\boldsymbol{r}_1,\hat{\mathbf{s}}_1;\boldsymbol{r}_2,\hat{\mathbf{s}}_2) = p(\hat{\mathbf{s}}_1,\hat{\mathbf{s}}_2)\int_0^\infty \mathrm{d}R\,R^2 g(\boldsymbol{r}_2,\boldsymbol{r}_1+R\hat{\mathbf{s}}_1)g(\boldsymbol{r}_1+R\hat{\mathbf{s}}_1,\boldsymbol{r}_1)\\ \times\delta\left(\frac{\boldsymbol{r}_2-\boldsymbol{r}_1-R\hat{\mathbf{s}}_1}{|\boldsymbol{r}_2-\boldsymbol{r}_1-R\hat{\mathbf{s}}_1|}-\hat{\mathbf{s}}_2\right)\mu_s(\boldsymbol{r}_1+R\hat{\mathbf{s}}_1).\end{aligned} \tag{3.39}$$

Suppose that $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ are located on the boundary of a bounded domain and $\hat{\mathbf{s}}_1$ points into and $\hat{\mathbf{s}}_2$ points out of the domain. Then the rays in the directions $\hat{\mathbf{s}}_1$ and $\hat{\mathbf{s}}_2$ must intersect at a point $\boldsymbol{R}$ that lies in the interior of the domain. In addition, the delta function in (3.39) implements the constraint that $\hat{\mathbf{s}}_1$, $\hat{\mathbf{s}}_2$ and $\boldsymbol{r}_1-\boldsymbol{r}_2$ all lie in the same plane. Using this fact and carrying out the integration in (3.39), we find that

$$\delta I(\boldsymbol{r}_1,\hat{\mathbf{s}}_1;\boldsymbol{r}_2,\hat{\mathbf{s}}_2) \propto \exp\left[-\int_0^{L_1}\mu_t(\boldsymbol{r}_1+\ell\hat{\mathbf{s}}_1)\,\mathrm{d}\ell - \int_0^{L_2}\mu_t(\boldsymbol{R}+\ell\hat{\mathbf{s}}_2)\,\mathrm{d}\ell\right], \tag{3.40}$$

where $L_1 = |\boldsymbol{R}-\boldsymbol{r}_1|$, $L_2 = |\boldsymbol{R}-\boldsymbol{r}_2|$ and we have omitted overall geometric prefactors. Note that the argument of the exponential corresponds to the integral of μ_t along the broken ray which begins at $\boldsymbol{r}_1$, passes through $\boldsymbol{R}$, and terminates at $\boldsymbol{r}_2$. The significance of such broken rays will be discussed in section 6.3.

The terms in the collision expansion can be classified by their smoothness. The lowest order term is the most singular. In the absence of scattering, according to (3.34), this term leads to a Radon transform relationship between the absorption coefficient and the specific intensity, under the condition that the source and detector are collinear. Inversion of the Radon transform is the basis for optical projection tomography [21, 22]. The first-order term is also singular, as is evident from the presence of a delta function in (3.39). Terms of higher order are of increasing smoothness. This observation has been exploited to prove uniqueness of the

inverse transport problem and to study its stability. A comprehensive review is presented in [23].

The above discussion has implicitly assumed that the angular dependence of the specific intensity is measurable. In practice, such measurements are extremely difficult to obtain. The experimentally measurable intensity is often an angular average of the specific intensity over the aperture of an optical system. The effect of averaging is to remove the singularities that are present in the specific intensity. The resulting inverse problem is then highly ill-posed [24]. To illustrate this point, we observe that it follows from (3.18) and (3.31) that the energy density Φ obeys the integral equation

$$\Phi(\boldsymbol{r}) = \Phi^{(0)}(\boldsymbol{r}) + \frac{1}{4\pi}\int \mathrm{d}\boldsymbol{r}' \frac{\mathrm{e}^{-\mu_t|\boldsymbol{r}-\boldsymbol{r}'|}}{|\boldsymbol{r}-\boldsymbol{r}'|^2} p(\boldsymbol{r}')\Phi(\boldsymbol{r}'), \tag{3.41}$$

where the scattering is assumed to be isotropic and

$$\Phi^{(0)}(\boldsymbol{r}) = \frac{1}{c}\int I^{(0)}(\boldsymbol{r}, \hat{\mathbf{s}})\,\mathrm{d}\hat{\mathbf{s}}. \tag{3.42}$$

Here the kernel appearing in (3.41) is smoothing with a Fourier transform that decays algebraically at high frequencies.

3.4. Transport regime

We now develop the scattering theory for the RTE in an inhomogeneously absorbing medium. We proceed by decomposing μ_a into a constant part $\bar{\mu}_a$ and a spatially varying part $\delta\mu_a$:

$$\mu_a(\boldsymbol{r}) = \bar{\mu}_a + \delta\mu_a(\boldsymbol{r}). \tag{3.43}$$

The RTE (3.10) can be rewritten in the form

$$\hat{\mathbf{s}}\cdot\nabla I + \bar{\mu}_t I - \mu_s \int p(\hat{\mathbf{s}}', \hat{\mathbf{s}}) I(\boldsymbol{r}, \hat{\mathbf{s}}')\,\mathrm{d}\hat{\mathbf{s}}' = -\delta\mu_a(\boldsymbol{r}) I, \tag{3.44}$$

where $\bar{\mu}_t = \bar{\mu}_a + \mu_s$. The solution to (3.44) is given by

$$I(\boldsymbol{r}, \hat{\mathbf{s}}) = I_0(\boldsymbol{r}, \hat{\mathbf{s}}) - \int \mathrm{d}\boldsymbol{r}'\,\mathrm{d}\hat{\mathbf{s}}'\, G(\boldsymbol{r}, \hat{\mathbf{s}}; \boldsymbol{r}', \hat{\mathbf{s}}')\delta\mu_a(\boldsymbol{r}') I(\boldsymbol{r}', \hat{\mathbf{s}}'), \tag{3.45}$$

where G denotes Green's function for a homogeneous medium with absorption $\bar{\mu}_a$ and I_0 is the unperturbed specific intensity. Equation (3.45) is the analogue of the Lippmann–Schwinger equation for the RTE. It describes the 'multiple scattering' of the incident specific intensity from inhomogeneities in $\delta\mu_a$. If only one absorption event is considered, then the intensity I on the right-hand side of (3.45) can be replaced by the incident intensity I_0. This result describes the linearization of the integral equation (3.45) with respect to $\delta\mu_a$. If the incident field is generated by a point source at $\boldsymbol{r}_1$ pointing in the direction $\hat{\mathbf{s}}_1$, then the change in specific intensity due to spatial fluctuations in absorption δI can be obtained from the relation

$$\delta I(\boldsymbol{r}_1, \hat{\mathbf{s}}_1; \boldsymbol{r}_2, \hat{\mathbf{s}}_2) = S_0 \int \mathrm{d}\boldsymbol{r}\,\mathrm{d}\hat{\mathbf{s}}\, G(\boldsymbol{r}_1, \hat{\mathbf{s}}_1; \boldsymbol{r}, \hat{\mathbf{s}}) G(\boldsymbol{r}, \hat{\mathbf{s}}; \boldsymbol{r}_2, \hat{\mathbf{s}}_2)\delta\mu_a(\boldsymbol{r}). \tag{3.46}$$

Here S_0 denotes the source power and $\boldsymbol{r}_2, \hat{\mathbf{s}}_2$ are the position and orientation of a point detector.

To make further progress requires knowledge of Green's function for the RTE. Green's function for an infinite medium with isotropic scattering can be obtained explicitly [15]. In three dimensions we have

$$G(\boldsymbol{r}, \hat{\mathbf{s}}; \boldsymbol{r}', \hat{\mathbf{s}}') = G_0(\boldsymbol{r}, \hat{\mathbf{s}}; \boldsymbol{r}', \hat{\mathbf{s}}') + \frac{\mu_s}{4\pi}\int \frac{\mathrm{d}\boldsymbol{k}}{(2\pi)^3}\,\mathrm{e}^{\mathrm{i}\boldsymbol{k}\cdot(\boldsymbol{r}-\boldsymbol{r}')} \frac{1}{(\mu_t + i\hat{\mathbf{s}}\cdot\boldsymbol{k})(\mu_t + \mathrm{i}\hat{\mathbf{s}}'\cdot\boldsymbol{k})} \times \frac{1}{1 - \frac{\mu_s}{|\boldsymbol{k}|}\tan^{-1}\left(\frac{|\boldsymbol{k}|}{\mu_t}\right)}, \tag{3.47}$$

where G_0 is the ballistic Green's function which is defined by (3.33). More generally, numerical procedures such as the discrete ordinate method or the P_N approximation may be employed to compute Green's function in a bounded domain with anisotropic scattering, as described in section 5.

The method of rotated reference frames is a spectral method for the computing the Green's function for the three-dimensional RTE in a homogeneous medium with anisotropic scattering and planar boundaries [25]. It is derived by considering the plane-wave modes for the RTE which are of the form

$$I(\boldsymbol{r}, \hat{\mathbf{s}}) = A_{\boldsymbol{k}}(\hat{\mathbf{s}})\, \mathrm{e}^{\boldsymbol{k}\cdot\boldsymbol{r}}, \tag{3.48}$$

where the amplitude A is to be determined. Evidently, the components of the wave vector $\boldsymbol{k}$ cannot be purely real; otherwise the above modes would have exponential growth in the $\hat{\boldsymbol{k}}$ direction. We thus consider evanescent modes with

$$\boldsymbol{k} = \mathrm{i}\boldsymbol{q} \pm \sqrt{q^2 + 1/\lambda^2}\, \hat{\boldsymbol{z}}, \tag{3.49}$$

where $\boldsymbol{q} \cdot \hat{\boldsymbol{z}} = 0$ and $\boldsymbol{k} \cdot \boldsymbol{k} \equiv 1/\lambda^2$. These modes are oscillatory in the transverse direction and decay in the $\pm z$-directions. By inserting (3.48) into the RTE (3.10), we find that $A_{\boldsymbol{k}}(\hat{\mathbf{s}})$ satisfies the equation

$$(\hat{\mathbf{s}} \cdot \boldsymbol{k} + \mu_a + \mu_s)\, A_{\boldsymbol{k}}(\hat{\mathbf{s}}) = \mu_s \int p(\hat{\mathbf{s}}, \hat{\mathbf{s}}') A_{\boldsymbol{k}}(\hat{\mathbf{s}}')\, \mathrm{d}\hat{\mathbf{s}}'. \tag{3.50}$$

To solve the eigenproblem defined by (3.50) it will prove useful to expand $A_{\boldsymbol{k}}(\hat{\mathbf{s}})$ into a basis of spherical functions defined in a rotated reference frame whose z-axis coincides with the direction $\hat{\boldsymbol{k}}$. We denote such functions by $Y_{lm}(\hat{\mathbf{s}}; \hat{\boldsymbol{k}})$ and define them via the relation

$$Y_{lm}(\hat{\mathbf{s}}; \hat{\boldsymbol{k}}) = \sum_{m'=-l}^{l} D^l_{mm'}(\varphi, \theta, 0) Y_{lm'}(\hat{\mathbf{s}}), \tag{3.51}$$

where $Y_{lm}(\hat{\mathbf{s}})$ are the spherical harmonics in the laboratory frame, $D^l_{mm'}$ is the Wigner D-function and φ, θ are the polar angles of $\hat{\boldsymbol{k}}$ in the laboratory frame. We thus expand $A_{\boldsymbol{k}}$ as

$$A_{\boldsymbol{k}}(\hat{\mathbf{s}}) = \sum_{l,m} C_{lm} Y_{lm}(\hat{\mathbf{s}}; \hat{\boldsymbol{k}}), \tag{3.52}$$

where the coefficients C_{lm} are to be determined. Note that since the phase function $p(\hat{\mathbf{s}}, \hat{\mathbf{s}}')$ is invariant under simultaneous rotation of $\hat{\mathbf{s}}$ and $\hat{\mathbf{s}}'$, it may be expanded into rotated spherical functions according to

$$p(\hat{\mathbf{s}}, \hat{\mathbf{s}}') = \sum_{l,m} p_l Y_{lm}(\hat{\mathbf{s}}; \hat{\boldsymbol{k}}) Y^*_{lm}(\hat{\mathbf{s}}'; \hat{\boldsymbol{k}}), \tag{3.53}$$

where the expansion coefficients p_l are independent of $\hat{\boldsymbol{k}}$. An alternative approach that may be used to solve the eigenproblem (3.50) is to employ the method of discrete ordinates [26–28].

Substituting (3.52) into (3.50) and making use of the orthogonality properties of the spherical functions, we find that the coefficients C_{lm} satisfy the equation

$$\sum_{l',m'} \mathsf{R}^{lm}_{l'm'} C_{l'm'} = \lambda \sigma_l C_{lm}. \tag{3.54}$$

Here the matrix R is defined by

$$\begin{aligned} \mathsf{R}^{lm}_{l'm'} &= \int \mathrm{d}\hat{\mathbf{s}}\, \hat{\mathbf{s}} \cdot \hat{\boldsymbol{k}} Y_{lm}(\hat{\mathbf{s}}; \hat{\boldsymbol{k}}) Y^*_{l'm'}(\hat{\mathbf{s}}; \hat{\boldsymbol{k}}) \\ &= \delta_{mm'}(b_{lm}\delta_{l',l-1} + b_{l+1,m}\delta_{l',l-1}), \end{aligned} \tag{3.55}$$

where

$$b_{lm} = \sqrt{(l^2 - m^2)/(4l^2 - 1)}, \tag{3.56}$$

and

$$\sigma_l = \mu_a + \mu_s(1 - p_l). \tag{3.57}$$

Equation (3.54) defines a generalized eigenproblem which can be transformed into a standard eigenproblem as follows. Define the diagonal matrix $\mathsf{S}^{lm}_{l'm'} = \delta_{mm'}\delta_{ll'}\sqrt{\sigma_l}$. Note that $\sigma_l > 0$ since $p_l \leqslant 1$ and thus S is well defined. We then pre- and post-multiply R by S^{-1} and find that $\mathsf{W}\psi = \lambda\psi$ where $\mathsf{W} = \mathsf{S}^{-1}\mathsf{R}\mathsf{S}^{-1}$ and $\psi = \mathsf{S}C$. It can be shown that W is symmetric and block tridiagonal with both a discrete and continuous spectrum of eigenvalues λ_μ and a corresponding complete orthonormal set of eigenvectors ψ_μ, indexed by μ [25]. We thus see that the modes (3.48), which are labeled by μ, the transverse wave vector $\boldsymbol{q}$, and the direction of decay, are of the form

$$I^{\pm}_{q\mu}(\boldsymbol{r}, \hat{\mathbf{s}}) = \sum_{l,m}\sum_{m'} \frac{1}{\sqrt{\sigma_l}} \psi^{\mu}_{lm} D^{l}_{mm'}(\varphi, \theta, 0) Y_{lm'}(\hat{\mathbf{s}})\, \mathrm{e}^{\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{\rho} \mp Q_\mu(\boldsymbol{q})z}, \tag{3.58}$$

where

$$Q_\mu(\boldsymbol{q}) = \sqrt{q^2 + 1/\lambda^2_\mu}. \tag{3.59}$$

Green's function for the RTE in the $z \geqslant 0$ half-space may be constructed as a superposition of the above modes:

$$G(\boldsymbol{r}, \hat{\mathbf{s}}; \boldsymbol{r}', \hat{\mathbf{s}}') = \int \frac{\mathrm{d}\boldsymbol{q}}{(2\pi)^2} \sum_{\mu} \mathcal{A}_{q\mu} I^{\pm}_{q\mu}(\boldsymbol{r}, \hat{\mathbf{s}}) I^{\mp}_{-q\mu}(\boldsymbol{r}', -\hat{\mathbf{s}}), \tag{3.60}$$

where the upper sign is chosen if $z > z'$, the lower sign is chosen if $z < z'$ and the coefficients $\mathcal{A}_{q\mu}$ are found from the boundary conditions. Using this result, we see that G can be written as the plane-wave decomposition

$$G(\boldsymbol{r}, \hat{\mathbf{s}}; \boldsymbol{r}', \hat{\mathbf{s}}') = \int \frac{\mathrm{d}\boldsymbol{q}}{(2\pi)^2} \sum_{lm,l'm'} g^{lm}_{l'm'}(z, z'; \boldsymbol{q})\, \mathrm{e}^{\mathrm{i}\boldsymbol{q}\cdot(\boldsymbol{\rho}-\boldsymbol{\rho}')} Y_{lm}(\hat{\mathbf{s}}) Y^*_{l'm'}(\hat{\mathbf{s}}'), \tag{3.61}$$

where

$$\begin{aligned} g^{lm}_{l'm'}(z, z'; \boldsymbol{q}) &= \frac{1}{\sqrt{\sigma_l \sigma'_l}} \sum_{\mu} \sum_{M,M'} \mathcal{A}_{q\mu} \psi^{\mu}_{lm} \psi^{\mu}_{l'm'} \\ &\quad \times D^{l}_{mM}(\varphi, \theta, 0) D^{l'}_{m'M'}(\varphi, \theta, 0)\, \mathrm{e}^{-Q_\mu(\boldsymbol{q})|z-z'|} \qquad (3.62) \\ &\equiv \sum_{\mu} \mathcal{B}^{lm}_{l'm'}(\boldsymbol{q}, \mu)\, \mathrm{e}^{-Q_\mu(\boldsymbol{q})|z-z'|}, \qquad (3.63) \end{aligned}$$

which defines $\mathcal{B}^{lm}_{l'm'}$. It is important to note that the dependence of G on the coordinates $\boldsymbol{r}, \boldsymbol{r}'$ and directions $\hat{\mathbf{s}}, \hat{\mathbf{s}}'$ is *explicit* and that the expansion is computable for any rotationally invariant phase function.

3.5. Diffuse light

The diffusion approximation (DA) to the RTE is widely used in applications. It is valid in the regime where the scattering length $l_s = 1/\mu_s$ is small compared to the distance of propagation. The standard approach to the DA is through the P_N approximation, in which the angular dependence of the specific intensity is expanded in spherical harmonics. The DA is obtained if the expansion is truncated at first order. The DA may also be derived using

asymptotic methods [29]. The advantage of this approach is that it leads to error estimates and treats the problem of boundary conditions for the resulting diffusion equation in a natural way.

We begin from the time-dependent form of the RTE

$$\frac{1}{c}\frac{\partial I}{\partial t} + \hat{\mathbf{s}} \cdot \nabla I + \mu_a I + \mu_s I = \mu_s \int p(\hat{\mathbf{s}}, \hat{\mathbf{s}}') I(\boldsymbol{r}, \hat{\mathbf{s}}')\, \mathrm{d}\hat{\mathbf{s}}'. \tag{3.64}$$

The DA holds when the scattering coefficient is large, the absorption coefficient is small, the point of observation is far from the boundary of the medium and the time scale is sufficiently long. Accordingly, we perform the rescaling

$$\mu_a \to \epsilon \mu_a, \qquad \mu_s \to \frac{1}{\epsilon}\mu_s, \qquad t \to t/\epsilon, \tag{3.65}$$

where $\epsilon \ll 1$. Thus, (3.64) becomes

$$\frac{\epsilon^2}{c}\frac{\partial I}{\partial t} + \epsilon \hat{\mathbf{s}} \cdot \nabla I + \epsilon^2 \mu_a I + \mu_s I = \mu_s \int p(\hat{\mathbf{s}}, \hat{\mathbf{s}}') I(\boldsymbol{r}, \hat{\mathbf{s}}')\, \mathrm{d}\hat{\mathbf{s}}'. \tag{3.66}$$

We then introduce the asymptotic expansion for the specific intensity

$$I(\boldsymbol{r}, \hat{\mathbf{s}}) = I_0(\boldsymbol{r}, \hat{\mathbf{s}}) + \epsilon I_1(\boldsymbol{r}, \hat{\mathbf{s}}) + \epsilon^2 I_2(\boldsymbol{r}, \hat{\mathbf{s}}) + \cdots \tag{3.67}$$

which we substitute into (3.66). Upon collecting terms of $O(1)$, $O(\epsilon)$ and $O(\epsilon^2)$ we have

$$\int p(\hat{\mathbf{s}}, \hat{\mathbf{s}}') I_0(\boldsymbol{r}, \hat{\mathbf{s}}')\, \mathrm{d}\hat{\mathbf{s}}' = I_0(\boldsymbol{r}, \hat{\mathbf{s}}), \tag{3.68}$$

$$\hat{\mathbf{s}} \cdot \nabla I_0 + \mu_s I_1 = \mu_s \int p(\hat{\mathbf{s}}, \hat{\mathbf{s}}') I_1(\boldsymbol{r}, \hat{\mathbf{s}}')\, \mathrm{d}\hat{\mathbf{s}}', \tag{3.69}$$

$$\frac{1}{c}\frac{\partial I_0}{\partial t} + \hat{\mathbf{s}} \cdot \nabla I_1 + \mu_a I_0 + \mu_s I_2 = \mu_s \int p(\hat{\mathbf{s}}, \hat{\mathbf{s}}') I_2(\boldsymbol{r}, \hat{\mathbf{s}}')\, \mathrm{d}\hat{\mathbf{s}}'. \tag{3.70}$$

The normalization condition (3.12) forces I_0 to depend only upon the spatial coordinate $\boldsymbol{r}$. If the phase function $p(\hat{\mathbf{s}}, \hat{\mathbf{s}}')$ depends only upon the quantity $\hat{\mathbf{s}} \cdot \hat{\mathbf{s}}'$, it can be seen that

$$I_1(\boldsymbol{r}, \hat{\mathbf{s}}) = -\frac{1}{1-g}\hat{\mathbf{s}} \cdot \nabla I_0(\boldsymbol{r}), \tag{3.71}$$

where the anisotropy g is given by

$$g = \int \hat{\mathbf{s}} \cdot \hat{\mathbf{s}}' p(\hat{\mathbf{s}} \cdot \hat{\mathbf{s}}')\, \mathrm{d}\hat{\mathbf{s}}', \tag{3.72}$$

with $-1 < g < 1$. Note that $g = 0$ corresponds to isotropic scattering and $g = 1$ to extreme forward scattering. If we insert the above expression for I_1 into (3.70) and integrate over $\hat{\mathbf{s}}$, we obtain the diffusion equation for the energy density Φ:

$$\frac{\partial}{\partial t}\Phi(\boldsymbol{r}, t) = \nabla \cdot [D(\boldsymbol{r})\nabla\Phi(\boldsymbol{r}, t)] - c\mu_a(\boldsymbol{r})u(\boldsymbol{r}, t), \tag{3.73}$$

where $I_0 = c\Phi/(4\pi)$. Here the diffusion coefficient is defined by

$$D = \frac{1}{3}c\ell^*, \qquad \ell^* = \frac{1}{(1-g)\mu_t}, \tag{3.74}$$

where ℓ^* is known as the transport mean free path. The usual expression for ℓ^* obtained from the P_N method

$$\ell^* = \frac{1}{(1-g)\mu_s + \mu_a}, \tag{3.75}$$

is asymptotically equivalent to (3.74) since $\mu_a = \epsilon^2 \mu_s$.

The above derivation of the DA holds in an infinite medium and at long times. In a bounded domain, it is necessary to account for both boundary and initial layers, since the boundary conditions for the diffusion equation and the RTE are not compatible [29]. For the remainder of this review, we will make use of the Robin boundary condition for a bounded domain Ω, which is of the form

$$\mathcal{B}^- \Phi := c\Phi + 2\zeta D \hat{\nu} \cdot \nabla \Phi = J^- \quad \text{on} \quad \partial\Omega, \tag{3.76}$$

where $\hat{\nu}$ is the outward unit normal and the extrapolation length $\ell_{\text{ext}} = 2\zeta D/c$ can be computed from radiative transport theory [15]. We note that $\zeta = 0$ corresponds to an absorbing boundary and $\zeta \to \infty$ to a reflecting boundary. In the case of a boundary between diffusing media of differing refractive index n_0 and n_1 then ζ and ℓ_{ext} take the form

$$\zeta = \frac{1+R}{1-R}, \qquad \ell_{\text{ext}} = \frac{2}{3}\ell^* \frac{1+R}{1-R}, \tag{3.77}$$

where R is the unpolarized Fresnel reflection coefficient. We also define the outgoing current

$$\mathcal{B}^+ \Phi := J^+ = -D\hat{\nu} \cdot \nabla \Phi = \frac{1}{2\zeta}(c\Phi - J^-). \tag{3.78}$$

If the phase function is not rotationally invariant, then the diffusion coefficient in (3.73) becomes a second rank tensor, whose elements are given by angular moments of $p(\hat{\mathbf{s}}, \hat{\mathbf{s}}')$. The diffusion process is then anisotropic. We return to anisotropic diffusion briefly in section 9.1. We also note that if the speed of light is not constant, then the above derivation of the DA can be modified beginning from (3.27) [19]. The resulting diffusion equation depends explicitly upon the index of refraction n and is of the form

$$\frac{\partial}{\partial t}\Phi(\boldsymbol{r}, t) = \nabla \cdot \left[n^2(\boldsymbol{r}) D(\boldsymbol{r}) \nabla \left(\frac{\Phi(\boldsymbol{r}, t)}{n^2(\boldsymbol{r})} \right) \right] - c\mu_a(\boldsymbol{r})\Phi(\boldsymbol{r}, t). \tag{3.79}$$

We now consider the scattering theory of time-harmonic diffuse waves. Assuming an $e^{i\omega t}$ time dependence with modulation frequency ω, the energy density obeys the equation

$$-\nabla \cdot [D(\boldsymbol{r})\nabla\Phi(\boldsymbol{r})] + (c\mu_a(\boldsymbol{r}) + i\omega)\Phi(\boldsymbol{r}) = 0, \tag{3.80}$$

In addition to (3.80), the energy density must satisfy the boundary condition (3.76). The solution to (3.80) obeys the Lippmann–Schwinger equation

$$\Phi = \Phi_0 - \mathcal{G}\mathcal{V}\Phi_0, \qquad \Phi_0 = \mathcal{G}J^-, \tag{3.81}$$

where $\mathcal{G}$ is Green's function for a homogeneous medium with absorption $\bar{\mu}_a$ and diffusion coefficient D_0. We have also made use of the specific form of the generalized potential operator introduced in (2.18), which for the diffusion equation is given by

$$\mathcal{V} = c\delta\mu_a - \nabla \cdot (\delta D \nabla), \tag{3.82}$$

where $\delta\mu_a = \mu_a - \bar{\mu}_a$ and $\delta D = D - \bar{D}$, with $\bar{\mu}_a$ and $\bar{D}$ constant. The unperturbed Green's function $G(\boldsymbol{r}, \boldsymbol{r}')$ satisfies

$$(\nabla^2 - k^2)G(\boldsymbol{r}, \boldsymbol{r}') = -\frac{1}{D_0}\delta(\boldsymbol{r} - \boldsymbol{r}'), \tag{3.83}$$

where the diffuse wavenumber k is given by

$$k^2 = \frac{c\bar{\mu}_a + \mathrm{i}\omega}{\bar{D}}. \tag{3.84}$$

We note here that the fundamental solution to the diffusion equation is given by

$$G(\boldsymbol{r}, \boldsymbol{r}') = \frac{1}{4\pi D} \frac{\mathrm{e}^{-k|\boldsymbol{r}-\boldsymbol{r}'|}}{|\boldsymbol{r}-\boldsymbol{r}'|}. \tag{3.85}$$

By iterating (3.81) beginning with $\Phi = \Phi_0$, we obtain (compare to (2.31))

$$\Phi = \Phi_0 - \mathcal{G}\mathcal{V}\Phi_0 + \mathcal{G}\mathcal{V}\mathcal{G}\mathcal{V}\Phi_0 + \cdots, \tag{3.86}$$

which is the analogue of the Born series for diffuse waves.

4. Optical tomography modalities

We may now define some of the different problems of optical tomography

4.1. Optical tomography based on the diffusion equation

In diffuse optical tomography the space X consists of pairs of functions $x \equiv (\mu_\mathrm{a}, D)$ and the governing PDE is the diffusion equation. We consider the domain as a simply connected set in $\mathbb{R}^n$ with

$$\Xi \equiv \Omega, \qquad \partial\Xi \equiv \partial\Omega. \tag{4.1}$$

The source functions J^- and their resultant outgoing currents J^+ belong to the space $H^{-1/2}(\partial\Omega)$. For the time-dependent case the PDE is given by (3.73) and in the frequency domain case by (3.80). In both cases Φ must satisfy the Robin boundary condition $\mathcal{B}^-$ given by (3.76) and Neumann boundary condition $\mathcal{B}^+$ given by (3.78); therefore the mapping

$$\Lambda : H^{-1/2}(\partial\Omega) \to H^{-1/2}(\partial\Omega) \tag{4.2}$$

is the Robin-to-Neumann map.

4.1.1. Diffuse optical tomography (DOT). The nonlinear inverse problem in DOT is to recover functions $\mu_\mathrm{a} \in \mathrm{X}_{\mu_\mathrm{a}}$, $D \in \mathrm{X}_D$ from measurements $y \in \mathrm{Y}$. The data may be collected either in the time domain or in the frequency domain, either at one or several modulation frequencies ω [3]. If the former is used, it is typical to Fourier transform the data and to develop the analysis in the frequency domain. At the zero frequency $\omega = 0$ ('DC' measurements), the data simply represent a total photon count without any phase information, and the recovery of both μ_a and D suffers from non-uniqueness [30] (but see [31]). In this case a simpler problem is commonly defined in which one parameter (usually scattering) is assumed known and the forward mapping is restricted to recovery of the second function.

For the DOT problem the Fréchet derivative (2.12) has a kernel based on the density functions for absorption and diffusion

$$\begin{pmatrix} \rho_{\mu_\mathrm{a}}(\boldsymbol{r}; \omega) \\ \rho_D(\boldsymbol{r}; \omega) \end{pmatrix} = \begin{pmatrix} \Phi^*(\boldsymbol{r}; \omega)\Phi(\boldsymbol{r}; \omega) \\ \nabla\Phi^*(\boldsymbol{r}; \omega) \cdot \nabla\Phi(\boldsymbol{r}; \omega) \end{pmatrix}, \tag{4.3}$$

$$\begin{pmatrix} \rho_{\mu_\mathrm{a}}(\boldsymbol{r}, t) \\ \rho_D(\boldsymbol{r}, t) \end{pmatrix} = \begin{pmatrix} \int_0^t \Phi^*(\boldsymbol{r}, -t')\Phi(\boldsymbol{r}, t-t')\,\mathrm{d}t' \\ \int_0^t \nabla\Phi^*(\boldsymbol{r}, -t') \cdot \nabla\Phi(\boldsymbol{r}, t-t')\,\mathrm{d}t' \end{pmatrix}, \tag{4.4}$$

where we note that the adjoint problem leads to a backward-time equation.

Although DOT is a nonlinear inverse problem, a linearized version is often considered, which we refer to as *difference diffuse optical tomography* (DDOT). Here we take differences in measurements $\delta y = y_2 - y_1$ and formulate a linear mapping as in (2.24). Clearly the relevant linear mapping is given by the linear Fréchet derivative operator, evaluated at $x_1 = (\mu_{a_1}, D_1)$. As in nonlinear DOT, the simplified case of absorption only imaging is frequently considered. The success of this approach depends on how sensitive the reconstruction is to the correct choice of linearization point x_1. This typically involves careful calibration and consideration of the modelling errors. We return to this point in section 7.4.

4.1.2. Fluorescence diffuse optical tomography (FDOT). In fluorescence optical tomography sources are introduced at an *excitation* wavelength λ^e giving rise to an excitation field Φ^e. Fluorescence is regarded as a function $h(\boldsymbol{r})$ which governs the absorption of radiation at wavelength λ^e and (partial) re-emission as a source at the longer wavelength (lower energy) λ^f. Measurements are taken at λ^f and λ^e. Using the frequency domain formulation, we consider two coupled PDEs

$$\mathcal{L}(x^e)\Phi^e \equiv -\nabla \cdot D^e(\boldsymbol{r})\nabla\Phi^e(\boldsymbol{r};\omega) + \left(c\mu_a^e(\boldsymbol{r}) + \mathrm{i}\omega\right)\Phi^e(\boldsymbol{r};\omega) = 0 \tag{4.5}$$

$$\mathcal{L}(x^f)\Phi^{e\to f} \equiv -\nabla \cdot D^f(\boldsymbol{r})\nabla\Phi^{e\to f}(\boldsymbol{r};\omega) + \left(c\mu_a^f(\boldsymbol{r}) + \mathrm{i}\omega\right)\Phi^{e\to f}(\boldsymbol{r};\omega) = h(\boldsymbol{r};\omega)\Phi^e(\boldsymbol{r};\omega) \tag{4.6}$$

$$\mathcal{B}^-\Phi^e \equiv c\Phi^e(\boldsymbol{r}_m;\omega) + 2\zeta D^e(\boldsymbol{r}_m)\hat{\nu}\cdot\nabla\Phi^e(\boldsymbol{r}_m;\omega) = J^{e-}(\boldsymbol{r}_s;\omega) \tag{4.7}$$

$$\mathcal{B}^-\Phi^{e\to f} \equiv c\Phi^{e\to f}(\boldsymbol{r}_m;\omega) + 2\zeta D^f(\boldsymbol{r}_m)\hat{\nu}\cdot\nabla\Phi^{e\to f}(\boldsymbol{r}_m;\omega) = 0 \tag{4.8}$$

$$\mathcal{B}^+\Phi^e \equiv J^{e+}(\boldsymbol{r}_m;\omega) = -D^e(\boldsymbol{r}_m)\hat{\nu}\cdot\nabla\Phi^e(\boldsymbol{r}_m;\omega) \tag{4.9}$$

$$\mathcal{B}^+\Phi^{e\to f} \equiv J^{e\to f+}(\boldsymbol{r}_m;\omega) = -D^f(\boldsymbol{r}_m)\hat{\nu}\cdot\nabla\Phi^{e\to f}(\boldsymbol{r}_m;\omega) \tag{4.10}$$

The quantity h is a product of the concentration of fluorescent material and its conversion efficiency. Since re-emission is a Poisson process it is characterized by a lifetime τ. In the frequency domain this leads to a complex valued parameter

$$h(\boldsymbol{r},\omega) = h_0(\boldsymbol{r})\frac{1}{1+\mathrm{i}\omega\tau(\boldsymbol{r})}. \tag{4.11}$$

The simplest problem considered is to recover $h \in \mathrm{X}_h$ from measurements $y \in \mathrm{Y}^{e\to f}$ and the linear operator

$$\mathcal{F}_h^{e\to f} : \mathrm{X}_h \to \mathrm{Y}^{e\to f}. \tag{4.12}$$

Consider the frequency-domain version. The kernel of the Fréchet derivative (2.12) in this case is the density function with respect to h given by

$$\rho_h(\boldsymbol{r};\omega) = \Phi^{e\to f*}(\boldsymbol{r};\omega)\Phi^e(\boldsymbol{r};\omega), \tag{4.13}$$

giving a complex-valued reconstruction from which h_0, τ can be recovered using (4.11). Clearly, for DC measurements, only the fluorescence h_0 can be recovered, not the lifetime; *fluorescence lifetime imaging tomography* (FLIM tomography) is a term used to emphasize that both h_0 and τ are being recovered. Time-domain measurements may of course be used to solve this problem. If the data are Fourier transformed into the frequency domain the techniques for image reconstruction are unchanged. It may also be expressed directly in the time domain. In this case the kernel is given by a double convolution

$$\rho_h(\boldsymbol{r};t) = \int_0^t\int_0^{t'} \Phi^{f*}(\boldsymbol{r};-t'')\Phi^e(\boldsymbol{r};t-t')h(\boldsymbol{r},t''-t')\,\mathrm{d}t'\,\mathrm{d}t'' \tag{4.14}$$

where

$$h(\boldsymbol{r}, t'' - t') = h_0(\boldsymbol{r})\, \mathrm{e}^{-\frac{|t''-t'|}{\tau(\boldsymbol{r})}} \tag{4.15}$$

gives explicitly the exponential decay of the Poisson process. In practice the half-life does not have a discrete value but a distribution, and the linear integral equation with (4.14) as its kernel has to also be convolved with the finite time spread function of the measurement system [32].

4.1.3. Diffuse optical tomography with fluorescence (DOT-FDOT). The discussion of the linear FDOT problem in section 4.1.2 is predictated on knowing the parameters x^e and x^f used to construct the kernel in (4.13). These could themselves be recovered using uncoupled DOT reconstructions at the excitation and fluorescence wavelengths. Therefore, it is natural to consider the joint problem (using the notation of (4.5)–(4.10))

$$\mathcal{L}(x^e)\Phi^e = 0 \qquad \mathcal{B}^-\Phi^e = J^{e-} \qquad \mathcal{B}^+\Phi^e = J^{e+}, \tag{4.16}$$

$$\mathcal{L}(x^f)\Phi^f = 0 \qquad \mathcal{B}^-\Phi^f = J^{f-} \qquad \mathcal{B}^+\Phi^f = J^{f+}, \tag{4.17}$$

$$\mathcal{L}(x^f)\Phi^{e\to f} = h\Phi^e \qquad \mathcal{B}^-\Phi^{e\to f} = 0 \qquad \mathcal{B}^+\Phi^{e\to f} = J^{e\to f+}, \tag{4.18}$$

with the joint forward operator

$$\begin{pmatrix} \mathcal{F}_x^e \\ \mathcal{F}_x^f \\ \mathcal{F}_h^{e\to f} \end{pmatrix} : \begin{pmatrix} \mathrm{X}_{\mu_\mathrm{a}}^e, \mathrm{X}_D^e \\ \mathrm{X}_{\mu_\mathrm{a}}^f, \mathrm{X}_D^f \\ \mathrm{X}_h \end{pmatrix} \to \begin{pmatrix} \mathrm{Y}^e \\ \mathrm{Y}^f \\ \mathrm{Y}^{e\to f} \end{pmatrix}. \tag{4.19}$$

The solution of the joint problem is computationally more demanding than the three separate problems. This problem has hardly been addressed in the frequency domain in [33, 34, 68] and in the time domain in [35–38].

4.1.4. Multispectral diffuse optical tomography. Whereas each of the above problems could be considered at a range of spectral samples and the spectral variation of the recovered solutions determined, the idea in multispectral DOT is to reformulate the problem into the recovery of a set of images of known chromophores with well characterized spectral dependence:

$$\mu_\mathrm{a}(\lambda_j) = \sum_i \epsilon_i(\lambda_j) c_i \to \mu_\mathrm{a}(\lambda) = \epsilon \mathbf{c}, \tag{4.20}$$

where ϵ is a known matrix. Similarly a spectral dependence of scattering can be written as

$$\mu_\mathrm{s}'(\lambda) = a\lambda^{-b}. \tag{4.21}$$

We note that the above model for the wavelength dependence of μ_s' corresponds to Rayleigh scattering when $b = 4$. In general, subdominant power-law corrections may be necessary to accurately represent the scattering behaviour of tissue.

Using the spectral models (4.20)–(4.21) leads to a formulation

$$\mathcal{F}_{\{c,a,b\}}^{\lambda} : \mathrm{X}_{\{c,a,b\}} \to \mathrm{Y}^{\lambda}. \tag{4.22}$$

The spectrally decoupled linearization is represented as

$$\begin{pmatrix} \delta \boldsymbol{y}(\lambda_1) \\ \delta \boldsymbol{y}(\lambda_2) \\ \vdots \\ \delta \boldsymbol{y}(\lambda_N) \end{pmatrix} = \begin{pmatrix} \mathsf{A}_{\mu_\mathrm{a}}(\lambda_1) & \dots & 0 & \mathsf{A}_{\mu_\mathrm{s}'}(\lambda_1) & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \mathsf{A}_{\mu_\mathrm{a}}(\lambda_N) & 0 & \dots & \mathsf{A}_{\mu_\mathrm{s}'}(\lambda_N) \end{pmatrix} \begin{pmatrix} \delta\mu_\mathrm{a}(\lambda_1) \\ \vdots \\ \delta\mu_\mathrm{a}(\lambda_N) \\ \delta\mu_\mathrm{s}'(\lambda_1) \\ \vdots \\ \delta\mu_\mathrm{s}'(\lambda_N) \end{pmatrix}. \tag{4.23}$$

The spectral component at each pixel is independent so we introduce the diagonal matrices

$$\mathsf{C}_k(\lambda_j) := c_k(\lambda_j)\mathsf{I} \qquad \mathsf{B}_a(\lambda_j) := \lambda_j^{-b}\mathsf{I} \qquad \mathsf{B}_b(\lambda_n) := a\lambda_j^{-b}\ln\lambda_j\mathsf{I}, \tag{4.24}$$

and we can now represent (4.23) as

$$\begin{pmatrix} \delta\boldsymbol{y}(\lambda_1) \\ \delta\boldsymbol{y}(\lambda_2) \\ \vdots \\ \delta\boldsymbol{y}(\lambda_N) \end{pmatrix} = \begin{pmatrix} \mathsf{A}_{\mu_\mathrm{a}}(\lambda_1) & \dots & 0 & \mathsf{A}_{\mu_\mathrm{s}'}(\lambda_1) & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & \mathsf{A}_{\mu_\mathrm{a}}(\lambda_N) & 0 & \dots & \mathsf{A}_{\mu_\mathrm{s}'}(\lambda_N) \end{pmatrix}$$
$$\times \begin{pmatrix} \mathsf{C}_1(\lambda_1) & \dots & \mathsf{C}_K(\lambda_1) & 0 & 0 & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathsf{C}_1(\lambda_n) & \dots & \mathsf{C}_K(\lambda_n) & 0 & 0 & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mathsf{C}_1(\lambda_N) & \dots & \mathsf{C}_K(\lambda_N) & 0 & 0 & \\ 0 & \dots & 0 & \mathsf{B}_a(\lambda_1) & \mathsf{B}_b(\lambda_1) & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & \mathsf{B}_a(\lambda_n) & \mathsf{B}_b(\lambda_n) & \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & \mathsf{B}_a(\lambda_N) & \mathsf{B}_b(\lambda_1) & \end{pmatrix} \begin{pmatrix} \delta\boldsymbol{c}_1 \\ \vdots \\ \delta\boldsymbol{c}_k \\ \vdots \\ \delta\boldsymbol{c}_K \\ \delta\boldsymbol{a} \\ \delta\boldsymbol{b} \end{pmatrix}, \tag{4.25}$$

see [39, 40].

Similarly to MSDOT in *multispectral fluorescence DOT* (MSFDOT) the fluorescence may be considered a linear combination of fluorophores emitting radiation in a known spectral pattern

$$h(\lambda_j) = \sum_i \epsilon_i(\lambda_j)p_i \to h(\lambda) = \epsilon^f\,\mathbf{p}. \tag{4.26}$$

In this case a linear forward operator may be constructed

$$\mathcal{F}_{\mathbf{p}}^{\lambda^e\to\lambda^f} : \mathrm{X}_{\mathbf{p}} \to \mathrm{Y}^{\lambda^e\to\lambda^f}, \tag{4.27}$$

see [41]. Finally, a very general problem could be considered in the recovery of all chromophores and flurophores from the operator

$$\begin{pmatrix} \mathcal{F}_{\{\boldsymbol{c},a,b\}}^{\lambda^e,\lambda^f} \\ \mathcal{F}_{\mathbf{p}}^{\lambda^e\to\lambda^f} \end{pmatrix} : \begin{pmatrix} \mathrm{X}_{\{\boldsymbol{c},a,b\}} \\ \mathrm{X}_{\mathbf{p}} \end{pmatrix} \to \begin{pmatrix} \mathrm{Y}^{\lambda^e,\lambda^f} \\ \mathrm{Y}^{\lambda^e\to\lambda^f} \end{pmatrix}. \tag{4.28}$$

4.1.5. Bioluminescence (BDOT). Bioluminescence optical tomography detects the light emitted from within tissues through the action of luciferase on its substrate, luciferin. The luciferase emission spectrum is quite wide [42] (500–800 nm) and overlaps with so-called near infrared window of biological tissues (650–850 nm), where the light scattering dominates over the light absorption. Red and near infrared components of the emission spectrum penetrate biological tissue appreciably well allowing to image bioluminescence objects imbedded deeply inside tissue.

In terms of the general framework established in section 2 the bioluminescence tomography forward problem is posed as

$$\mathcal{L}(x)\Phi = q \tag{4.29}$$

$$\mathcal{B}^{-}\Phi = 0, \tag{4.30}$$

$$J^{+} = \mathcal{B}^{+}\Phi. \tag{4.31}$$

There are no incoming radiation sources and spontaneously created light is subject to the same boundary conditions as in all other problems considered in this review. In the inverse problem, the unknown function q is to be reconstructed from measured values of the function J^{+} on the surface. This is a linear inverse problem which can be formulated in terms of a Fredholm integral equation of the first kind

$$J^{+}(\xi_{\mathrm{m}}) = \int_{\Xi} G(\xi_{\mathrm{m}}, \xi) q(\xi)\, \mathrm{d}\xi, \tag{4.32}$$

In [43] the bioluminescent source distribution was recovered from a monochromatic data set and theoretical aspects of uniqueness of solutions were discussed in [44, 45]. The idea of using multiple wavelengths in bioluminescence imaging was introduced by Chaudhari *et al* [46] in which this approach is applied for 3D localization of deep sources within a mouse phantom *in vivo*. Combination of bioluminescence and positron emission tomography (PET) is discussed in [47].

4.2. Optical tomography based on the radiative transfer equation

Each of the problems in section 4.1 could be considered based on the RTE instead of the diffusion equation. The parameters considered in the inverse problem are the same but the measurements are different. There are two possibilities: angularly resolved measurements or angularly independent measurements. In the latter case the size of the measured data set is the same as in the diffusion problems, but the construction of the linearization is different. In the case of angularly dependent data the data set is much bigger. Issues regarding uniqueness depend critically on which measurements are made [23].

In the RTE-based problems, the generalized domain becomes $\Xi = \Omega \times S^{n-1}$ with boundary $\partial\Xi = B_{-} \cup B_{+}$ where $B_{\pm} = \partial\Omega \times S^{n-1}_{\pm}(\nu)$. The governing PDE is given by (3.10) with parameters μ_{a}, μ_{s} and the boundary condition $\mathcal{B}^{-}$ specifies the specific intensity on B_{-} as given by (3.11). The transfer function (2.5) is known as the *albedo operator*

$$\Lambda : L_1(B_{-}) \to L_1(B_{+}), \tag{4.33}$$

which maps directional incoming radiation on $\partial\Omega$ to directional outgoing radiation

$$J^{+}_{i,i'}(\boldsymbol{r}_{\mathrm{m}}, \hat{\boldsymbol{s}})|_{\hat{\boldsymbol{s}}\cdot\hat{\nu}>0} = \Lambda J^{-}_{i,i'}(\boldsymbol{r}_{\mathrm{s}}, \hat{\boldsymbol{s}})|_{\hat{\boldsymbol{s}}\cdot\hat{\nu}<0}, \tag{4.34}$$

where the set of source functions $\mathbb{S}$ is indexed by position index i and direction index i'.

As in the diffusion case, the specific problems may be time dependent, time independent, or frequency domain. The role of the aperture function is crucial. In angular resolved measurements

$$\begin{aligned} y_{j,j',i,i'} &= \left\langle w_{j,j'}, J^{+}_{i,i'} \right\rangle_{\partial\Xi} \\ &= \int_{S^2} w_{j'}(\hat{\boldsymbol{s}}) \int_{\partial\Omega} w_j(\boldsymbol{r}_{\mathrm{m}}) J^{+}_{i,i'}(\boldsymbol{r}_{\mathrm{m}}, \hat{\boldsymbol{s}})\, \mathrm{d}\boldsymbol{r}_{\mathrm{m}}\, \mathrm{d}\hat{\boldsymbol{s}}. \end{aligned} \tag{4.35}$$

The sensitivity functions are given by

$$\rho_{\mu_{\mathrm{a}}}(\boldsymbol{r}) = \int_{S^2} I^{*}(\boldsymbol{r}, \hat{\boldsymbol{s}}) I(\boldsymbol{r}, \hat{\boldsymbol{s}})\, \mathrm{d}\hat{\boldsymbol{s}} \tag{4.36}$$

$$\rho_{\mu_{\mathrm{s}}}(\boldsymbol{r}) = \int_{S^2} I^{*}(\boldsymbol{r}, \hat{\boldsymbol{s}}) I(\boldsymbol{r}, \hat{\boldsymbol{s}})\, \mathrm{d}\hat{\boldsymbol{s}} - \int_{S^2}\int_{S^2} p(\hat{\boldsymbol{s}} \cdot \hat{\boldsymbol{s}}\,') I^{*}(\boldsymbol{r}, \hat{\boldsymbol{s}})\, \mathrm{d}\hat{\boldsymbol{s}}\, I(\boldsymbol{r}, \hat{\boldsymbol{s}})\, \mathrm{d}\hat{\boldsymbol{s}} \tag{4.37}$$

where $I^*(\boldsymbol{r}, \hat{\boldsymbol{s}})$ is the solution to the adjoint transport equation with source boundary condition $w_{j'}(\hat{\boldsymbol{s}})w_j(\boldsymbol{r}_\mathrm{m})$ on B_+.

Note that the equivalent right-hand side for the fluorescence problem in (4.6) is assumed isotropic

$$q^f(\boldsymbol{r}) = h(\boldsymbol{r}) \int_{S^2} I^e(\boldsymbol{r}, \hat{\boldsymbol{s}})\, \mathrm{d}\hat{\boldsymbol{s}}. \tag{4.38}$$

For detailed discussion of the uniqueness and stability of the different problems in inverse transport arising from angularly resolved versus angularly averaged and time-dependent versus time-independent measurements we refer to the recent review article [23].

5. Forward modelling methods

The inversion methods described in the sequel are dependent on the accuracy and efficiency with which Green's functions can be computed. The analytic form of such functions for the DA is known for several problems and geometries. The format for the RTE is generally limited to one-dimensional and other special cases (see section 3.4).

5.1. Volume discretization methods

In the finite element method (FEM) the volume Ω is discretized to a mesh

$$\Omega \to \{\mathbb{T}_\Omega, \mathbb{N}_\Omega, \mathbb{U}_\Omega\} \tag{5.1}$$

where $\mathbb{T}_\Omega$ is the set of P elements $\triangle_e$; $e = 1, \ldots, \mathbb{N}_\Omega$ is the set of N vertices $\boldsymbol{N}_p$; $p = 1, \ldots, N$, and $\mathbb{U}_\Omega$ is the set of N locally supported basis functions $u_p(\boldsymbol{r})$; $p = 1, \ldots, N$. For the diffusion equation, writing

$$\Phi(\boldsymbol{r}) \simeq \Phi^h(\boldsymbol{r}) = \sum_p \Phi_p u_p(\boldsymbol{r}), \tag{5.2}$$

results in a discrete system

$$\mathsf{K}(x)\Phi = \boldsymbol{q}, \tag{5.3}$$

where $\mathsf{K}(x)$ has matrix elements

$$K_{lm} = \int_\Omega \left(D(\boldsymbol{r})\nabla u_l(\boldsymbol{r}) \cdot \nabla u_l(\boldsymbol{r}) + (c\mu_\mathrm{a}(\boldsymbol{r}) + \mathrm{i}\omega)\, u_l(\boldsymbol{r})u_m(\boldsymbol{r})\right) \mathrm{d}\boldsymbol{r} + \frac{1}{2\zeta} \int_{\partial\Omega} u_l(\boldsymbol{r})u_m(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r}, \tag{5.4}$$

and $\boldsymbol{q}$ represents the discretization of the boundary conditions. Making use of the basis expansion (2.22) allows the representation of $\mathsf{K}(x)$ as

$$\mathsf{K} = \mathsf{S} + \mathrm{i}\omega\mathsf{B} + \sum_k \left(D_k \mathsf{K}_k^D + c\mu_k \mathsf{K}_k^\mu\right), \tag{5.5}$$

where B is the mass matrix, S is the matrix of surface integrals with elements given by the last term on the right in (5.5) and $\mathsf{K}_k^D \equiv \frac{\partial \mathsf{K}}{\partial D_k}$, $\mathsf{K}_k^\mu \equiv \frac{\partial \mathsf{K}}{\partial c\mu_k}$ are given by

$$\mathsf{K}_{k,lm}^D = \int_\Omega b_k(\boldsymbol{r})\nabla u_l(\boldsymbol{r}) \cdot \nabla u_m(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r} \tag{5.6}$$

$$\mathsf{K}_{k,lm}^\mu = \int_\Omega b_k(\boldsymbol{r})u_l(\boldsymbol{r})u_m(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r}. \tag{5.7}$$

In RTE FEM, a basis is also defined for the angular directions

$$S^{n-1} \to \mathbb{V}_{S^{n-1}} = \{\mathbb{T}_{S^{n-1}}, \mathbb{N}_{S^{n-1}}, \mathbb{U}_{S^{n-1}}\}. \tag{5.8}$$

Different schemes result by choosing different basis on the unit circle S^1 or sphere S^2

(i) The discrete ordinate method chooses a discrete set: $\mathbb{U}_{S^n} = \{\delta(\hat{\boldsymbol{s}} - \hat{\boldsymbol{s}}_p\}$ which are chosen to give exact quadrature points for a spherical harmonic expansion of the angular variable. The radiance is thus represented explicitly in a set of ray directions [48].

(ii) The P_N method chooses the spherical harmonics (rotated into real functions) directly: $\mathbb{U}_{S^n} = \{\tilde{Y}_p(\hat{\boldsymbol{s}})\}$. This allows explicitly the representation as diffusion in the lowest order with higher terms representing higher order effects [49].

(iii) The local basis function method chooses a set of locally supported basis functions on $\mathbb{T}_{S^{n-1}}$. This allows essentially the same machinery of sparse matrix manipulation as for the spatial variables [50–52].

(iv) The wavelet basis chooses a hierarchical set of wavelets on the sphere to allow for variable and adaptive angular discretization [53].

The RTE has been utilized as the forward model for optical tomography reconstruction in a few studies. In most of these papers, the forward solution of the RTE has been based either on the finite difference method (FDM) or on the finite volume method (FVM) and the discrete ordinate formulation of the RTE, see [54–58]. In the FEM solution of the RTE in low-scattering or non-scattering regions, the ray-effect may become visible [59, 60]. In [61], the FEM model of [51, 52] was augmented with the streamline diffusion modification to stabilize the forward solution in this case. The streamline diffusion modification has been found to stabilize numerical solutions of the RTE in situations in which standard techniques produce oscillating results [50, 61, 62]. Due to the heavy computational and memory requirements of meshes, adaptive techniques have been used within a FVM approach [63, 64], as well as the FEM scheme [65–69]. In order to overcome the difficulty of 3D meshing, meshless methods can be used [70].

5.2. Boundary discretization methods

Rather than meshing a volume we may consider the domain as the union of a number of subdomains

$$\Omega = \cup_\ell \Omega_\ell, \qquad \ell = 1 \ldots L, \tag{5.9}$$

together with constant within subdomain optical parameters $\{x_\ell\}$ and a set of interfaces between domains

$$\Sigma_j = \partial\Omega_{\ell,\ell'}, \qquad j = 1 \ldots J. \tag{5.10}$$

We consider the diffusion approximation and introduce the following notation

$$U_i := \Phi_{i-1}|_{\Sigma_i} = \Phi_i|_{\Sigma_i}, \tag{5.11}$$

$$J_i := D_{i-1} \left.\frac{\partial \Phi_{i-1}}{\partial \nu_{i-1}}\right|_{\Sigma_i} = -D_i \left.\frac{\partial \Phi_i}{\partial \nu_i}\right|_{\Sigma_i} \equiv D_{i-1} \left.\frac{\partial \Phi_{i-1}}{\partial \nu}\right|_{\Sigma_i} = D_i \left.\frac{\partial \Phi_i}{\partial \nu}\right|_{\Sigma_i}. \tag{5.12}$$

For simplicity we develop the discussion using only two regions. For a more general implementation we refer to [71]. For homogeneous region Ω_1, Ω_2, Green's second theorem provides the following

$$\Phi_1(\boldsymbol{r}) + \int_{\partial\Omega} \left(\frac{\partial G_1(\boldsymbol{r}, \boldsymbol{r}'_{\mathrm{m}})}{\partial \nu} + \frac{G_1(\boldsymbol{r}, \boldsymbol{r}'_{\mathrm{m}})}{2\zeta D_1} \right) U_1(\boldsymbol{r}'_{\mathrm{m}})\, \mathrm{d}\boldsymbol{r}'_{\mathrm{m}} - \int_{\Sigma} \left(\frac{\partial G_1(\boldsymbol{r}, \boldsymbol{r}'_{\mathrm{m}})}{\partial \nu} U_2(\boldsymbol{r}'_{\mathrm{m}}) - \frac{G_1(\boldsymbol{r}, \boldsymbol{r}'_{\mathrm{m}})}{D_1} J_2(\boldsymbol{r}'_{\mathrm{m}}) \right) \mathrm{d}\boldsymbol{r}'_{\mathrm{m}} = \mathrm{Q}_1(\boldsymbol{r}) \tag{5.13}$$

$$\Phi_2(\boldsymbol{r}) + \int_\Sigma \left(\frac{\partial G_2(\boldsymbol{r}, \boldsymbol{r}'_{\rm m})}{\partial \nu} U_2(\boldsymbol{r}'_{\rm m}) - \frac{G_2(\boldsymbol{r}, \boldsymbol{r}'_{\rm m})}{D_2} J_2(\boldsymbol{r}'_{\rm m}) \right) {\rm d}\boldsymbol{r}'_{\rm m} = {\rm Q}_2(\boldsymbol{r}), \tag{5.14}$$

where

$${\rm Q}_\ell(\boldsymbol{r}) = \int_{\Omega_\ell} G_\ell(\boldsymbol{r}, \boldsymbol{r}') q_\ell(\boldsymbol{r}') \, {\rm d}\boldsymbol{r}', \tag{5.15}$$

and the fundamental solutions G_i are the three-dimensional Green's functions of the diffusion equation in an infinite medium as in (3.85). Taking the limit as $\boldsymbol{r} \in \Omega_1 \to \partial\Omega$, $\boldsymbol{r} \in \Omega_1 \to \Sigma$, and $\boldsymbol{r} \in \Omega_2 \to \Sigma$ results in three coupled boundary integral equations (BIEs) in the three unknown functions

$$\{{\rm f}\} = \{U_1, U_2, J_2\}. \tag{5.16}$$

In contrast to (5.1) only the interfaces Σ_i are discretized to meshes

$$\mathbb{V}_i = \{\mathbb{T}_i, \mathbb{N}_i, \mathbb{U}_i\}, \tag{5.17}$$

where $\mathbb{T}_i$ is the set of P_i surface elements $\triangle_{i,e}$; $e = 1, \ldots, P_i$, $\mathbb{N}_i$ is the set of N_i vertices $\boldsymbol{N}_{i,p}$; $p = 1, \ldots, N_i$, and $\mathbb{U}_i$ is the set of N_i locally supported basis functions $u_{i,p}(\boldsymbol{r})$; $p = 1, \ldots, N_i$. Then the functions (5.11)–(5.12) are represented as

$$U_i(\boldsymbol{r}) \approx \sum_{p=1}^{N_i} U_{i,p} u_{i,p}(\boldsymbol{r}), \qquad J_i(\boldsymbol{r}) \approx \sum_{p=1}^{N_i} J_{i,p} u_{i,p}(\boldsymbol{r}). \tag{5.18}$$

Using the representation (5.18) in the BIE system (5.13)–(5.14) followed by sampling at the nodal points, we obtain a discrete system for the collocation Boundary Element Method [72–76] as a linear matrix equation

$$\begin{pmatrix} \varrho_1^+ {\rm I} + {\rm A}_{11}^1 + \frac{1}{2\zeta} {\rm B}_{11}^1 & -{\rm A}_{12}^1 & {\rm B}_{12}^1 \\ {\rm A}_{11}^2 + \frac{1}{2\zeta} {\rm B}_{11}^2 & \varrho_1^- {\rm I} - {\rm A}_{12}^2 & {\rm B}_{12}^2 \\ 0 & \varrho_2^+ {\rm I} + {\rm A}_{22}^2 & -{\rm B}_{22}^2 \end{pmatrix} \begin{pmatrix} \boldsymbol{U}_1 \\ \boldsymbol{U}_2 \\ \boldsymbol{J}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{Q}_1|_{\partial\Omega} \\ \boldsymbol{Q}_1|_\Sigma \\ 0 \end{pmatrix} \tag{5.19}$$

where we assumed that the source term is only in Ω_1, and where $\boldsymbol{U}_i$, $\boldsymbol{J}_i$ are the vectors of coefficients $U_{i,p}$, $J_{i,p}$ in (5.18), and the matrix elements are given by

$$A_{11}^j(p, p') = \int_{\partial\Omega} \frac{\partial G_1(\boldsymbol{N}_{j,p}, \boldsymbol{r}'_{\rm m})}{\partial \nu} u_{1,p'}(\boldsymbol{r}'_{\rm m}) \, {\rm d}\boldsymbol{r}'_{\rm m} \tag{5.20}$$

$$A_{12}^j(p, p') = \int_\Sigma \frac{\partial G_1(\boldsymbol{N}_{j,p}, \boldsymbol{r}'_{\rm m})}{\partial \nu} u_{2,p'}(\boldsymbol{r}'_{\rm m}) \, {\rm d}\boldsymbol{r}'_{\rm m} \tag{5.21}$$

$$A_{22}^j(p, p') = \int_\Sigma \frac{\partial G_2(\boldsymbol{N}_{j,p}, \boldsymbol{r}'_{\rm m})}{\partial \nu} u_{2,p'}(\boldsymbol{r}'_{\rm m}) \, {\rm d}\boldsymbol{r}'_{\rm m} \tag{5.22}$$

$$B_{11}^j(p, p') = \int_{\partial\Omega} \frac{G_1(\boldsymbol{N}_{j,p}, \boldsymbol{r}'_{\rm m})}{D_1} u_{1,p'}(\boldsymbol{r}'_{\rm m}) \, {\rm d}\boldsymbol{r}'_{\rm m} \tag{5.23}$$

$$B_{12}^j(p, p') = \int_\Sigma \frac{G_1(\boldsymbol{N}_{j,p}, \boldsymbol{r}'_{\rm m})}{D_1} u_{2,p'}(\boldsymbol{r}'_{\rm m}) \, {\rm d}\boldsymbol{r}'_{\rm m} \tag{5.24}$$

$$B_{22}^j(p, p') = \int_\Sigma \frac{G_2(\boldsymbol{N}_{j,p}, \boldsymbol{r}'_{\rm m})}{D_2} u_{2,p'}(\boldsymbol{r}'_{\rm m}) \, {\rm d}\boldsymbol{r}'_{\rm m}. \tag{5.25}$$

The extra function $\varrho_i^\pm$ in (5.19), arises due to singularities on the boundary. These terms can be calculated by surrounding the point $\boldsymbol{r}_{\rm m}$, which lies on the boundary, by a small hemisphere

σ_ε of radius ε and taking the limit when $\varepsilon \to 0$. However, as shown in [73, 77], this term can be obtained indirectly by utilizing some simple physical considerations. In particular, we have $\varrho_i^+(\boldsymbol{r}) = \varrho_i^-(\boldsymbol{r}) = \frac{1}{2}$ when the observation point lies on a smooth surface, which is the case considered here.

Note that appropriate numerical techniques for handling the singularity of the kernels is required for the diagonal elements of these matrices [71]. We represent (5.19) as

$$\mathsf{T}\boldsymbol{f} = \boldsymbol{Q}, \tag{5.26}$$

where $\boldsymbol{f} = [\boldsymbol{U}_1, \boldsymbol{U}_2, \boldsymbol{V}_2]$ is the discrete version of $\{\mathrm{f}\}$ in (5.16). The matrix T takes the form of a dense unsymmetric block matrix.

The normal flux is not always required. In this case we may take the Schur complement of (5.19) to obtain

$$\begin{pmatrix} \frac{1}{2}\mathsf{I} + \mathsf{A}^1_{11} + \frac{1}{2\zeta}\mathsf{B}^1_{11} & -\mathsf{A}^1_{12} + \mathsf{B}^1_{12}\left(\mathsf{B}^2_{22}\right)^{-1}\left(\frac{1}{2}\mathsf{I} + \mathsf{A}^2_{22}\right) \\ \mathsf{A}^2_{11} + \frac{1}{2\zeta}\mathsf{B}^2_{11} & \frac{1}{2}\mathsf{I} - \mathsf{A}^2_{12} + \mathsf{B}^2_{12}\left(\mathsf{B}^2_{22}\right)^{-1}\left(\frac{1}{2}\mathsf{I} + \mathsf{A}^2_{22}\right) \end{pmatrix} \begin{pmatrix} \boldsymbol{U}_1 \\ \boldsymbol{U}_2 \end{pmatrix} = \begin{pmatrix} \boldsymbol{Q}_1|_{\partial\Omega} \\ \boldsymbol{Q}_1|_{\Sigma} \end{pmatrix}. \tag{5.27}$$

This is a smaller system than (5.19) and therefore computationally cheaper. However, it is still usually lengthy to solve large dense matrix systems.

After calculation of the boundary functions $\{U, J\}$ they can be used to construct internal fields through the application of (5.13)–(5.14), and thence to the construction of a linearized problem such as (2.24). This is sometimes referred to as a *semi-analytic* method for Jacobian construction. In particular, for a single homogeneous region of arbitrary shape we would write, for an internal point $\boldsymbol{r} \in \Omega$ and surface points $\boldsymbol{r}_p \in \partial\Omega$

$$\Phi(\boldsymbol{r}) = G(\boldsymbol{r}, \boldsymbol{r}_\mathrm{s}) - \sum_{p=1}^{N} \left(\frac{\partial G(\boldsymbol{r}, \boldsymbol{r}_p)}{\partial \nu} + \frac{G(\boldsymbol{r}, \boldsymbol{r}_p)}{2\zeta D} \right) U_p. \tag{5.28}$$

Another approximation is obtained if we write (5.27) in the form

$$(\mathsf{I} - \mathsf{G})\,\boldsymbol{U} = \boldsymbol{Q}, \tag{5.29}$$

together with a Neumann series approximation

$$\boldsymbol{U} = (\mathsf{I} + \mathsf{G} + \mathsf{G}^2 + \ldots)\boldsymbol{Q}. \tag{5.30}$$

Truncation of this series at the nth power in G constitutes the nth-order Kirchhoff approximation [78, 79]

$$\begin{aligned} \Phi(\boldsymbol{r}) = G(\boldsymbol{r}, \boldsymbol{r}_\mathrm{s}) &+ \sum_{p=1}^{N} \left(\frac{\partial G(\boldsymbol{r}, \boldsymbol{r}_p)}{\partial \nu} + \frac{G(\boldsymbol{r}, \boldsymbol{r}_p)}{2\zeta D} \right) G(\boldsymbol{r}_p, \boldsymbol{r}_\mathrm{s}) \\ &+ \sum_{p=1}^{N} \sum_{p'=1}^{N} \left(\frac{\partial G(\boldsymbol{r}, \boldsymbol{r}_p)}{\partial \nu} + \frac{G(\boldsymbol{r}, \boldsymbol{r}_p)}{2\zeta D} \right) \left(\frac{\partial G(\boldsymbol{r}_p, \boldsymbol{r}_{p'})}{\partial \nu} + \frac{G(\boldsymbol{r}_p, \boldsymbol{r}_{p'})}{2\zeta D} \right) G(\boldsymbol{r}_{p'}, \boldsymbol{r}_\mathrm{s}) \\ &\vdots \end{aligned} \tag{5.31}$$

The Kirchhoff approximations are thus seen to be of the same form as the Born series, with the perturbations being taken as single and double layer potential operators on the boundary.

5.3. Monte Carlo methods

Monte Carlo methods are highly prevalent in many models of radiation in tissue. Within optical tomography, one of the early cited works include Prahl *et al* [80] dating from 1989 and

Wang *et al* [81] describing multi-layered tissues. For the more complex geometries involved in modelling light propagation in the head, [82, 83] describe 3D voxel-based MC models including anisotropy [84].

Use of the MC model within image reconstruction is usually limited to the construction of a linear perturbation model for reconstructing the difference in optical properties from changes in the data [83, 85] with [86] describing the use of MC to recover optical absorption changes in a layered model

Because of the heavy computational cost of Monte Carlo methods, GPU-based fast MC was investigated in [87]. Another way to 'parallelize' computation of MC solutions with different optical parameters (μ_s' and μ_a) is to use white Monte Carlo. Absorption can always be scaled, but the problem of photon termination remains (if absorption is not known when photon packets are followed). In a homogeneous semi-infinite model also scattering can be scaled, but this is not applicable to heterogeneous models [88].

5.4. Hybrid methods

5.4.1. Coupled radiative transport and diffusion. To overcome the limitations of diffusion theory in the proximity of the light sources, hybrid methods which combine Monte Carlo simulation with diffusion theory have been reported [89, 90] and have been applied to media with low-scattering and non-scattering regions [91]. This approach still suffers from the time-consuming nature of the Monte Carlo methods and often require iterative mapping between the models which increases computation times even more.

The Fokker–Planck equation (3.13), which has been found to describe light propagation accurately when the scattering is strongly peaked in the forward direction, can be utilized at the small depths below highly collimated light sources [92].

A coupled radiative transfer equation and diffusion approximation model was used in [93] for media which contained strongly absorbing and low-scattering regions and in [51, 52] to describe light propagation in turbid medium containing highly collimated light sources. In [61] this model was extended to describe light propagation in turbid medium with low-scattering and non-scattering regions. In the coupled approach, the RTE is used as a forward model in sub-domains in which the assumptions of the DA are not valid. This includes the regions in the proximity of the source and boundary and the low-scattering and non-scattering regions. The DA is used as a forward model elsewhere in the domain. The RTE and DA are coupled through boundary conditions between the RTE and DA sub-domains and they are solved simultaneously using the FEM.

5.4.2. The void problem. The void problem refers to the difficulty of modelling a non-scattering region within a highly scattering one. This has particular relevance in optical tomography applied to the brain where the presence of the cerebral spinal fluid (CSF) has just such a property [94].

The effect of void regions on particle transport was identified in early studies of the Boltzmann equation for the propagation of neutrons in [59, 60, 95, 96]. In applications of optical tomography the effect of the voids has been studied more recently in [55, 61, 93, 97–101], and it has been found that under these circumstances, most numerical solutions to the transport equation, apart from Monte Carlo, run into difficulties so that special methods need to be developed. One proposed solution is the *radiosity-diffusion* model [97, 99, 102]. The principle here is to model propagation in voids through geometrical optics, assuming that the distribution of directions is given by a *non-local* boundary condition at a diffuse/non-diffuse

interface

$$\Phi(\boldsymbol{r}) + 2\zeta D\hat{\nu}\cdot\nabla\Phi(\boldsymbol{r}) = \frac{1}{\pi}\int_{\partial\overline{\Omega}} \cos\theta\cos\theta'\left(\Phi(\boldsymbol{r}') - 2D\left(\frac{1+R^{(1)}}{1-R^{(0)}}\right)\hat{\nu}\cdot\nabla\Phi(\boldsymbol{r}')\right) \times (1-|R(\theta)|^2)h(\boldsymbol{r},\boldsymbol{r}')\frac{\exp\left[-\left(\mu_{\mathrm{a}} + \mathrm{i}\frac{\omega}{c}\right)|\boldsymbol{r}-\boldsymbol{r}'|\right]}{|\boldsymbol{r}-\boldsymbol{r}'|^2}\,\mathrm{d}\boldsymbol{r}' \tag{5.32}$$

$$\boldsymbol{r},\boldsymbol{r}'\in\partial\overline{\Omega} \quad \cos\theta = \hat{\nu}(\boldsymbol{r})\cdot\frac{\boldsymbol{r}-\boldsymbol{r}'}{|\boldsymbol{r}-\boldsymbol{r}'|},\ \cos\theta' = \hat{\nu}(\boldsymbol{r}')\cdot\frac{\boldsymbol{r}-\boldsymbol{r}'}{|\boldsymbol{r}-\boldsymbol{r}'|}$$

where $\partial\overline{\Omega}$ is the surface of the void region and $h(\boldsymbol{r},\boldsymbol{r}')$ is a *visibility flag* that is unity if $\boldsymbol{r},\boldsymbol{r}'$ are in line of sight across the void, and zero otherwise. See [103] for a detailed derivation. Both a FEM-radiosity [99] and BEM-radiosity [104] implementation have been presented. The radiosity-diffusion model was used in [105–108] where a boundary recovery method was also developed for the inverse problem.

In [101, 109] a different approach was proposed in which the void was replaced by an interface condition with an anisotropic Laplace–Beltrami operator. The tangential diffusion induced by this anisotropy was related to the void width and curvature.

6. Direct inversion methods

By direct inversion we mean the use of inversion formulas and associated fast image reconstruction algorithms. In optical tomography such formulas exist for particular experimental geometries, including those with planar, cylindrical and spherical boundaries. In this section, we consider several different inverse problems organized by length scale. We begin with the macroscopic case and proceed downwards.

6.1. Diffuse optical tomography

As described in section 4, our aim is to reconstruct the absorption and diffusion coefficients of a macroscopic medium from boundary measurements. We first consider the linearized inverse problem in one dimension and then discuss direct methods for linear and nonlinear inversion in higher dimensions.

6.1.1. One-dimensional problem. We start by studying the time-dependent inverse problem in one dimension, which illustrates many features of the three-dimensional case. Let Ω be the half-line $x \geqslant 0$. The energy density Φ obeys the one-dimensional version of the time-dependent diffusion equation (3.73)

$$\frac{\partial}{\partial t}\Phi(x,t) = D\frac{\partial^2}{\partial x^2}\Phi(x,t) - c\mu_a(x)\Phi(x,t), \qquad x\in\Omega, \tag{6.1}$$

where the diffusion coefficient D is assumed to be constant, an assumption that will be relaxed later. The energy density is taken to obey the initial and boundary conditions

$$\Phi(x,0) = \delta(x-x_1), \tag{6.2}$$

$$\Phi(0,t) - \ell_{\mathrm{ext}}\frac{\partial\Phi}{\partial x}(0,t) = 0. \tag{6.3}$$

Here the initial condition imposes the presence of a point source of unit-strength at x_1. Since Φ decays exponentially, we consider for $k \geqslant 0$ the Laplace transform

$$\Phi(x,k) = \int_0^\infty \mathrm{e}^{-k^2Dt}\Phi(x,t)\,\mathrm{d}t, \tag{6.4}$$

which obeys the equation

$$-\frac{\mathrm{d}^2\Phi(x)}{\mathrm{d}x^2} + k^2(1+\eta(x))\Phi(x) = \frac{1}{D}\delta(x-x_1), \tag{6.5}$$

where η is the spatially varying part of the absorption, which is defined by $\eta = c\mu_a/(Dk^2) - 1$. The solution to the forward problem is given by the integral equation

$$\Phi(x) = \Phi_0(x) - k^2 \int_\Omega G(x, y)\Phi(y)\eta(y)\,\mathrm{d}y, \tag{6.6}$$

where Green's function is of the form

$$G(x, y) = \frac{1}{2Dk}\left(\mathrm{e}^{-k|x-y|} + \frac{1-k\ell_{\mathrm{ext}}}{1+k\ell_{\mathrm{ext}}}\,\mathrm{e}^{-k|x+y|}\right), \tag{6.7}$$

and Φ_0 is the unperturbed field, which obeys (3.3) with $\eta = 0$. The above integral equation may be linearized with respect to η by replacing Φ on the right-hand side by Φ_0. This approximation is accurate when $\mathrm{supp}(\eta)$ and η are small. If we introduce the scattering data $\Phi_s = \Phi_0 - \Phi$ and perform the above linearization, we obtain

$$\Phi_s(x_1, x_2) = k^2 \int_\Omega G(x_1, y)G(y, x_2)\eta(y)\,\mathrm{d}y. \tag{6.8}$$

Here $\Phi_s(x_1, x_2)$ is proportional to the change in intensity due to a point source at x_1 that is measured by a detector at x_2.

If the source and detector are placed at the origin ($x_1 = x_2 = 0$), (6.8) becomes, upon using (6.7) and omitting overall constants

$$\Phi_s(k) = \int_0^\infty \mathrm{e}^{-kx}\eta(x)\,\mathrm{d}x, \tag{6.9}$$

where the dependence of Φ_s on k has been made explicit. Thus, the linearized inverse problem can be seen to correspond to inverting the Laplace transform of η. Inversion of the Laplace transform is the paradigmatic exponentially ill-posed problem. It can be analysed following [110]. Equation (6.9) defines an operator $A : \eta \mapsto \Phi_s$ which is bounded and self-adjoint on $L^2([0, \infty])$. The singular functions f and g of A satisfy

$$A^*Af = \sigma^2 f, \qquad AA^*g = \sigma^2 g, \tag{6.10}$$

where σ is the corresponding singular value. In addition, f and g are related by

$$Af = \sigma g, \qquad A^*g = \sigma f. \tag{6.11}$$

If we observe that $A^*A(x, y) = 1/(x + y)$ and use the identity

$$\int_0^\infty \frac{y^a}{1+y}\,\mathrm{d}y = \frac{\pi}{\sin(a+1)\pi}, \qquad -1 \leqslant \mathrm{Re}(a) < 0, \tag{6.12}$$

we see that

$$f_s(x) = g_s^*(x) = \frac{1}{\sqrt{2\pi}}x^{-\frac{1}{2}+\mathrm{i}s}, \qquad s \in \mathbb{R} \tag{6.13}$$

and

$$\sigma_s^2 = \frac{\pi}{\cosh(\pi s)} \sim \mathrm{e}^{-\pi|s|}. \tag{6.14}$$

Note that the singular values of A are exponentially small, which gives rise to severe ill-posedness. Using the above, we can write an inversion formula for (6.7) in the form

$$\eta(x) = \int_0^\infty \mathrm{d}k \int_{-\infty}^\infty \mathrm{d}s\, \mathcal{R}\left(\frac{1}{\sigma_s}\right) f_s(x)g_s^*(k)\Phi_s(k), \tag{6.15}$$

where the regularizer $\mathcal{R}$ has been introduced to control the contribution of small singular values.

6.1.2. Linearized inverse problem. We now consider the linearized inverse problem in three dimensions. To indicate the parallels with the one-dimensional case, we will find it convenient to work in the half-space geometry. Extensions to other geometries, including those with planar, cylindrical and spherical boundaries is also possible [111–113]. In particular, we note that the slab geometry is often employed in optical mammography and small animal imaging. As before, we define the scattering data $\Phi_s = \Phi_{\text{inc}} - \Phi$. Linearizing (3.81) with respect to $\delta\mu_a$ and δD we find that, up to an overall constant, Φ_s obeys the integral equation

$$\Phi_s(\boldsymbol{r}_1, \boldsymbol{r}_2) = \int \mathrm{d}\boldsymbol{r}\,[G(\boldsymbol{r}_1, \boldsymbol{r})G(\boldsymbol{r}, \boldsymbol{r}_2)c\delta\mu_a(\boldsymbol{r}) + \nabla_r G(\boldsymbol{r}_1, \boldsymbol{r}) \cdot \nabla_r G(\boldsymbol{r}, \boldsymbol{r}_2)\delta D(\boldsymbol{r})]\,, \tag{6.16}$$

where $\boldsymbol{r}_1$ is the position of the source, $\boldsymbol{r}_2$ is the position of the detector and we have integrated by parts to evaluate the action of the operator V. Green's function in the half-space $z \geqslant 0$ is given by the plane-wave decomposition

$$G(\boldsymbol{r}, \boldsymbol{r}') = \int \frac{\mathrm{d}\boldsymbol{q}}{(2\pi)^2} g(\boldsymbol{q}; z, z') \exp[\mathrm{i}\boldsymbol{q} \cdot (\boldsymbol{\rho} - \boldsymbol{\rho}')], \tag{6.17}$$

where we have used the notation $\boldsymbol{r} = (\boldsymbol{\rho}, z)$. If either $\boldsymbol{r}$ or $\boldsymbol{r}'$ lies in the plane $z = 0$, then

$$g(\boldsymbol{q}; z, z') = \frac{\ell_{\text{ext}}}{D_0} \frac{\exp[-Q(\boldsymbol{q})|z - z'|]}{Q(\boldsymbol{q})\ell_{\text{ext}} + 1}, \tag{6.18}$$

where

$$Q(\boldsymbol{q}) = \sqrt{q^2 + k^2}. \tag{6.19}$$

The inverse problem is to recover $\delta\mu_a$ and δD from boundary measurements. To proceed, we introduce the Fourier transform of Φ_s with respect to the source and detector coordinates according to

$$\tilde{\Phi}_s(\boldsymbol{q}_1, \boldsymbol{q}_2) = \int \mathrm{d}\boldsymbol{\rho}_1\, \mathrm{d}\boldsymbol{\rho}_2\, \mathrm{e}^{\mathrm{i}(\boldsymbol{q}_1 \cdot \boldsymbol{\rho}_1 + \boldsymbol{q}_2 \cdot \boldsymbol{\rho}_2)} \Phi_s(\boldsymbol{\rho}_1, 0; \boldsymbol{\rho}_2, 0). \tag{6.20}$$

If we define

$$\psi(\boldsymbol{q}_1, \boldsymbol{q}_2) = (Q(\boldsymbol{q}_1)\ell_{\text{ext}} + 1)(Q(\boldsymbol{q}_2)\ell_{\text{ext}} + 1)\tilde{\Phi}_s(\boldsymbol{q}_1, \boldsymbol{q}_2) \tag{6.21}$$

and make use of (6.17) and (6.16), we find that

$$\psi(\boldsymbol{q}_1, \boldsymbol{q}_2) = \int \mathrm{d}\boldsymbol{r} \mathrm{e}^{\mathrm{i}(\boldsymbol{q}_1 + \boldsymbol{q}_2) \cdot \boldsymbol{\rho}} \left[\kappa_A(\boldsymbol{q}_1, \boldsymbol{q}_2; z)\delta\mu_a(\boldsymbol{r}) + \kappa_D(\boldsymbol{q}_1, \boldsymbol{q}_2; z)\delta D(\boldsymbol{r})\right]. \tag{6.22}$$

Here

$$\kappa_A(\boldsymbol{q}_1, \boldsymbol{q}_2; z) = c \exp[-(Q(\boldsymbol{q}_1) + Q(\boldsymbol{q}_2))z], \tag{6.23}$$

$$\kappa_D(\boldsymbol{q}_1, \boldsymbol{q}_2; z) = -(\boldsymbol{q}_1 \cdot \boldsymbol{q}_2 + Q(\boldsymbol{q}_1)Q(\boldsymbol{q}_2)) \exp[-(Q(\boldsymbol{q}_1) + Q(\boldsymbol{q}_2))z]. \tag{6.24}$$

We now change variables according to

$$\boldsymbol{q}_1 = \boldsymbol{q} + \boldsymbol{p}/2, \qquad \boldsymbol{q}_2 = \boldsymbol{q} - \boldsymbol{p}/2, \tag{6.25}$$

where $\boldsymbol{q}$ and $\boldsymbol{p}$ are independent two-dimensional vectors and rewrite (6.22) as

$$\psi(\boldsymbol{q} + \boldsymbol{p}/2, \boldsymbol{q} - \boldsymbol{p}/2) = \int \mathrm{d}\boldsymbol{r} \exp(-\mathrm{i}\boldsymbol{q} \cdot \boldsymbol{\rho})[\kappa_A(\boldsymbol{p}, \boldsymbol{q}; z)\delta\mu_a(\boldsymbol{r}) + \kappa_D(\boldsymbol{p}, \boldsymbol{q}; z)\delta D(\boldsymbol{r})], \tag{6.26}$$

The above result has the structure of a Fourier–Laplace transform in which $\delta\mu_a$ and δD are related to ψ. This relation can be used to obtain an inversion formula for the integral equation (6.22). To proceed, we note that the Fourier transform in the transverse direction can be

inverted separately from the Laplace transform in the longitudinal direction. We thus arrive at the result

$$\psi(\boldsymbol{q}+\boldsymbol{p}/2, \boldsymbol{q}-\boldsymbol{p}/2) = \int [\kappa_A(\boldsymbol{p}, \boldsymbol{q}; z)\delta\tilde{\mu}_a(\boldsymbol{q}, z) + \kappa_D(\boldsymbol{p}, \boldsymbol{q}; z)\delta\tilde{D}(\boldsymbol{q}, z)]\,\mathrm{d}z, \tag{6.27}$$

where $\delta\tilde{\mu}_a$ and $\delta\tilde{D}$ denote the two-dimensional Fourier transforms. For fixed $\boldsymbol{q}$, equation (6.27) defines an integral equation for $\delta\tilde{\mu}_a(\boldsymbol{q}, z)$ and $\delta\tilde{D}(\boldsymbol{q}, z)$. It is readily seen that the minimum L^2 norm solution to (6.27) has the form

$$\delta\tilde{\mu}_a(\boldsymbol{q}, z) = \int \mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{p}'\kappa_A^*(\boldsymbol{p}, \boldsymbol{q}; z)M^{-1}(\boldsymbol{p}, \boldsymbol{p}'; \boldsymbol{q})\psi(\boldsymbol{p}'+\boldsymbol{q}/2, \boldsymbol{p}'-\boldsymbol{q}/2), \tag{6.28}$$

$$\delta\tilde{D}(\boldsymbol{q}, z) = \int \mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{p}'\kappa_D^*(\boldsymbol{p}, \boldsymbol{q}; z)M^{-1}(\boldsymbol{p}, \boldsymbol{p}'; \boldsymbol{q})\psi(\boldsymbol{p}'+\boldsymbol{q}/2, \boldsymbol{p}'-\boldsymbol{q}/2), \tag{6.29}$$

where the matrix elements of M are given by the integral

$$M(\boldsymbol{p}, \boldsymbol{p}'; \boldsymbol{q}) = \int \left[\kappa_A(\boldsymbol{p}, \boldsymbol{q}; z)\kappa_A^*(\boldsymbol{p}', \boldsymbol{q}; z) + \kappa_D(\boldsymbol{p}, \boldsymbol{q}; z)\kappa_D^*(\boldsymbol{p}', \boldsymbol{q}; z)\right]\mathrm{d}z. \tag{6.30}$$

Finally, we apply the inverse Fourier transform in the transverse direction to arrive at the inversion formula

$$\delta\mu_a(\boldsymbol{r}) = \int \frac{\mathrm{d}\boldsymbol{q}}{(2\pi)^2}\,\mathrm{e}^{-\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{\rho}} \int \mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{p}'\kappa_A^*(\boldsymbol{p}, \boldsymbol{q}; z)M^{-1}(\boldsymbol{p}, \boldsymbol{p}'; \boldsymbol{q})\psi(\boldsymbol{q}'+\boldsymbol{p}/2, \boldsymbol{q}'-\boldsymbol{p}/2), \tag{6.31}$$

$$\delta D(\boldsymbol{r}) = \int \frac{\mathrm{d}\boldsymbol{q}}{(2\pi)^2}\,\mathrm{e}^{-\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{\rho}} \int \mathrm{d}\boldsymbol{p}\,\mathrm{d}\boldsymbol{p}'\kappa_D^*(\boldsymbol{p}, \boldsymbol{q}; z)M^{-1}(\boldsymbol{p}, \boldsymbol{p}'; \boldsymbol{q})\psi(\boldsymbol{q}'+\boldsymbol{p}/2, \boldsymbol{q}'-\boldsymbol{p}/2). \tag{6.32}$$

Several remarks on the above result are necessary. First, implementation of (6.31) and (6.32) requires regularization to stabilize the computation of the inverse of the matrix M. Second, sampling of the data function Φ_s is easily incorporated. The Fourier transform in (6.20) is replaced by a lattice Fourier transform. If the corresponding wave vectors $\boldsymbol{q}$, $\boldsymbol{q}_2$ are restricted to the first Brillouin zone of the lattice, then the inversion formula (6.31) and (6.32) recovers a bandlimited approximation to the coefficients $\delta\mu_a$ and δD [113, 114]. Third, the resolution of reconstructed images in the transverse and longitudinal directions is, in general, quite different. The transverse resolution is controlled by sampling and is determined by the highest spatial frequency that is present in the data. The longitudinal resolution is much lower due to the severe ill-posedness of the Laplace transform inversion which is implicit in (6.26), similar to the one-dimensional case discussed in section 6.1.1. Finally, the inversion formula (6.31) and (6.32) can be used to develop a fast image reconstruction algorithm whose computational complexity scales as $O(NM\log M)$, where M is the number of detectors, N is the number of sources and $M \gg N$ [113]. The algorithm has recently been tested in noncontact optical tomography experiments. Quantitative reconstructions of complex phantoms with millimeter-scale features located centimeters within a highly scattering medium have been reported [115, 116]. Data sets of order 10^8 source–detector pairs can be reconstructed in approximately 1 min of CPU time on a 1.5 GHz computer [115].

6.1.3. Boundary removal. The direct inversion method described in section 6.1.2 is applicable to relatively simple geometries, including those with planar, cylindrical and spherical boundaries. An extension of the method can be employed to handle more complex geometries [79]. Consider a bounded domain Ω in which the diffusion equation (3.80) for the

energy density Φ holds. Making use of the Kirchhoff integral as described in section 5.2, we obtain

$$\Phi(\boldsymbol{r}) = \Phi_0(\boldsymbol{r}) - c\int_\Omega G(\boldsymbol{r},\boldsymbol{r}')\delta\mu_a(\boldsymbol{r}')\Phi(\boldsymbol{r}')\,\mathrm{d}\boldsymbol{r}' + D\int_{\partial\Omega}[G(\boldsymbol{r},\boldsymbol{r}')\nabla\Phi(\boldsymbol{r}') - \Phi(\boldsymbol{r}')\nabla_{\boldsymbol{r}'}G(\boldsymbol{r},\boldsymbol{r}')]\cdot\hat{\nu}\,\mathrm{d}\boldsymbol{r}', \tag{6.33}$$

where, for simplicity, we have assumed that the diffusion coefficient is constant throughout Ω. Here

$$\Phi_0(\boldsymbol{r}) = \int_\Omega G(\boldsymbol{r},\boldsymbol{r}')J^-(\boldsymbol{r}')\,\mathrm{d}\boldsymbol{r}' \tag{6.34}$$

and G is the fundamental solution to the diffusion equation defined in (3.85). Using the boundary condition (3.76) and introducing the approximation $\Phi = \Phi_0$ in Ω, we find that (6.33) becomes

$$(\Phi_0 - \Phi - M)(\boldsymbol{r}) = \int_\Omega G(\boldsymbol{r},\boldsymbol{r}')\Phi_0(\boldsymbol{r}')c\delta\mu_a(\boldsymbol{r}')\,\mathrm{d}\boldsymbol{r}', \tag{6.35}$$

where

$$M(\boldsymbol{r}) = \int_{\partial\Omega}\left[\frac{1}{\ell_{\mathrm{ext}}}G(\boldsymbol{r},\boldsymbol{r}') + \hat{\nu}\cdot\nabla_{\boldsymbol{r}'}G(\boldsymbol{r},\boldsymbol{r}')\right]\Phi(\boldsymbol{r}')\,\mathrm{d}\boldsymbol{r}'. \tag{6.36}$$

We observe that the left-hand side of (6.35) is known from measurements. Thus, within the accuracy of the aforementioned approximation, the inverse problem becomes that of recovering $\delta\mu_a$ in an infinite medium with measurements on the surface $\partial\Omega$. Let B denote a ball that contains Ω. Then the field on $\partial\Omega$ can be propagated to ∂B by using the Rayleigh–Sommerfeld formula

$$\Phi(\boldsymbol{r}) = D\int_{\partial\Omega}\hat{\nu}\cdot\nabla_{\boldsymbol{r}'}G(\boldsymbol{r},\boldsymbol{r}')\Phi(\boldsymbol{r}')\,\mathrm{d}\boldsymbol{r}', \qquad \boldsymbol{r}\in\partial B. \tag{6.37}$$

Thus, it is possible to reformulate the inverse problem for a bounded domain in terms of one for an infinite domain, with measurements prescribed on a spherical surface. This latter problem is one for which direct inversion can be performed.

6.1.4. Nonlinear inverse problem. We now consider the nonlinear inverse problem of DOT. The Born series (3.86) can be rewritten in the form

$$\Phi_s(\boldsymbol{r}_1,\boldsymbol{r}_2) = \int \mathrm{d}\boldsymbol{r}\,K_1^i(\boldsymbol{r}_1,\boldsymbol{r}_2;\boldsymbol{r})\eta_i(\boldsymbol{r}) + \int \mathrm{d}\boldsymbol{r}\,\mathrm{d}\boldsymbol{r}'K_2^{ij}(\boldsymbol{r}_1,\boldsymbol{r}_2;\boldsymbol{r},\boldsymbol{r}')\eta_i(\boldsymbol{r})\eta_j(\boldsymbol{r}') + \cdots, \tag{6.38}$$

where

$$\eta(\boldsymbol{r}) = \begin{pmatrix}\eta_1(\boldsymbol{r})\\ \eta_2(\boldsymbol{r})\end{pmatrix} = \begin{pmatrix}\delta\mu_a(\boldsymbol{r})\\ \delta D(\boldsymbol{r})\end{pmatrix}, \tag{6.39}$$

summation over repeated indices is implied with $i, j = 1, 2$. The components of the operators K_1 and K_2 are given by

$$K_1^1(\boldsymbol{r}_1,\boldsymbol{r}_2;\boldsymbol{r}) = cG(\boldsymbol{r}_1,\boldsymbol{r})G(\boldsymbol{r},\boldsymbol{r}_2), \tag{6.40}$$

$$K_1^2(\boldsymbol{r}_1,\boldsymbol{r}_2;\boldsymbol{r}) = \nabla_{\boldsymbol{r}}G(\boldsymbol{r}_1,\boldsymbol{r})\cdot\nabla_{\boldsymbol{r}}G(\boldsymbol{r},\boldsymbol{r}_2), \tag{6.41}$$

$$K_2^{11}(\boldsymbol{r}_1,\boldsymbol{r}_2;\boldsymbol{r},\boldsymbol{r}') = -c^2G(\boldsymbol{r}_1,\boldsymbol{r})G(\boldsymbol{r},\boldsymbol{r}')G(\boldsymbol{r}',\boldsymbol{r}_2), \tag{6.42}$$

$$K_2^{12}(\boldsymbol{r}_1,\boldsymbol{r}_2;\boldsymbol{r},\boldsymbol{r}') = -cG(\boldsymbol{r}_1,\boldsymbol{r})\nabla_{\boldsymbol{r}'}G(\boldsymbol{r},\boldsymbol{r}')\cdot\nabla_{\boldsymbol{r}'}G(\boldsymbol{r}',\boldsymbol{r}_2), \tag{6.43}$$

$$K_2^{21}(\boldsymbol{r}_1, \boldsymbol{r}_2; \boldsymbol{r}, \boldsymbol{r}') = -c\nabla_{\boldsymbol{r}} G(\boldsymbol{r}_1, \boldsymbol{r}) \cdot \nabla_{\boldsymbol{r}} G(\boldsymbol{r}, \boldsymbol{r}') G(\boldsymbol{r}', \boldsymbol{r}_2), \tag{6.44}$$

$$K_2^{22}(\boldsymbol{r}_1, \boldsymbol{r}_2; \boldsymbol{r}, \boldsymbol{r}') = -\nabla_{\boldsymbol{r}} G(\boldsymbol{r}_1, \boldsymbol{r}) \cdot \nabla_{\boldsymbol{r}} [\nabla_{\boldsymbol{r}'} G(\boldsymbol{r}, \boldsymbol{r}') \cdot \nabla_{\boldsymbol{r}'} G(\boldsymbol{r}', \boldsymbol{r}_2)]. \tag{6.45}$$

It will prove useful to express the Born series as a formal power series in tensor powers of η of the form

$$\Phi_s = K_1\eta + K_2\eta \otimes \eta + K_3\eta \otimes \eta \otimes \eta + \cdots. \tag{6.46}$$

The solution to the nonlinear inverse problem of DOT may be formulated from the ansatz that η may be expressed as a series in tensor powers of Φ_s of the form

$$\eta = \mathcal{K}_1\Phi_s + \mathcal{K}_2\Phi_s \otimes \Phi_s + \mathcal{K}_3\Phi_s \otimes \Phi_s \otimes \Phi_s + \cdots, \tag{6.47}$$

where $\mathcal{K}_j$s are operators which are to be determined [117–120]. To proceed, we substitute the expression (6.46) for Φ_s into (6.47) and equate terms with the same tensor power of η. We thus obtain the relations

$$\mathcal{K}_1 K_1 = I, \tag{6.48}$$

$$\mathcal{K}_2 K_1 \otimes K_1 + \mathcal{K}_1 K_2 = 0, \tag{6.49}$$

$$\mathcal{K}_3 K_1 \otimes K_1 \otimes K_1 + \mathcal{K}_2 K_1 \otimes K_2 + \mathcal{K}_2 K_2 \otimes K_1 + \mathcal{K}_1 K_3 = 0, \tag{6.50}$$

$$\sum_{m=1}^{j-1} \mathcal{K}_m \sum_{i_1+\cdots+i_m=j} K_{i_1} \otimes \cdots \otimes K_{i_m} + \mathcal{K}_j K_1 \otimes \cdots \otimes K_1 = 0, \tag{6.51}$$

which may be solved for $\mathcal{K}_j$'s with the result

$$\mathcal{K}_1 = K_1^+, \tag{6.52}$$

$$\mathcal{K}_2 = -\mathcal{K}_1 K_2 \mathcal{K}_1 \otimes \mathcal{K}_1, \tag{6.53}$$

$$\mathcal{K}_3 = -\left(\mathcal{K}_2 K_1 \otimes K_2 + \mathcal{K}_2 K_2 \otimes K_1 + \mathcal{K}_1 K_3\right) \mathcal{K}_1 \otimes \mathcal{K}_1 \otimes \mathcal{K}_1, \tag{6.54}$$

$$\mathcal{K}_j = -\left(\sum_{m=1}^{j-1} \mathcal{K}_m \sum_{i_1+\cdots+i_m=j} K_{i_1} \otimes \cdots \otimes K_{i_m}\right) \mathcal{K}_1 \otimes \cdots \otimes \mathcal{K}_1. \tag{6.55}$$

We will refer to (6.47) as the inverse series. Here we note several of its properties. First, K_1^+ is the regularized pseudoinverse of the operator K_1. The singular value decomposition of the operator K_1^+ can be computed analytically for particular geometries, as explained in section 3.5. Since the operator $\mathcal{K}_1$ is unbounded, regularization of K_1^+ is required to control the ill-posedness of the inverse problem. Second, the coefficients in the inverse series have a recursive structure. The operator $\mathcal{K}_j$ is determined by the coefficients of the Born series $K_1, K_2, \ldots, K_j$. Third, the inverse scattering series can be represented in diagrammatic form as shown in figure 2. A solid line corresponds to a factor of G, a wavy line to the incident field and a solid vertex ($\bullet$) to $\mathcal{K}_1\Phi_s$, and the effect of the final application of $\mathcal{K}_1$ is to join the ends of the diagrams. Note that the recursive structure of the series is evident in the diagrammatic expansion which is shown to be third order. Finally, inversion of only the linear term in the Born series is required to compute the inverse series to all orders. Thus, an ill-posed nonlinear inverse problem is reduced to an ill-posed linear inverse problem plus a well-posed nonlinear problem, namely the computation of the higher order terms in the series.

We now characterize the convergence of the inverse series. We restrict our attention to the case of a uniformly scattering medium for which $\eta = \delta\mu_a$. To proceed, we require an estimate on the L^2 norm of the operator $\mathcal{K}_j$. We define the constants μ and ν by

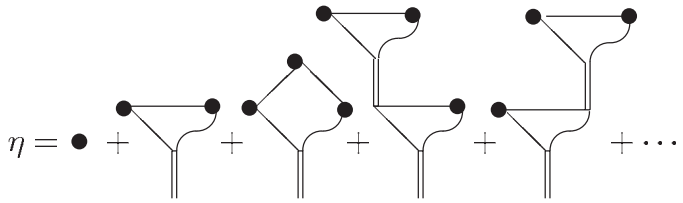

Figure 2. Diagrammatic representation of the inverse scattering series.

$$\mu = \sup_{\boldsymbol{r}\in B_a} k^2 \|G_0(\boldsymbol{r}, \cdot)\|_{L^2(B_a)}. \tag{6.56}$$

$$\nu = k^2 |B_a|^{1/2} \sup_{\boldsymbol{r}\in B_a} \|G_0(\boldsymbol{r}, \cdot)\|_{L^2(\partial\Omega)}. \tag{6.57}$$

Here B_a denotes a ball of radius a which contains the support of η. It can be shown [120] that if $(\mu_p + \nu_p)\|\mathcal{K}_1\|_2 < 1$ then the operator

$$\mathcal{K}_j : L^2(\partial\Omega \times \cdots \times \partial\Omega) \longrightarrow L^2(B_a) \tag{6.58}$$

defined by (6.55) is bounded and

$$\|\mathcal{K}_j\|_2 \leqslant C(\mu + \nu)^j \|\mathcal{K}_1\|_2^j, \tag{6.59}$$

where $C = C(\mu_p, \nu_p, \|\mathcal{K}_1\|_p)$ is independent of j.

We can now state the main result on the convergence of the inverse series.

Theorem 6.1 ([120]). *The inverse scattering series converges in the L^2 norm if $\|\mathcal{K}_1\|_2 < 1/(\mu + \nu)$ and $\|\mathcal{K}_1\Phi_s\|_{L^2(B_a)} < 1/(\mu + \nu)$. Furthermore, the following estimate for the series limit $\tilde{\eta}$ holds*

$$\left\| \tilde{\eta} - \sum_{j=1}^{N} \mathcal{K}_j \Phi_s \otimes \cdots \otimes \Phi_s \right\|_{L^p(B_a)} \leqslant C \frac{\left[(\mu_p + \nu_p)\|\mathcal{K}_1\Phi_s\|_{L^2(B_a)}\right]^{N+1}}{1 - (\mu + \nu)\|\mathcal{K}_1\Phi_s\|_{L^2(B_a)}},$$

where $C = C(\mu, \nu, \|\mathcal{K}_1\|_2)$ does not depend on N nor on the scattering data Φ_s.

The stability of the limit of the inverse series under perturbations in the scattering data can be analysed as follows:

Theorem 6.2 ([120]). *Let $\|\mathcal{K}_1\|_2 < 1/(\mu + \nu)$ and let Φ_{s1} and Φ_{s2} be scattering data for which $M\|\mathcal{K}_1\|_2 < 1/(\mu + \nu)$, where $M = \max(\|\Phi_{s1}\|_2, \|\Phi_{s2}\|_2)$. Let η_1 and η_2 denote the corresponding limits of the inverse scattering series. Then the following estimate holds*

$$\|\eta_1 - \eta_2\|_{L^2(B_a)} < \tilde{C}\|\Phi_{s1} - \Phi_{s2}\|_{L^2(\partial\Omega\times\partial\Omega)},$$

where $\tilde{C} = \tilde{C}(\mu, \nu, \|\mathcal{K}_1\|_2, M)$ is a constant that is otherwise independent of Φ_{s1} and Φ_{s2}.

It is a consequence of the proof of theorem 6.2 that $\tilde{C}$ is proportional to $\|\mathcal{K}_1\|_2$. Since regularization sets the scale of $\|\mathcal{K}_1\|_2$, it follows that the stability of the nonlinear inverse problem is controlled by the stability of the linear inverse problem.

The limit of the inverse scattering series does not, in general, coincide with η. We characterize the approximation error as follows.

Theorem 6.3 ([120]). *Suppose that $\|\mathcal{K}_1\|_2 < 1/(\mu + \nu)$, $\|\mathcal{K}_1\Phi_s\|_{L^2(B_a)} < 1/(\mu + \nu)$. Let $\mathcal{M} = \max(\|\eta\|_{L^2(B_a)}, \|\mathcal{K}_1 K_1 \eta\|_{L^2(B_a)})$ and assume that $\mathcal{M} < 1/(\mu + \nu)$. Then the norm of*

the difference between the partial sum of the inverse series and the true absorption obeys the estimate

$$\left\| \eta - \sum_{j=1}^{N} \mathcal{K}_j \Phi_s \otimes \cdots \otimes \Phi_s \right\|_{L^2(B_a)} \leqslant C \|(I - \mathcal{K}_1 K_1)\eta\|_{L^2(B_a)} + \tilde{C} \frac{[(\mu_p + \nu_p)\|\mathcal{K}_1\|_2 \|\Phi_s\|]^N}{1 - (\mu + \nu)\|\mathcal{K}_1\|_2 \|\Phi_s\|_2},$$

where $C = C(\mu, \nu, \|\mathcal{K}_1\|_2, \mathcal{M})$ *and* $\tilde{C} = \tilde{C}(\mu, \nu \|\mathcal{K}_1\|_2)$ *are independent of* N *and* Φ_s.

We note that, as expected, the above result shows that regularization of $\mathcal{K}_1$ creates an error in the reconstruction of η. For a fixed regularization, the relation $\mathcal{K}_1 K_1 = I$ holds on a subspace of $L^2(B_a)$ which, in practice, is finite dimensional. By regularizing $\mathcal{K}_1$ more weakly, the subspace becomes larger, eventually approaching all of L^2. However, in this instance, the estimate in theorem 6.3 would not hold since $\|\mathcal{K}_1\|_2$ is so large that the inverse scattering series would not converge. Nevertheless, theorem 6.3 does describe what can be reconstructed exactly, namely those η for which $\mathcal{K}_1 K_1 \eta = I$. That is, if we know *a priori* that η belongs to a particular finite-dimensional subspace of L^2, we can choose $\mathcal{K}_1$ to be a true inverse on this subspace. Then, if $\|\mathcal{K}_1\|_2$ and $\|\mathcal{K}_1 \Phi_s\|_{L^2}$ are sufficiently small, the inverse series will recover η exactly.

It is straightforward to compute the constants μ and ν in dimension three. We have

$$\mu = k^2 \, \mathrm{e}^{-ka/2} \left(\frac{\sinh(ka)}{4\pi k} \right)^{1/2} \tag{6.60}$$

$$\nu \leqslant k^2 |\partial\Omega\| B_a|^{1/2} \frac{\mathrm{e}^{-2k \operatorname{dist}(\partial\Omega, B_a)}}{(4\pi \operatorname{dist}(\partial\Omega, B_a))^2}. \tag{6.61}$$

Note that ν is exponentially small. It can be seen that the radius of convergence of the inverse series $R = 1/(\mu + \nu) \lesssim O(1/(ka)^{3/2})$ when $ka \gg 1$.

Numerical studies of the inverse scattering series have been performed [121]. For inhomogeneities with radial symmetry, exact solutions to the forward problem were used as scattering data and reconstructions were computed to fifth order in the inverse series. It was found that the series appears to converge quite rapidly for low-contrast objects. As the contrast is increased, the higher order terms systematically improve the reconstructions until, at sufficiently large contrast, the series diverges.

6.2. *Inverse transport*

We now turn our attention to the inverse problem for the RTE. This is a very large subject in its own right. Indeed, a recent topical review has already covered the key mathematical issues regarding existence, uniqueness and stability of the inverse transport problem [23]. Here we aim to discuss the inverse problem in some special cases which lead to direct inversion procedures and fast algorithms. See [122] for further details. As in section 6.1.2, we will work in the $z \geqslant 0$ half-space with the source and detector located on the $z = 0$ plane. The source is assumed to be pointlike and oriented in the inward normal direction. The light exiting the medium is further assumed to pass through a normally oriented angularly selective aperture which collects all photons with intensity

$$\mathcal{I}(\boldsymbol{r}) = \int_{\hat{\nu} \cdot \hat{\mathbf{s}} > 0} \hat{\nu} \cdot \hat{\mathbf{s}} w(\hat{\mathbf{s}}) I(\boldsymbol{r}, \hat{\mathbf{s}}) \, \mathrm{d}\hat{\mathbf{s}}, \tag{6.62}$$

where w accounts for the effect of the aperture and the integration is carried out over all outgoing directions. When the aperture selects only photons travelling in the normal direction,

then $w(\hat{\mathbf{s}}) = \delta(\hat{\mathbf{s}} - \hat{\nu})$ and $\mathcal{I}(\boldsymbol{r}) = I(\boldsymbol{r}, \hat{\nu})$. This case is relevant to noncontact measurements in which the lens is focused at infinity. The case of complete angularly averaged data corresponds to $w(\hat{\mathbf{s}}) = 1$.

We now consider the linearized inverse problem. If the medium is inhomogeneously absorbing, it follows from (3.46) and (6.62) that the change in intensity measured relative to a homogeneous reference medium with absorption $\bar{\mu}_a$ is proportional to the data function $\Phi_s(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2)$ which is given by

$$\Phi_s(\boldsymbol{\rho}_1, \boldsymbol{\rho}_2) = \int_{\hat{\nu}\cdot\hat{\mathbf{s}}>0} \hat{\nu}\cdot\hat{\mathbf{s}} w(\hat{\mathbf{s}}) \Phi_s(\boldsymbol{\rho}_1, 0, \hat{z}; \boldsymbol{\rho}_2, 0, -\hat{\mathbf{s}})\, \mathrm{d}\hat{\mathbf{s}}. \tag{6.63}$$

Following the development in section 6.1.2, we consider the Fourier transform of Φ_s with respect to the source and detector coordinates. Upon inserting the plane-wave decomposition for G given by (3.61) into (3.46) and carrying out the Fourier transform, we find that

$$\begin{aligned}\tilde{\Phi}_s(\boldsymbol{q}_1, \boldsymbol{q}_2) = \sum_{\mu_1,\mu_2} M_{\mu_1\mu_2}(\boldsymbol{q}_1, \boldsymbol{q}_2) \int \mathrm{d}\boldsymbol{r} \exp\big[\mathrm{i}(\boldsymbol{q}_1 + \boldsymbol{q}_2)\cdot\boldsymbol{\rho} \\ - (Q_{\mu_1}(\boldsymbol{q}_1) + Q_{\mu_2}(\boldsymbol{q}_2) z\big] \delta\mu_a(\boldsymbol{r}),\end{aligned} \tag{6.64}$$

where

$$M_{\mu_1\mu_2}(\boldsymbol{q}_1, \boldsymbol{q}_2) = \sum_{l_1 m_1, l_1' m_1'} \sum_{l_2 m_2} \mathcal{B}^{l_1 m_1}_{l_1' m_1'}(\boldsymbol{q}_1, \mu_1) \mathcal{B}^{l_1' m_1'}_{l_2 m_2}(\boldsymbol{q}_2, \mu_2) \int_{\hat{\nu}\cdot\hat{\mathbf{s}}>0} \hat{\nu}\cdot\hat{\mathbf{s}} w(\hat{\mathbf{s}}) Y_{l_2 m_2}(\hat{\mathbf{s}})\, \mathrm{d}\hat{\mathbf{s}}. \tag{6.65}$$

Equation (6.64) is a generalization of the Fourier–Laplace transform which holds for the diffusion approximation, as discussed in section 6.1.2. It can be seen that (6.64) reduces to the appropriate form in the diffusion limit, since only the smallest discrete eigenvalue contributes.

The inverse problem now consists of recovering $\delta\mu_a$ from $\tilde{\Phi}_s$. To proceed, we make use of the change of variables (6.25) and rewrite (6.64) as

$$\tilde{\Phi}_s(\boldsymbol{q} + \boldsymbol{p}/2, \boldsymbol{q} - \boldsymbol{p}/2) = \int \mathrm{d}z\, K(\boldsymbol{q}, \boldsymbol{p}; z) \tilde{\delta}\mu_a(\boldsymbol{q}, z), \tag{6.66}$$

where $\delta\tilde{\mu}_a(\boldsymbol{q}, z)$ denotes the two-dimensional Fourier transform of $\delta\mu_a$ with respect to its transverse argument and

$$\begin{aligned}K(\boldsymbol{q}, \boldsymbol{p}; z) = \sum_{\mu_1,\mu_2} M_{\mu_1\mu_2}(\boldsymbol{q} + \boldsymbol{p}/2, \boldsymbol{q} - \boldsymbol{p}/2) \\ \times \exp[-(Q_{\mu_1}(\boldsymbol{q} + \boldsymbol{p}/2) + Q_{\mu_2}(\boldsymbol{q} - \boldsymbol{p}/2))z].\end{aligned} \tag{6.67}$$

This change of variables can be used to separately invert the transverse and longitudinal functional dependences of $\delta\mu_a$ since for fixed $\boldsymbol{q}$, (6.66) defines a one-dimensional integral equation for $\delta\tilde{\mu}_a(\boldsymbol{q}, z)$ whose pseudoinverse solution can in principle be computed. We thus obtain a solution to the inverse problem in the form

$$\delta\mu_a(\boldsymbol{r}) = \int \frac{\mathrm{d}\boldsymbol{q}}{(2\pi)^2} \mathrm{e}^{-\mathrm{i}\boldsymbol{q}\cdot\boldsymbol{\rho}} \int \mathrm{d}\boldsymbol{p}\, K^+(z; \boldsymbol{q}, \boldsymbol{p}) \tilde{\Phi}_s(\boldsymbol{q} + \boldsymbol{p}/2, \boldsymbol{q} - \boldsymbol{p}/2), \tag{6.68}$$

where K^+ denotes the pseudoinverse of K. It is important to note the ill-posedness due to the exponential decay of the evanescent modes (3.58) for large z. Therefore, we expect that the resolution in the z direction will degrade with depth, but that sufficiently close to the surface the transverse resolution will be controlled by sampling.

6.3. Single-scattering tomography

Consider an experiment in which a narrow collimated beam is normally incident on a highly scattering medium which has the shape of a slab. Suppose that the slab is sufficiently thin that the incident beam undergoes predominantly single-scattering and that the intensity of transmitted light is measured by an angularly selective detector. If the detector is collinear with the incident beam and its aperture is set to collect photons travelling in the normal direction, then only unscattered photons will be measured. Now, if the aperture is set away from the normal direction, then the detector will not register any photons. However, if the detector is no longer collinear and only photons which exit the slab at a fixed angle are collected, then it is possible to selectively measure single-scattered photons. Note that the contribution of single-scattered photons is described by (3.40).

The inverse problem for single-scattered light is to recover μ_a from measurements of δI as given by (3.40), assuming μ_s and p are known. The more general problem of simultaneously reconstructing μ_a, μ_s and p can also be considered. In either case, what must be investigated is the inversion of the broken-ray Radon transform which is defined as follows. Let f be a sufficiently smooth function. The broken-ray Radon transform is defined by

$$\mathcal{R}_b f(\boldsymbol{r}_1, \hat{\mathbf{s}}_1; \boldsymbol{r}_2, \hat{\mathbf{s}}_2) = \int_{BR(\boldsymbol{r}_1, \hat{\mathbf{s}}_1; \boldsymbol{r}_2, \hat{\mathbf{s}}_2)} f(\boldsymbol{r})\, \mathrm{d}\boldsymbol{r}. \tag{6.69}$$

Here $BR(\boldsymbol{r}_1, \hat{\mathbf{s}}_1; \boldsymbol{r}_2, \hat{\mathbf{s}}_2)$ denotes the broken ray which begins at $\boldsymbol{r}_1$, travels in the direction $\hat{\mathbf{s}}_1$ and ends at $\boldsymbol{r}_2$ in the direction $\hat{\mathbf{s}}_2$. Note that if $\boldsymbol{r}_1, \boldsymbol{r}_2, \hat{\mathbf{s}}_1$ and $\hat{\mathbf{s}}_2$ all lie in the same plane and $\hat{\mathbf{s}}_1$ and $\hat{\mathbf{s}}_2$ point into and out of the slab, then the point of intersection is uniquely determined. Thus, it will suffice to consider the inverse problem in the plane and to reconstruct the function f from two-dimensional slices.

Evidently, the problem of inverting (6.69) is overdetermined. However, if the directions $\hat{\mathbf{s}}_1$ and $\hat{\mathbf{s}}_2$ are taken to be fixed, then the inverse problem is formally determined. To this end, we consider transmission measurements in a slab of width L and choose a coordinate system in which $\hat{\mathbf{s}}_1$ points in the $\hat{\boldsymbol{z}}$ direction. The sources are chosen to be located on the line $z = 0$ in the yz-plane and are taken to point in the $\hat{\boldsymbol{z}}$ direction. The detectors are located on the line $z = L$ and we assume that the angle θ between $\hat{\mathbf{s}}_1$ and $\hat{\mathbf{s}}_2$ is fixed. Under these conditions, it can be seen that the solution to (6.69) is given by

$$f(y, z) = \int \frac{\mathrm{d}k}{2\pi}\, \mathrm{e}^{-\mathrm{i}ky} \left[\cot\left(\frac{\theta}{2}\right) F(k, z) - \mathrm{i}\cot^2\left(\frac{\theta}{2}\right) k\, \mathrm{e}^{-\mathrm{i}\cot(\theta/2)k\xi} \right. \tag{6.70}$$

$$\left. \times \int_0^z \mathrm{e}^{\mathrm{i}\cot(\theta/2)k\xi} F(k, \xi)\, \mathrm{d}\xi \right], \tag{6.71}$$

where

$$F(k, \xi) = \lim_{y_2 \to \Delta y(\xi)} \left(\frac{\partial}{\partial y_2} + \mathrm{i}k \right) \int \mathrm{e}^{\mathrm{i}ky_1} \mathcal{R}_b f(y_1, y_1 + y_2)\, \mathrm{d}y_1 \tag{6.72}$$

and $\Delta y(\xi) = (L - \xi)\tan\theta$.

Equation (6.70) is the inversion formula for the broken-ray Radon transform. We note that in contrast to x-ray computed tomography, it is unnecessary to collect projections along rays which are rotated about the sample. This considerably reduces the complexity of an imaging experiment. We also note that the presence of a derivative in (6.72), as in the Radon inversion formula, means that regularization is essential. Finally, by making use of measurements from multiple detector orientations it is possible to simultaneously reconstruct μ_a and μ_s. See [123] for further details.

7. Numerical inversion methods

The starting point for numerical inversion methods is the definition of a variational form whose minimum represents the solution. The most commonly used is a regularized weighted least-squares functional

$$\mathcal{E}(y, \mathcal{F}(x)) = \tfrac{1}{2}\|y - \mathcal{F}(x)\|^2_{\Gamma_e^{-1}} + \tfrac{1}{2}\|x\|^2_{\Gamma_x^{-1}}. \tag{7.1}$$

From the Bayesian point of view (7.1) represents the negative log of the posterior probability density function

$$\pi(x|y) = \exp(-\mathcal{E}(y, \mathcal{F}(x))) = \pi(y|x)\pi(x), \tag{7.2}$$

where $\pi(y|x)$ is the Gaussian (normal) probability density function of the distribution of the noise in the data with mean zero and covariance Γ_e

$$\pi(y|x) = \pi_{\text{noise}}(y - \mathcal{F}(x)) = \mathcal{N}(0, \Gamma_e) \tag{7.3}$$

and $\pi(x)$ is the Gaussian (normal) prior probability density function of the distribution of x with mean zero and covariance Γ_x. Thus, minimization of (7.1) represents the *maximum a posteriori* (MAP) estimate of (7.2).

The forms in (7.1) and (7.2) can be generalized in two main ways. Firstly the prior may be assumed not normal but of the form

$$\pi(x) = \exp(-\Psi(x)). \tag{7.4}$$

Secondly the noise need not be Gaussian distributed but could follow an alternative probability model such as Poisson. This consideration leads to alternative *data fit functionals* of which the commonest is the cross-entropy or KL divergence

$$KL(y, \mathcal{F}(x)) = \int_{\Upsilon} \left(y \log\left[\frac{y}{\mathcal{F}(x)}\right] - y + \mathcal{F}(x) \right) \mathrm{d}y. \tag{7.5}$$

From an optimization point of view the minimization of (7.1) is known as the regularized output least-squares solution and can be achieved with classical methods if $\mathcal{E}(y, \mathcal{F}(x))$ is convex. Non-convexity does not normally arise from the forward mappings but may arise if the prior is non-convex. An alternative formulation of the problem which has its origins in control theory is known as the 'all-at-once' or 'PDE-constrained method'. In this formulation the solution of the forward problem is treated as a constraint

$$\begin{aligned} &\text{minimize:} \quad \|y - \mathcal{M}U\|^2_{\Gamma_e^{-1}} + \alpha\Psi(x) \\ &\text{subject to:} \quad (2.1)\text{–}(2.3). \end{aligned} \tag{7.6}$$

Several of the problems in section 4 are linear. In addition it is frequently the case that a linearized problem is considered even when the forward problem is nonlinear. Suppose we assume a known linearization point x_0 and we consider the true solution to be a perturbation from this point $x = x_0 + \eta$ then we consider the minimization

$$\eta_* = \arg\min_\eta \left[\tfrac{1}{2}\|y - \mathcal{F}(x_0) - \mathcal{F}'(x_0)\eta\|^2_{\Gamma_e^{-1}} + \alpha\tfrac{1}{2}\|\eta\|^2_{\Gamma_x^{-1}} \right] \tag{7.7}$$

$$= \arg\min_\eta \left[\tfrac{1}{2}\|\delta y - \mathcal{F}'(x_0)\eta\|^2_{\Gamma_e^{-1}} + \alpha\tfrac{1}{2}\|\eta\|^2_{\Gamma_x^{-1}} \right]. \tag{7.8}$$

where $\delta y = y - \mathcal{F}(x_0)$ is the change in measurement assumed to be linearly related to η. Discretization of this problem as discussed in section 2 leads to the form (2.24). In the next section we consider several generic linear solvers as applied in optical tomography.

7.1. Linear methods

Consider the solution of a linear discrete problem

$$\delta \boldsymbol{y} = \mathsf{A}\eta \quad \delta \boldsymbol{y} \in \mathbb{R}^M, \qquad \eta \in \mathbb{R}^N, \qquad \mathsf{A} \in \mathbb{R}^M \times \mathbb{R}^N. \tag{7.9}$$

The weighted least-squares term to be minimized is

$$\begin{aligned} \eta_* &= \arg\min_\eta \left[\tfrac{1}{2} \|\delta \boldsymbol{y} - \mathsf{A}\eta\|^2_{\Gamma_e^{-1}} + \alpha \tfrac{1}{2} \|\eta\|^2_{\Gamma_x^{-1}} \right] \\ &= \arg\min_\eta \left[\tfrac{1}{2} (\delta \boldsymbol{y} - \mathsf{A}\eta)^{\mathrm{T}} \Gamma_e^{-1} (\delta \boldsymbol{y} - \mathsf{A}\eta) + \alpha \tfrac{1}{2} \eta^{\mathrm{T}} \Gamma_x^{-1} \eta \right]. \end{aligned} \tag{7.10}$$

We define factorizations

$$\mathsf{L}_e^{\mathrm{T}} \mathsf{L}_e = \Gamma_e^{-1} \qquad \mathsf{L}_x^{\mathrm{T}} \mathsf{L}_x = \Gamma_x^{-1} \tag{7.11}$$

and the canonical (dimensionless) variables

$$\tilde{\boldsymbol{y}} = \mathsf{L}_e \delta \boldsymbol{y} \qquad \tilde{\mathsf{A}} = \mathsf{L}_e \mathsf{A} \mathsf{L}_x^{-1} \qquad \tilde{\boldsymbol{x}} = \mathsf{L}_x \eta. \tag{7.12}$$

Then the solution to (7.10) is given by

$$\tilde{\boldsymbol{x}}_* = \arg\min_{\tilde{\boldsymbol{x}}} \left[\tilde{\mathcal{E}}_\alpha(\tilde{\boldsymbol{y}}, \tilde{\mathsf{A}}\tilde{\boldsymbol{x}}) := \tfrac{1}{2} \|\tilde{\boldsymbol{y}} - \tilde{\mathsf{A}}\tilde{\boldsymbol{x}}\|^2 + \alpha \tfrac{1}{2} \|\tilde{\boldsymbol{x}}\|^2 \right]. \tag{7.13}$$

In purely algebraic terms, the transformation $\mathsf{A} \to \tilde{\mathsf{A}}$ given in (7.12) is a conditioning step, with L_e, L_x^{-1} the left and right preconditioners respectively [124].

We now consider some linear inversion methods for (7.13).

7.1.1. Newton methods (overdetermined system). A Newton solution for an over determined system ($M > N$) is written

$$\tilde{\boldsymbol{x}}_* = (\tilde{\mathsf{A}}^{\mathrm{T}} \tilde{\mathsf{A}} + \alpha \mathsf{I})^{-1} \tilde{\mathsf{A}}^{\mathrm{T}} \tilde{\boldsymbol{y}} \tag{7.14}$$

$$\begin{aligned} \mathsf{L}_x \eta_* &= \left(\mathsf{L}_x^{-\mathrm{T}} \mathsf{A}^{\mathrm{T}} \Gamma_e^{-1} \mathsf{A} \mathsf{L}_x^{-1} + \alpha \mathsf{I} \right)^{-1} \mathsf{L}_x^{-\mathrm{T}} \mathsf{A}^{\mathrm{T}} \Gamma_e^{-1} \delta \boldsymbol{y} \\ \eta_* &= \left(\mathsf{A}^{\mathrm{T}} \Gamma_e^{-1} \mathsf{A} + \alpha \Gamma_x^{-1} \right)^{-1} \mathsf{A}^{\mathrm{T}} \Gamma_e^{-1} \delta \boldsymbol{y}. \end{aligned} \tag{7.15}$$

7.1.2. Newton methods (underdetermined). A Newton solution for an underdetermined system ($M < N$) is written

$$\tilde{\boldsymbol{x}}_* = \tilde{\mathsf{A}}^{\mathrm{T}} (\tilde{\mathsf{A}} \tilde{\mathsf{A}}^{\mathrm{T}} + \alpha \mathsf{I})^{-1} \tilde{\boldsymbol{y}} \tag{7.16}$$

$$\begin{aligned} \mathsf{L}_x \eta_* &= \mathsf{L}_x^{-\mathrm{T}} \mathsf{A}^{\mathrm{T}} \mathsf{L}_e^{\mathrm{T}} \left(\mathsf{L}_e \mathsf{A} \Gamma_x \mathsf{A}^{\mathrm{T}} \mathsf{L}_e^{\mathrm{T}} + \alpha \mathsf{I} \right)^{-1} \mathsf{L}_e \delta \boldsymbol{y} \\ \eta_* &= \Gamma_x \mathsf{A}^{\mathrm{T}} \left(\mathsf{A} \Gamma_x \mathsf{A}^{\mathrm{T}} + \alpha \Gamma_e \right)^{-1} \delta \boldsymbol{y}. \end{aligned} \tag{7.17}$$

7.1.3. Landweber method and steepest descent. The steepest descent method is an iterative reconstruction scheme. Beginning with an arbitrary initial estimate $\tilde{\boldsymbol{x}}^0$ (usually all zeros), we iterate solutions using

$$\tilde{\boldsymbol{x}}_{k+1} = \tilde{\boldsymbol{x}}_k + \tau_k \tilde{\boldsymbol{s}}_k, \tag{7.18}$$

where the steepest direction for (7.13) is given by

$$\tilde{\boldsymbol{s}}_k = \tilde{\mathsf{A}}^{\mathrm{T}} (\tilde{\boldsymbol{y}} - \tilde{\mathsf{A}} \tilde{\boldsymbol{x}}_k) - \alpha \tilde{\boldsymbol{x}}_k, \tag{7.19}$$

and

$$\tau_k = \frac{\|\tilde{\boldsymbol{s}}_k\|^2}{\|\tilde{\mathsf{A}} \tilde{\boldsymbol{s}}_k\|^2 + \alpha \|\boldsymbol{s}_k\|^2}, \tag{7.20}$$

is the step length in direction $\tilde{\boldsymbol{s}}^{(k)}$ that minimizes the one-dimensional error function

$$\tilde{\mathcal{E}}(\tau) := \tilde{\mathcal{E}}_\alpha(\tilde{\boldsymbol{y}}, \tilde{\mathsf{A}}(\tilde{\boldsymbol{x}} + \tau\tilde{\boldsymbol{s}}_k)). \tag{7.21}$$

The implication is that the vectors in the sequence $\{\tilde{\boldsymbol{s}}_1, \tilde{\boldsymbol{s}}_2, \ldots\}$ have the properties [125] that they can be obtained iteratively

$$\tilde{\boldsymbol{s}}_{k+1} = \tilde{\boldsymbol{s}}_k - \tau_k \tilde{\mathsf{A}}^{\mathrm{T}}\tilde{\mathsf{A}}\boldsymbol{s}_k, \tag{7.22}$$

and satisfy an orthogonality condition

$$(\tilde{\boldsymbol{s}}_{k+1}, \tilde{\boldsymbol{s}}_k) = 0. \tag{7.23}$$

The Landweber method replaces (7.18) with

$$\tilde{\boldsymbol{x}}_{k+1} = \tilde{\boldsymbol{x}}_k + \tau\tilde{\mathsf{A}}^{\mathrm{T}}(\tilde{\boldsymbol{y}} - \tilde{\mathsf{A}}\tilde{\boldsymbol{x}}_k), \tag{7.24}$$

where τ is a relaxation parameter. Rather than take the exact step that would move the solution to the minimum of the error function in the direction $\tilde{\boldsymbol{s}}_k$, the Landweber method takes a fixed step. Thus, its convergence is slower than the steepest descent method.

Considering the form of (7.24) we have

$$\eta^{(k+1)} = \eta^{(k)} + \tau\Gamma_x\mathsf{A}^{\mathrm{T}}\Gamma_e^{-1}(\delta\boldsymbol{y} - \mathsf{A}\eta^{(k)}). \tag{7.25}$$

7.1.4. Krylov methods. Krylov methods are the method of choice for large-scale linear ill-posed problems due to their ability to reach a solution in a subspace of relatively low dimension. Consider the general, overdetermined quasi-Newton scheme,

$$\eta = \mathsf{H}_x^{-1}\mathsf{A}^{\mathrm{T}}\delta\boldsymbol{y}, \tag{7.26}$$

or the underdetermined scheme

$$\eta = \mathsf{A}^{\mathrm{T}}\mathsf{H}_e^{-1}\delta\boldsymbol{y}, \tag{7.27}$$

where $\mathsf{H}_x \in \mathbb{R}^N \times \mathbb{R}^N$, $\mathsf{H}_e \in \mathbb{R}^M \times \mathbb{R}^M$ are symmetric positive definite. Then the Krylov spaces may be defined as the spaces spanned by the set of vectors

$$\mathrm{K}_x := \mathrm{span}\left\{\mathsf{A}^{\mathrm{T}}\delta\boldsymbol{y}, \mathsf{H}_x\mathsf{A}^{\mathrm{T}}\delta\boldsymbol{y} \ldots \mathsf{H}_x^j\mathsf{A}^{\mathrm{T}}\delta\boldsymbol{y} \ldots\right\}, \tag{7.28}$$

$$\mathrm{K}_e := \mathrm{span}\left\{\mathsf{A}^{\mathrm{T}}\delta\boldsymbol{y}, \mathsf{A}^{\mathrm{T}}\mathsf{H}_e\delta\boldsymbol{y} \ldots \mathsf{A}^{\mathrm{T}}\mathsf{H}_e^j\delta\boldsymbol{y}, \ldots\right\}. \tag{7.29}$$

Deriving these Krylov sequences for the problems (7.14), or (7.16) both give rise to the same set

$$\begin{aligned}\tilde{\mathbb{K}}_\alpha &= \left\{\tilde{\mathsf{A}}^{\mathrm{T}}\tilde{\boldsymbol{y}}, \tilde{\mathsf{A}}^{\mathrm{T}}\tilde{\mathsf{A}}\tilde{\mathsf{A}}^{\mathrm{T}}\tilde{\boldsymbol{y}} + \alpha\tilde{\mathsf{A}}^{\mathrm{T}}\tilde{\boldsymbol{y}}, \ldots, \sum_{j'=0}^{j}\alpha^{j'}(\tilde{\mathsf{A}}^{\mathrm{T}}\tilde{\mathsf{A}})^{j-j'}\tilde{\mathsf{A}}^{\mathrm{T}}\tilde{\boldsymbol{y}} \ldots\right\} \\ &=: \left\{\tilde{\boldsymbol{v}}_\alpha^{(0)}, \tilde{\boldsymbol{v}}_\alpha^{(1)}, \ldots, \tilde{\boldsymbol{v}}_\alpha^{(j)} \ldots\right\}.\end{aligned} \tag{7.30}$$

Defining the unregularized Krylov space of dimension J, via the basis set

$$\tilde{\mathbb{K}} \equiv \{\tilde{\boldsymbol{v}}^{(j)}\} := \{(\tilde{\mathsf{A}}^{\mathrm{T}}\tilde{\mathsf{A}})^j\tilde{\mathsf{A}}^{\mathrm{T}}\tilde{\boldsymbol{y}}, \qquad j = 0 \ldots J-1\}, \tag{7.31}$$

we may derive the Krylov space for any regularization parameter α via the basis set

$$\tilde{\mathbb{K}}_\alpha = \begin{cases}\tilde{\boldsymbol{v}}_\alpha^{(0)} = \tilde{\boldsymbol{v}}^{(0)} \\ \tilde{\boldsymbol{v}}_\alpha^{(1)} = \tilde{\boldsymbol{v}}^{(1)} + \alpha\tilde{\boldsymbol{v}}_\alpha^{(0)} \\ \vdots \\ \tilde{\boldsymbol{v}}_\alpha^{(j)} = \tilde{\boldsymbol{v}}^{(j)} + \alpha\tilde{\boldsymbol{v}}_\alpha^{(j-1)}.\end{cases} \tag{7.32}$$

Finally we may construct the Krylov space for the original parameters:

$$\mathbb{K}_\alpha = \begin{cases} \boldsymbol{v}_\alpha^{(0)} = \Gamma_x \mathsf{A}^{\mathrm{T}} \Gamma_e^{-1} \delta \boldsymbol{y} \\ \boldsymbol{v}_\alpha^{(1)} = \left(\Gamma_x \mathsf{A}^{\mathrm{T}} \Gamma_e^{-1} \mathsf{A} + \alpha \right) \boldsymbol{v}_\alpha^{(0)} \\ \vdots \\ \boldsymbol{v}_\alpha^{(j)} = \left(\Gamma_x \mathsf{A}^{\mathrm{T}} \Gamma_e^{-1} \mathsf{A} + \alpha \right) \boldsymbol{v}_\alpha^{(j-1)}. \end{cases} \tag{7.33}$$

Contrast this with the basis defined from (7.15)

$$\mathbb{K}_\alpha^{(\mathrm{ON})} = \begin{cases} \boldsymbol{v}_\alpha^{(0)} = \mathsf{A}^{\mathrm{T}} \Gamma_e^{-1} \delta \boldsymbol{y} \\ \boldsymbol{v}_\alpha^{(1)} = \mathsf{A}^{\mathrm{T}} \Gamma_e^{-1} \mathsf{A} \mathsf{A}^{\mathrm{T}} \Gamma_e^{-1} \delta \boldsymbol{y} + \alpha \Gamma_x^{-1} \mathsf{A}^{\mathrm{T}} \Gamma_e^{-1} \delta \boldsymbol{y} \\ \vdots \\ \boldsymbol{v}_\alpha^{(j)} = \left(\mathsf{A}^{\mathrm{T}} \Gamma_e^{-1} \mathsf{A} + \alpha \Gamma_x^{-1} \right)^{j-1} \mathsf{A}^{\mathrm{T}} \Gamma_e^{-1} \delta \boldsymbol{y} \end{cases} \tag{7.34}$$

or that from (7.17)

$$\mathbb{K}_\alpha^{(\mathrm{UN})} = \begin{cases} \boldsymbol{v}_\alpha^{(0)} = \Gamma_x \mathsf{A}^{\mathrm{T}} \delta \boldsymbol{y} \\ \boldsymbol{v}_\alpha^{(1)} = \Gamma_x \mathsf{A}^{\mathrm{T}} \mathsf{A} \Gamma_x \mathsf{A}^{\mathrm{T}} \delta \boldsymbol{y} + \alpha \Gamma_x \mathsf{A}^{\mathrm{T}} \Gamma_e \delta \boldsymbol{y} \\ \vdots \\ \boldsymbol{v}_\alpha^{(j)} = \Gamma_x \mathsf{A}^{\mathrm{T}} \sum_{j'}^{j} \alpha^{j'} \left(\mathsf{A} \Gamma_x \mathsf{A}^{\mathrm{T}} \right)^{j-j'} \Gamma_e^{j'} \delta \boldsymbol{y}. \end{cases} \tag{7.35}$$

In the conjugate gradient method a second sequence of $\tilde{\mathsf{A}}$-conjugate directions $\{\tilde{\boldsymbol{p}}_1, \tilde{\boldsymbol{p}}_2, \ldots\}$ is constructed and (7.19) becomes

$$\tilde{\boldsymbol{x}}_{k+1} = \tilde{\boldsymbol{x}}_k + \tau_k \tilde{\boldsymbol{p}}_k. \tag{7.36}$$

The direction $\tilde{\boldsymbol{p}}_k$ is given by

$$\tilde{\boldsymbol{p}}_{k+1} = \tilde{\boldsymbol{s}}_{k+1} + \beta_k \tilde{\boldsymbol{p}}_k, \tag{7.37}$$

where the Gram–Schmidt constants β_k are given by one of the methods

$$\beta_k = \begin{cases} \text{Fletcher–Reeves:} & \frac{\|\tilde{\boldsymbol{s}}_{k+1}\|^2}{\|\tilde{\boldsymbol{s}}_k\|^2} \\ \text{Polak–Ribiere:} & \frac{\langle \tilde{\boldsymbol{s}}_{k+1}, \tilde{\boldsymbol{s}}_{k+1} - \tilde{\boldsymbol{s}}_k \rangle}{\langle \tilde{\boldsymbol{p}}_{k+1}, \tilde{\boldsymbol{s}}_{k+1} - \tilde{\boldsymbol{s}}_k \rangle} \end{cases} \tag{7.38}$$

and the step length (cf (7.20)) is given by

$$\tau_k = \frac{\|\boldsymbol{s}_k\|^2}{\|\tilde{\mathsf{A}} \tilde{\boldsymbol{p}}_k\|^2 + \alpha \|\tilde{\boldsymbol{p}}_k\|^2}. \tag{7.39}$$

We note that given a Krylov vector (an image in parameter space) the next vector is formed by the steps

$$\text{forward project} \rightarrow \text{filter in data space} \rightarrow \text{back project} \rightarrow \text{filter in image space} \tag{7.40}$$

Thus, the sequence can be formed without building the matrix A and constitutes a *matrix free* approach.

7.1.5. Row and column normalization. Let us define $\boldsymbol{a}_i^R \in \mathbb{R}^N$ as the ith row of A and $\boldsymbol{a}_j^C \in \mathbb{R}^M$ as the jth column. Then we can define a number of purely algebraic conditioning matrices

$$\begin{aligned} \mathsf{R}_1 &= \operatorname{diag}\left[\left\{ \|\boldsymbol{a}_i^R\|_1; i = 1 \ldots M \right\} \right] \\ \mathsf{C}_1 &= \operatorname{diag}\left[\left\{ \|\boldsymbol{a}_j^C\|_1; j = 1 \ldots N \right\} \right] \end{aligned} \tag{7.41}$$

$$\begin{aligned}\mathsf{R}_2^2 &= \operatorname{diag}\left[\left\{\left\|\boldsymbol{a}_i^R\right\|_2^2;\, i = 1 \ldots M\right\}\right] = \operatorname{diag}[\mathsf{A}\mathsf{A}^{\mathrm{T}}]\\ \mathsf{C}_2^2 &= \operatorname{diag}\left[\left\{\left\|\boldsymbol{a}_j^C\right\|_2^2;\, j = 1 \ldots N\right\}\right] = \operatorname{diag}[\mathsf{A}^{\mathrm{T}}\mathsf{A}],\end{aligned} \tag{7.42}$$

where the notation $\|\cdot\|_p := \left(\sum .^p\right)^{1/p}$ defines the p-norm of a vector.

For the special case where A contains purely nonnegative elements the forms R_1, C_1 can be constructed as row and columns sums:

$$\mathsf{R}_1 = \mathsf{A}\mathbf{1}, \qquad \mathsf{C}_1 = \mathsf{A}^{\mathrm{T}}\mathbf{1}. \tag{7.43}$$

The choice

$$\Gamma_e \leftarrow \mathsf{R}_1 \qquad \Gamma_x^{-1} \leftarrow \mathsf{C}_1 \tag{7.44}$$

leads to the *simultaneous algebraic reconstruction technique* (SART)

$$\eta^{(k+1)} = \eta^{(k)} + \tau \mathsf{C}_1^{-1}\mathsf{A}^{\mathrm{T}}\mathsf{R}_1^{-1}(\delta\boldsymbol{y} - \mathsf{A}\eta^{(k)}). \tag{7.45}$$

This is suitable for emission tomography and also for fluorescence optical tomography, where the sought for parameters are positive and the matrix A represents a probability of a photon emitted at voxel i arriving at detector j.

The choice

$$\Gamma_e \leftarrow \mathsf{R}_2^2 \qquad \Gamma_x \leftarrow \mathsf{I} \tag{7.46}$$

leads to the *simultaneous iterative reconstruction technique* (SIRT)

$$\eta^{(k+1)} = \eta^{(k)} + \tau \mathsf{A}^{\mathrm{T}}\mathsf{R}_2^{-2}(\delta\boldsymbol{y} - \mathsf{A}\eta^{(k)}). \tag{7.47}$$

A particular preconditioner that is used in conjugate gradient solvers is given by [126]

$$\Gamma_x^{-1} = \mathsf{M} = \operatorname{diag}\left[\mathsf{A}^{\mathrm{T}}\Gamma_e^{-1}\mathsf{A}\right], \tag{7.48}$$

which seeks to equalize the curvature of the objective function along the coordinate axes ('sphering').

The SIRT and SART algorithms derived from the Landweber method can also be thought of as derived from the algebraic reconstruction technique (ART), or Kaczmarz method [127]. This is a row-action method that computes an update to the solution by processing one row of the linearized system at a time. The inner loop generates an update from row i by

$$\eta^{(k+1)} = \eta^{(k)} + \varpi_k \frac{\delta\boldsymbol{y}_i - \left\langle \boldsymbol{a}_i^R, \eta^{(k)}\right\rangle}{\left\|\boldsymbol{a}_i^R\right\|^2}\boldsymbol{a}_i^R, \tag{7.49}$$

where ϖ_k is a relaxation parameter. The loop over rows of A is repeated n times. The rows i are processed in randomized order, which has been shown to improve the convergence rate of the method [128].

7.2. Multiplicative methods

Consider the substitution

$$y = \mathrm{e}^{\zeta} \leftrightarrow \zeta = \log y. \tag{7.50}$$

An optimization problem in terms of (7.50) may be defined

$$x_* = \arg\min_x\left[\tfrac{1}{2}\|\log y - \log \mathcal{F}(x)\|_{\Gamma_e^{-1}}^2 + \alpha\Psi(x)\right]. \tag{7.51}$$

Taking the linearization of this as in (7.7) leads to a problem

$$\eta_* = \arg\min_\eta\left[\frac{1}{2}\|\log y - \log \mathcal{F}(x_0) - \frac{\mathcal{F}'(x_0)}{\mathcal{F}(x_0)}\eta\|_{\Gamma_e^{-1}}^2 + \alpha\Psi(x)\right] \tag{7.52}$$

$$= \arg\min_\eta\left[\frac{1}{2}\|\zeta^\delta - \hat{A}\eta\|_{\Gamma_e^{-1}}^2 + \alpha\Psi(x)\right] \tag{7.53}$$

where $\zeta^\delta = \log y - \log \mathcal{F}(x_0)$ and the linear operator

$$\hat{A} = \frac{\mathcal{F}'(x_0)}{y_0}, \tag{7.54}$$

with $y_0 = \mathcal{F}(x_0)$. Similarly, considering

$$x = \mathrm{e}^\xi \leftrightarrow \xi = \log x, \tag{7.55}$$

leads to a linearized problem

$$\xi_*^\delta = \arg\min_{\xi^\delta} \left[\frac{1}{2} \| \log y - \log \mathcal{F}(x_0) - \frac{\mathcal{F}'(x_0)x_0}{\mathcal{F}(x_0)} \xi^\delta \|^2_{\Gamma_e^{-1}} + \alpha \Psi(x) \right] \tag{7.56}$$

$$= \arg\min_{\xi^\delta} \left[\frac{1}{2} \| \zeta^\delta - \check{A} \xi^\delta \|^2_{\Gamma_e^{-1}} + \alpha \Psi(x) \right], \tag{7.57}$$

with

$$\check{A} = \frac{\mathcal{F}'(x_0)x_0}{y_0}. \tag{7.58}$$

Taking the discrete version we arrive at

$$\tilde{\mathsf{A}} = \mathrm{diag}[1/\boldsymbol{y}_0] \mathsf{L}_e \, \mathsf{A} \, \mathsf{L}_x^{-1} \, \mathrm{diag}[\boldsymbol{x}_0], \tag{7.59}$$

which means that the covariances have been transformed:

$$\Gamma_e \to \mathrm{diag}[\boldsymbol{y}_0] \Gamma_e \, \mathrm{diag}[\boldsymbol{y}_0], \qquad \Gamma_x \to \mathrm{diag}[\boldsymbol{x}_0] \Gamma_x \, \mathrm{diag}[\boldsymbol{x}_0]. \tag{7.60}$$

The steepest descent scheme for (7.57) is given by

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k \odot \exp \left[\tilde{\mathsf{A}}^{\mathrm{T}} \log \left(\frac{\boldsymbol{y}}{\mathcal{F}(\boldsymbol{x}_k)} \right) - \alpha \Psi'(\boldsymbol{x}_k) \right], \tag{7.61}$$

where the notation $\odot$ refers to pointwise multiplication.

When the mapping $\mathcal{F}$ is linear (as in fluorescence DOT for example) and when regularization is only taken implicitly, we arrive at the *multiplicative ART* (MART) scheme

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k \odot \exp \left[\mathsf{A}^{\mathrm{T}} \log \left(\frac{\delta \boldsymbol{y}}{\mathsf{A} \boldsymbol{x}_k} \right) \right]. \tag{7.62}$$

If we start instead from the KL divergence

$$KL(\boldsymbol{y}, \mathcal{F}(\boldsymbol{x})) = \sum \boldsymbol{y} \log \left[\frac{\boldsymbol{y}}{\mathcal{F}(\boldsymbol{x})} \right] - \boldsymbol{y} + \mathcal{F}(\boldsymbol{x}), \tag{7.63}$$

we get to the *maximum likelihood expectation maximization* (MLEM) algorithm

$$\boldsymbol{x}_{k+1} = \frac{\boldsymbol{x}_k}{\mathsf{A}^{\mathrm{T}} \mathbf{1}} \odot \left[\mathsf{A}^{\mathrm{T}} \left(\frac{\boldsymbol{y}}{\mathcal{F}(\boldsymbol{x}_k)} \right) \right]. \tag{7.64}$$

In this case the explicit regularization leads to the MAP-EM algorithm

$$\boldsymbol{x}_{k+1} = \frac{\boldsymbol{x}_k}{\mathsf{A}^{\mathrm{T}} \mathbf{1} + \alpha \Psi'(\boldsymbol{x}_k)} \odot \left[\mathsf{A}^{\mathrm{T}} \left(\frac{\boldsymbol{y}}{\mathcal{F}(\boldsymbol{x}_k)} \right) \right]. \tag{7.65}$$

Alternatively a two step method reconstruction followed by filtering can be employed

$$\boldsymbol{x}_{k+\frac{1}{2}} = \frac{\boldsymbol{x}_k}{\mathsf{A}^{\mathrm{T}} \mathbf{1}} \odot \left[\mathsf{A}^{\mathrm{T}} \left(\frac{\boldsymbol{y}}{\mathcal{F}(\boldsymbol{x}_k)} \right) \right], \tag{7.66}$$

$$\boldsymbol{x}_{k+1} = \mathcal{P} \boldsymbol{x}_{k+\frac{1}{2}}, \tag{7.67}$$

where $\mathcal{P}$ is a projection such as positivity, scale–space filtering or frequency band pass.

7.3. Nonlinear methods

7.3.1. Gauss–Newton method. The Gauss–Newton method can be considered as the iterative minimization of the quadratic Taylor-series approximation to $\mathcal{E}(y, \mathcal{F}(x))$

$$\mathcal{E}(\delta \boldsymbol{y}, \mathcal{F}(\boldsymbol{x}_k + \eta)) \simeq \tfrac{1}{2}\|\boldsymbol{y} - \mathcal{F}(\boldsymbol{x}_k) - \mathcal{F}'(\boldsymbol{x}_k)\eta\|^2_{\Gamma_e^{-1}} + \alpha\left(\Psi(\boldsymbol{x}_k) + 2\langle \eta, \mathcal{W}\boldsymbol{x}_k\rangle + \langle \eta, \mathcal{W}\eta\rangle\right), \tag{7.68}$$

where $\mathcal{W}(\boldsymbol{x}_k) : \mathrm{X} \to \mathrm{X}$ represents the mapping induced by the linearization of the functional Ψ. Minimization of (7.68) is given by the solution to

$$\left(\mathsf{A}^{\mathrm{T}}\Gamma_e^{-1}\mathsf{A} + \alpha\mathcal{W}\right)\eta = \mathsf{A}^{\mathrm{T}}\Gamma_e^{-1}\left(\boldsymbol{y} - \mathcal{F}(\boldsymbol{x}_k)\right) - \mathcal{W}\boldsymbol{x}_k, \tag{7.69}$$

$$\mathsf{H}\eta = -\boldsymbol{g}, \tag{7.70}$$

with Hessian $\mathsf{H} = \left(\mathsf{A}^{\mathrm{T}}\Gamma_e^{-1}\mathsf{A} + \alpha\mathcal{W}\right)$ and gradient $\boldsymbol{g} = \mathcal{W}\boldsymbol{x}_k - \mathsf{A}^{\mathrm{T}}\Gamma_e^{-1}\left(\boldsymbol{y} - \mathcal{F}(\boldsymbol{x}_k)\right)$. From the form of the Hessian we can see that it is guaranteed to be symmetric non-negative definite provided that Ψ is convex, and therefore η is guaranteed to be in a descent direction for $\mathcal{E}$. Solution of (7.70) can be carried out with any of the methods in section 7.1. In particular the *Gauss–Newton–Krylov* method uses a Krylov solver for (7.70). For a badly ill-posed problem the size of the requisite Krylov space can be quite small, and may be constructed entirely through forward and adjoint solutions and image and data filtering operations as in (7.40), and therefore also as a matrix-free approach.

Since the approximation in (7.68) is only locally quadratic, the update given by solving (7.70) may (i) not be optimal or may (ii) not be a descent step (if the Hessian is not symmetric positive definite). There are two strategies for resolving these problems. The *Damped Gauss–Newton* method is a globalization strategy that addresses the first problem. In this approach the update direction is used in a one-dimensional line search to find an update step τ_k that minimizes $\mathcal{E}$ along this direction:

$$\tau_k = \arg\min_\tau \mathcal{E}(\boldsymbol{y}, \mathcal{F}(\boldsymbol{x}_k + \tau\eta)). \tag{7.71}$$

Note that the full nonlinear mapping $\mathcal{F}$ is used in (7.71), not its linearization. The *Levenberg–Marquardt* method can also address the second problem. In this approach a control parameter γ is used to modify the Hessian

$$\mathsf{H} \to \mathsf{H} + \gamma\mathsf{I}.$$

When γ is large the update η tends towards the steepest descent direction with increasingly shorter steps. When γ is small the update tends towards the Newton direction. The idea in the algorithm is to decrease γ whenever the preceding step reduced $\mathcal{E}$ but to increase it and re-solve (7.68) if the preceding step increased $\mathcal{E}$. The role of γ also serves to modify the eigenspectrum of H and force it to be symmetric positive definite. The role of column normalization of A (see section 7.1.5) is often crucial in the use of the Levenberg–Marquardt algorithm since it can 'sphere' the level sets of the local quadratic approximation. Typically the Levenberg–Marquardt method is required for highly nonlinear problems, which is not the case in optical tomography. A comparison of Levenberg–Marquardt and damped Gauss–Newton methods can be found in [129]. The Gauss–Newton method was used for the inverse RTE problem in optical tomography in [130].

7.3.2. Nonlinear conjugate gradient method (NCG). The algorithm is effectively the one in section 7.1.4 with line search as in (7.71) replacing the calculation of the step length give

in (7.39). Rather than present it in the canonical variables we can present it in the normal variables (Fletcher–Reeves version)

$$
\begin{aligned}
&\text{Set } \boldsymbol{g}_0 = \alpha\Psi'(\boldsymbol{x}_0) - \mathcal{F}'^{*}\Gamma_e^{-1}(\boldsymbol{y} - \mathcal{F}(\boldsymbol{x}_0))\\
&\boldsymbol{p}_0 = -\Gamma_x \boldsymbol{g}_0\\
&\text{for } k = 1, \ldots : \text{ do}\\
&\quad \tau_k = \arg\min_\tau \mathcal{E}(\boldsymbol{x}_{k-1} + \tau \boldsymbol{p}_{k-1})\\
&\quad \boldsymbol{x}_k = \boldsymbol{x}_{k-1} + \tau_k \boldsymbol{p}_{k-1}\\
&\quad \boldsymbol{g}_k = \alpha\Psi'(\boldsymbol{x}_k) - \mathcal{F}'^{*}\Gamma_e^{-1}(\boldsymbol{y} - \mathcal{F}(\boldsymbol{x}_k))\\
&\quad \beta_k = \frac{\langle \boldsymbol{g}_k, \Gamma_x \boldsymbol{g}_k\rangle}{\langle \boldsymbol{g}_{k-1}, \Gamma_x \boldsymbol{g}_{k-1}\rangle}\\
&\quad \boldsymbol{p}_k = -\Gamma_x \boldsymbol{g}_k + \beta_k \boldsymbol{p}_{k-1}\\
&\text{end for}
\end{aligned}
$$

Note that $\tilde{\boldsymbol{s}}_k = -\Gamma_x \boldsymbol{g}_k$ represents the application of the preconditioner $\boldsymbol{M}^{-1} = \Gamma_x$ to the gradient. If we assume $\Psi'(\boldsymbol{x}) \equiv \mathcal{W}\boldsymbol{x} = \mathsf{L}_x^{\mathrm{T}}\mathsf{L}_x\boldsymbol{x} = \Gamma_x^{-1}\boldsymbol{x}$ then this conditioned gradient will be

$$\tilde{\tilde{\boldsymbol{s}}}_k = \Gamma_x \mathcal{F}'^{*}\Gamma_e^{-1}\,(\boldsymbol{y} - \mathcal{F}(\boldsymbol{x}_k)) - \alpha \boldsymbol{x}_k.$$

Note the similarity to the general scheme in (7.40). Also note that the variables $\tilde{\tilde{\boldsymbol{s}}}$ are no longer dimensionless. They have the reciprocal dimensions of the original parameters themselves.

Nonlinear conjugate gradients is usually restarted after a certain number of iterations. As for any descent method, NCG may be combined with projection onto convex sets for enforcement of constraints such as lower and upper bounds, but at the expense of loss of its conjugacy properties. Nonlinear CG was used in optical tomography for the diffusion-based problem in [131] and for the RTE-based problem in [132]

7.3.3. Limited-memory BFGS method. The BFGS (Broyden–Fletcher–Goldfarb–Shanno) algorithm is a quasi-Newton approach that builds up estimates $\tilde{\mathsf{H}}^{-1}$ of the inverse of the Hessian matrix H [133, 134]. The update rule

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - \lambda_k \tilde{\mathsf{H}}_k^{-1} \boldsymbol{g}_k \tag{7.72}$$

is employed, where $\tilde{\mathsf{H}}_k$ is updated at each iteration by the formula

$$\tilde{\mathsf{H}}_{k+1}^{-1} = \mathsf{V}_k^{\mathrm{T}} \tilde{\mathsf{H}}_k^{-1} \mathsf{V}_k + \rho_k \boldsymbol{d}_k \boldsymbol{d}_k^{\mathrm{T}}, \tag{7.73}$$

$$\rho_k = 1/\langle \boldsymbol{z}_k, \boldsymbol{d}_k\rangle, \tag{7.74}$$

$$\mathsf{V}_k = \mathsf{I} - \rho_k \boldsymbol{z}_k \boldsymbol{d}_k^{\mathrm{T}}, \tag{7.75}$$

$$\boldsymbol{d}_k = \boldsymbol{x}_{k+1} - \boldsymbol{x}_k, \tag{7.76}$$

$$\boldsymbol{z}_k = \boldsymbol{g}_{k+1} - \boldsymbol{g}_k. \tag{7.77}$$

In the limited memory version of the algorithm (L-BFGS), the approximate matrices $\tilde{\mathsf{H}}_k^{-1}$ are not stored explicitly, but described implicitly by a limited number of pairs of vectors $\{\boldsymbol{d}_i, \boldsymbol{z}_i\}$, where in each iteration, a new vector pair is added, and the oldest pair is discarded.

7.3.4. Nonlinear Kaczmarz method. The nonlinear Kaczmarz method is widely used in nonlinear tomography [135]. It is applicable for problems such as DOT which use multiple sources. Using one source at a time, the subset of the data from this source is back-projected

and added to the solution

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \mathcal{C}\mathcal{F}'^{*}_{i(k)}\Gamma_e^{-1}(\boldsymbol{y}_{i(k)} - \mathcal{F}_{i(k)}(\boldsymbol{x}_k)) \tag{7.78}$$

where the index $i(k)$ refers to the subset of the data accessed on the kth iteration of the algorithm. The operator $\mathcal{C}$ is a simple operator playing the role of a right-preconditioner. A natural choice would be

$$\mathcal{C} = \left(\mathcal{F}_{i(k)}\mathcal{F}'^{*}_{i(k)} + \alpha \mathsf{I}\right)^{-1}, \tag{7.79}$$

but this may be hard to compute. Applications in optical tomography can be found in [54, 136].

7.3.5. Iterative coordinate descent. A relatively simple approach to generating step directions is to update one pixel at a time. It is equivalent to taking a descent direction $\boldsymbol{s}$, projecting to one unit axis, and iterating through the dimensions of X. This is effectively a Gibbs sampling procedure for the posterior distribution within the Bayesian framework.

Writing $\hat{\boldsymbol{e}}_k$ for a unit vector with 1 in the kth entry and 0 otherwise, the ICD method considers

$$\tau_k = \arg\min_k \left[\mathcal{E}(\boldsymbol{x}_k + \tau\hat{\boldsymbol{e}}_k)\right], \tag{7.80}$$

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \tau_k\hat{\boldsymbol{e}}_k. \tag{7.81}$$

Taking a local linearization around $\boldsymbol{x}_k$ the one-dimensional objective functional to be minimized is

$$\begin{aligned}\mathcal{E}(\boldsymbol{x}_k + \tau\hat{\boldsymbol{e}}_k) = {} & \tfrac{1}{2}\|\boldsymbol{y} - \mathcal{F}(\boldsymbol{x}_k)\|^2_{\Gamma_e^{-1}} - \tau\left\langle \boldsymbol{y} - \mathcal{F}(\boldsymbol{x}_k), \Gamma_e^{-1}\boldsymbol{a}_k^C\right\rangle \\ & + \tfrac{1}{2}\tau^2\left\|\boldsymbol{a}_k^C\right\|^2_{\Gamma_e^{-1}} + \alpha\Psi(\boldsymbol{x}_k + \tau\hat{\boldsymbol{e}}_k).\end{aligned} \tag{7.82}$$

If the prior is Gaussian then (7.82) is a weighted least-squares problem with minimum

$$\tau_k = \frac{\left\langle \boldsymbol{y} - \mathcal{F}(\boldsymbol{x}_k), \Gamma_e^{-1}\boldsymbol{a}_k^C\right\rangle - \left\langle \hat{\boldsymbol{e}}_k, \Gamma_x^{-1}\boldsymbol{x}_k\right\rangle}{\left\|\boldsymbol{a}_k^C\right\|^2_{\Gamma_e^{-1}} + \left\langle \hat{\boldsymbol{e}}_k, \Gamma_x^{-1}\hat{\boldsymbol{e}}_k\right\rangle}, \tag{7.83}$$

but it is relatively simple also for a non-Gaussian prior since evaluation of the prior usually only involves neighbouring pixels. Projection onto constraint sets is also efficient since once the constraints are imposed for one pixel it is not revisited. However, the evaluation of the likelihood is only efficient for an explicit matrix method and not for the matrix-free approach.

The ICD method was used in PET by Fessler [137] and in optical tomography by Bouman and Webb [138, 139]. Acceleration was achieved by a multiresolution strategy in [140].

7.3.6. PDE-constrained method. The PDE-constrained method considers the approach presented in (7.7). To implement the method we consider Lagrangian (dual) fields Z and define an objective function

$$\mathcal{J}_i(x, U_i, Z_i) = \|y_i - \mathcal{M}U_i\|^2_{\Gamma_e^{-1}} + \alpha\Psi(x) + \langle Z_i, (\mathcal{L}(x)U_i - q_i)\rangle_\Omega, \tag{7.84}$$

where q_i represents an equivalent source for the boundary condition (2.2). When considering all sources the full Lagrangian is

$$\mathcal{J}(x, U, Z) = \sum_i \mathcal{J}_i(x, U_i, Z_i). \tag{7.85}$$

The minimum of (7.84) occurs where the first variation

$$\mathcal{J}_{i,x} = \alpha\Psi'(x) + \langle Z_i, \mathcal{L}_x U_i\rangle_\Omega \tag{7.86}$$

$$\mathcal{J}_{i,U} = \mathcal{L}^* Z_i - \mathcal{M}^*\Gamma_e^{-1}(y_i - \mathcal{M}U_i) \tag{7.87}$$

$$\mathcal{J}_{i,Z} = \mathcal{L}U_i - q_i \tag{7.88}$$

becomes zero. We recognize (7.87) as the equation for the backprojected field of the residual difference between the data y_i and the measurement of the direct field U_i. We recognize $\mathcal{L}_x = \mathcal{V}(\eta)$ as the potential operator introduced in (2.18). Therefore, if (7.88) and (7.88) are both equated to zero, (7.86) is the gradient of the negative log posterior of a MAP estimation scheme. If instead $\{-\mathcal{J}_{i,x}, -\mathcal{J}_{i,U}, -\mathcal{J}_{i,Z}\}$ is used as an update direction for $\{x, U, Z\}$, then we may solve for all variables simultaneously without requiring convergence for any one variable until termination.

Taking the second variation of $\mathcal{J}$ results to a Newton system

$$\begin{pmatrix} \alpha\Psi''(x) & \mathcal{L}_x^* Z_i & \mathcal{L}_x U_i \\ \mathcal{L}_x^* Z_i & \mathcal{M}^*\Gamma_e^{-1}\mathcal{M} & \mathcal{L} \\ \mathcal{L}_x U_i & \mathcal{L} & 0 \end{pmatrix} \begin{pmatrix} x^\delta \\ U_i^\delta \\ Z_i^\delta \end{pmatrix} = - \begin{pmatrix} \alpha\Psi'(x) + \langle Z_i, \mathcal{L}_x U_i\rangle_\Omega \\ \mathcal{L}^* Z_i - \mathcal{M}^*\Gamma_e^{-1}(y - \mathcal{M}U_i) \\ \mathcal{L}U_i - q \end{pmatrix}. \tag{7.89}$$

In many applications the system in (7.89) is simplified by taking the Schur complement of the complete Hessian and by using the Gauss–Newton or other quasi-Newton approximations for the solution scheme [57, 141, 142]. For the time-domain problem, even with such Hessian reduction techniques, Newton methods are infeasible but first-order descent schemes can still be used [38].

7.4. Error and prior modelling

So far we did not discuss the forms of the data covariances Γ_e or the prior $\Psi(x)$. The format of noise will usually be predicted on physical grounds. It is often taken to be zero-mean Gaussian noise with a possibly non-white covariance. When considering photon counting measurements the implication is that the noise should be Poisson with variance $\sigma_j^2 = y_j$. Assuming sufficient signal to approach the central limit theorem would lead to an equivalent Gaussian model of additive noise with

$$\Gamma_e = \mathrm{diag}[y] \quad \Leftrightarrow \quad \mathsf{L}_e = \mathrm{diag}\left[\frac{1}{y^{1/2}}\right]. \tag{7.90}$$

A more commonly used model assumes that equal numbers of photons are collected at each detector leading to a constant relative error and the formal covariance structure

$$\Gamma_e = \mathrm{diag}[y^2] \quad \Leftrightarrow \quad \mathsf{L}_e = \mathrm{diag}\left[\frac{1}{y}\right] \tag{7.91}$$

which also corresponds to the error model implicit in using the Rytov series in place of the Born series for the linearized model. Within this photon counting paradigm there is no correlation between errors on different detectors. However, when considering the actual errors between measured and modelled data the correlation is far from being negligible, and their mean is far from being zero. This discrepancy can be understood by formally considering the modelling error as a random variable [143].

$$\pi(y^{\mathrm{meas}} - \mathcal{F}_h(x_{\mathrm{true}})) \sim \mathcal{N}(\overline{\epsilon}, \Gamma_e + \Gamma_\epsilon), \tag{7.92}$$

where $\mathcal{F}_h$ is an approximate model with the accuracy of the computational method employed, $\overline{\epsilon}$ is the model bias representing the discrepancy between this model and the real physics, and

Γ_ϵ is the model error covariance. In practical applications of optical tomography an estimate of the bias is made by measuring a reference problem and comparing it to a reference model

$$\bar{\epsilon} = y^{\text{ref}} - \mathcal{F}_h(x_{\text{ref}}), \tag{7.93}$$

which leads to a corrected model

$$\mathcal{F}_{\text{corrected}}(x) = \mathcal{F}_h(x) + y^{\text{ref}} - \mathcal{F}_h(x_{\text{ref}}). \tag{7.94}$$

Clearly the corrected model and the measured data agree exactly for the value $x = x_{\text{ref}}$ and the assumption is that they will also agree at nearby values. This is questionable.

A more principled approach is the *approximation error method* [143]. In this approach the statistical properties of the modelling error are estimated by sampling over a plausible distribution of solutions and comparing the modeled data from each sample with 'measured' data from this sample. In practice the measured data could be also be modelled but with a much more accurate technique. In [144] this technique was shown to result in reconstructed images using a relatively inaccurate forward model that were of almost equal quality to those using a more accurate forward model; the increase in computational efficiency was an order of magnitude.

Choice of regularization Ψ is critical and difficult to justify unequivocally. Assuming a form

$$\Psi(x) = \int_\Omega \psi\left(|\nabla x|\right) \mathrm{d}\boldsymbol{r}, \tag{7.95}$$

where $\psi : \mathrm{X} \to \mathrm{X}$ is an image to image mapping leads to the linearization

$$\Gamma_x^{-1} = \mathcal{W}(x) = \nabla^{\mathrm{T}} k(\boldsymbol{r}) \nabla, \tag{7.96}$$

where the *diffusivity* is given by

$$k := \frac{\psi'\left(|\nabla x|\right)}{|\nabla x|}. \tag{7.97}$$

Note that the matrix Γ_x is a covariance whose entries represent a correlation between pixels. On the other hand Γ_x^{-1} is sparse, representing the local relationship between neighbours. We could therefore specify Γ_x in a number of ways.

(i) From a database of representative images $\{x_k\}$ so that $\Gamma_x = E[(x - \bar{x})(x - \bar{x})^{\mathrm{T}}]$
(ii) By specifying a Markov random field $\Gamma_x^{-1} \equiv \mathsf{W}$
(iii) By specifying a PDE $\Gamma_x^{-1} \equiv \nabla^{\mathrm{T}} k \nabla$

Methods such as first-order Tikhonov (Phillips–Twomey), total variation (TV), or generalized Gaussian Markov random field (GGMRF) [145] make statistical assumptions about the distribution of edges. They pose assumptions about the regularity (e.g. smoothness) of the solution, the regularity being measured in terms of some norm of the solution and its differentials [146]. In addition the 'lumpy background' prior provides an effective method of estimating information content of different imaging systems [147].

Image processing methods for computer vision offer a large set of techniques for de-noising and segmentation based on anisotropic diffusion processes [148–151]. These techniques formulate directly a PDE for image flow, rather than as the Euler equation of a variational form and can be considered more general. More specific prior information can be incorporated if we assume that some approximate knowledge of the objects such as shape topology and intra-region parameter regions is available. [152]

8. Shape-based methods

Optical tomography as discussed so far in this article is an example of a *parameter identification* problem because the reconstructed images represent the parameters of a model of light propagation. In common with many medical imaging modalities the reconstructed images are not an end in themselves; they need to be analysed for structural and functional information, including classification of regions, segmentation and cross validation with other modalities.

In the particular case of OT, this post-processing step has some drawbacks. In reconstruction, we often use small sets of measurements to reconstruct a large image with many parameters; this is not the case in camera-based measurement systems, yet the size of the relevant measurement domain may still be small in comparison to the solution space. In the subsequent analysis stage we analyse these parameters to categorize the image into only a few discrete categories, e.g. to estimate a classification into discrete regions, or to determine the parameters of a low-dimensional shape model. In each case, the final result is of considerably smaller dimensionality than the intermediate reconstruction. An alternative approach would be to integrate the classification or segmentation with the reconstruction. This problem may be better posed than the two-stage approach, as we are only moving from the sparse measurements to another low-dimensional space.

In many biomedical applications, the parameter distribution, which is sought in the inverse problem, contains some sorts of interfaces. These can for example be boundaries of some tumour or hematoma, or the interfaces between different organs or tissue types, or the boundaries of regions filled with some tracer or marker substance. These interfaces are typically not recovered well in classical reconstruction schemes due to the above-mentioned need for relatively strong regularization tools. Most of these regularization tools penalize variations or gradients in the parameter distribution, which yields oversmoothed reconstructions. Therefore, interfaces and boundaries are blurred and region boundaries cannot easily be detected. However, in many applications it is important to be able to find the boundaries of certain inclusions as accurately as possible. Whereas inside and outside these subregions the parameters might not vary much, often across the interfaces significant jumps occur in the tissue parameters. This motivates the possibility of direct *shape-based* reconstruction methods, which generally can be categorized into *explicit* and *implicit* methods.

Shape reconstruction techniques for optical tomography using the DA as the forward model are presented in [109, 152–160] and using the RTE for the forward model in [161].

8.1. Explicit shape method

In shape-based methods it is assumed that the domain is represented as in (5.9)–(5.10). The assumption that the optical parameters are constant in subdomains is usually also imposed, but can in principle be relaxed. In the explicit shape method we represent the boundaries by an explicit parameterization and construct a forward problem in terms of the coefficients of the parameterization and possibly the optical parameters inside each domain.

8.1.1. Parametric representation of surfaces. A typical parametrization of a closed surface is in terms of trigonometric functions (2D) or spherical harmonics (3D)

$$b_k = \begin{pmatrix} \mathrm{Re} \\ \mathrm{Im} \end{pmatrix} \begin{cases} \mathrm{e}^{\mathrm{i}k\vartheta} & \text{in 2D} \\ Y_l^m(\vartheta, \varphi),\ k = (l+1)^2 + m & \text{in 3D,} \end{cases} \tag{8.1}$$

where a purely real basis is constructed from the real and imaginary parts of the complex functions.

These bases may be used to parameterize the radial distance of the surface from an internal point, restricting the analysis to star-shaped objects, or to develop a a harmonic mapping of the surface onto a circle or sphere, which is made possible by the specific representation of each cartesian component separately in the basis.

In general, the higher order basis functions are roughly assumed to represent more detailed characteristics of the surface, whereas the lower order ones describe more the overall features like volume, orientation, etc.

For simplicity we introduce the notation

$$S = \{\gamma_k\}, \qquad k = 1, \ldots, K, \tag{8.2}$$

to describe the finite set of basis coefficients for the surface Σ up to degree K.

8.1.2. The forward problem. As well as the set of surface coefficients we define the set of parameters $\{x_\ell\}$, $\ell = 1, \ldots, L$, for each domain Ω_ℓ. Then the nonlinear forward operator is defined as the mapping from the optical properties $x = \{x_\ell\}$ of the individual regions and the geometric parameters S, to the measurements on the surface of $\partial\Omega$

$$y = \mathcal{F}(S, x). \tag{8.3}$$

8.1.3. The inverse problem. The most general inverse problem to be considered in this framework would be the simultaneous reconstruction of shape parameters γ and optical parameters x from the data y. However, in many practical applications, good estimates for the optical parameters in most of the regions Ω_i are available, and only those optical parameters which correspond to unknown anomalies need to be recovered. In particular, the outer boundary shape $\partial\Omega$ is usually considered known as well as the optical parameters in the outermost region. Relaxation of these assumptions, in particular the first, leads to more complex considerations that have so far not been addressed in optical tomography, but for which methods have been developed in related fields [162, 163].

As in the parameter identification problem, the key aspect of using the shape-based method is to discover a formula for calculating shape sensitivities functions which form the kernel of the Fréchet derivative of (8.3). Roughly speaking, the shape sensitivity function for shape basis function b_k is given by the spatial sensitivity function ρ defined in (2.21), projected onto b_k on Σ

$$\delta y_{j,i,k} = \langle \rho_{j,i}, b_k \rangle_\Sigma. \tag{8.4}$$

The precise form of the sensitivity function depends on the forward model. In [152] a FEM model was used and in [160] a BEM model. In the latter the model is naturally represented as the union of subdomains as in (5.9) and the domain boundaries are explicitly meshed. The fields Φ and the normal currents $J_n = \hat{v} \cdot \nabla\Phi$ are explicitly represented, and sensitivity function ρ is defined in terms of both the field and the normal currents. Shape updating requires only movement of the mesh points of the internal surfaces. By contrast, in the FEM model the surface Σ intersects to a subset of elements of the mesh. The appropriate form of (8.4) calls for the integration of products of forward and adjoint fields along split elements. If the simplified BEM model is used, where the normal currents are removed by taking the Schur complement of the discrete system (5.27), then the sensitivity functions form FEM and BEM are the same.

A shape reconstruction scheme searches for the set $\hat{S}$ that minimizes the distance between computed and measured data

$$\hat{S} = \arg\min_S \left[\mathcal{E}(S) := \tfrac{1}{2}\|y - \mathcal{F}(S, x)\|^2\right]. \tag{8.5}$$

In [152–154, 160] a Levenberg–Marquardt scheme was used for the inversion step

$$S^{(\mathrm{ind}+1)} = S^{(\mathrm{ind})} + (\mathcal{F}'^{*}_{\mathrm{ind}}\mathcal{F}'_{\mathrm{ind}} + \lambda \mathrm{I})^{-1}\mathcal{F}'^{*}_{\mathrm{ind}}(y - \mathcal{F}(S^{(\mathrm{ind})}, x)). \tag{8.6}$$

8.2. Implicit shape method

The level set technique represents the boundary of a domain as the zero-set of a function φ. Rather than updating the boundary of the domain an update rule for φ is developed and the optical parameters are defined by

$$x = \begin{cases} x_{\mathrm{int}} & \varphi \leqslant 0 \\ x_{\mathrm{ext}} & \varphi > 0, \end{cases} \tag{8.7}$$

providing a tool which is able to reconstruct interfaces in the region of interest together with certain characteristics of the interior and exterior subregions. The key difference from the explicit methods is that topological changes in the domains are easily accommodated. A description of this approach for the DA with examples in 2D is provided in [159] and the 3D case in [164]. Level set methods based on the RTE model in optical tomography were presented in [157, 158]. In these papers the scattering function μ_{s} was assumed known *a priori* and only the subdomain boundaries of the absorption function μ_{a} were reconstructed.

Within the diffusion approximation, consideration of the two parameters μ_{a} and D independently calls for the use of two level set function $\varphi_{\mu_{\mathrm{a}}}, \varphi_D$. Assuming that the background absorption and diffusion parameters μ_{ext} and D_{ext} are constant and known, the unknowns of the inverse problem are therefore $\varphi_{\mu_{\mathrm{a}}}, \varphi_D$ and the interior absorption and diffusion parameters μ_{int} and D_{int}. The forward mapping is defined by

$$y = \mathcal{F}(D(\varphi_D, D_{\mathrm{int}}), \mu_{\mathrm{a}}(\varphi_{\mu_{\mathrm{a}}}, \mu_{\mathrm{int}})), \tag{8.8}$$

and the output least-squares cost functional is given by

$$\mathcal{E}(\varphi_D, \varphi_{\mu_{\mathrm{a}}}, D_{\mathrm{int}}, \mu_{\mathrm{int}}) = \tfrac{1}{2}\|y - \mathcal{F}(D(\varphi_D, D_{\mathrm{int}}), \mu_{\mathrm{a}}(\varphi_{\mu_{\mathrm{a}}}, \mu_{\mathrm{int}}))\|^2. \tag{8.9}$$

The goal is to find a minimizer $(\hat{\varphi}_D, \hat{\varphi}_{\mu_{\mathrm{a}}}, \hat{D}_{\mathrm{int}}, \hat{\mu}_{\mathrm{int}})$ of this cost functional.

For solving the inverse problem an evolution approach is adopted in which the unknowns of the inverse problem are assumed to depend on an artificial evolution time t and evolve during the reconstruction

$$\begin{aligned} \frac{\partial \varphi_D}{\partial t} &= f_D(t), & \frac{\partial \varphi_{\mu_{\mathrm{a}}}}{\partial t} &= f_{\mu_{\mathrm{a}}}(t), \\ \frac{\partial D_{\mathrm{int}}}{\partial t} &= h_D(t), & \frac{\partial \mu_{\mathrm{int}}}{\partial t} &= h_{\mu_{\mathrm{a}}}(t). \end{aligned} \tag{8.10}$$

where f_D, $f_{\mu_{\mathrm{a}}}$, h_D, $h_{\mu_{\mathrm{a}}}$ are forcing terms which point into a descent direction of the cost (8.9). We use $\Sigma := \varphi = 0$ to denote the zero set of φ and

$$s = \mathcal{F}'^{*}(y - \mathcal{F}\left(D(\varphi_D, D_{\mathrm{int}}), \mu_{\mathrm{a}}(\varphi_{\mu_{\mathrm{a}}}, \mu_{\mathrm{int}})\right) \tag{8.11}$$

for the 'classical' descent direction (compare (7.19)). Then the forcing terms are defined

$$f_x = \mathcal{W}^{-1}(x_{\mathrm{int}} - x_{\mathrm{ext}})s, \tag{8.12}$$

$$h_x = (x_{\mathrm{int}} - x_{\mathrm{ext}})s|_{\Sigma}, \tag{8.13}$$

where $\mathcal{W}$ is a regularization term as in section 7.3. Note that the update for the level set φ is defined on the whole domain, whereas the update for the parameters is only defined on the zero set interface. In practice the level set update may be restricted to a narrow band within a specified distance of this interface. For details we refer to [159, 164].

The level set approach can easily be generalized to more complex scenarios such as smoothly varying or parameterized interior and exterior parameter profiles.

9. Conclusions and further topics

Before drawing this review to a conclusion we briefly mention some important topics which space precludes us from developing in further detail.

9.1. Anisotropy

As mentioned in section 3.5, if we relax the assumption that the RTE phase function is angularly invariant, we can develop an anisotropic model. The derivation of this model in terms of a spherical harmonic expansion (i.e. the P_N method) leads to a diffusion equation with a tensor diffusion coefficient [165, 166]

$$\frac{\partial}{\partial t}\Phi(\boldsymbol{r},t) - \nabla\cdot\mathsf{K}(\boldsymbol{r})\nabla\Phi(\boldsymbol{r},t) + c\mu_{\mathrm{a}}(\boldsymbol{r})\Phi(\boldsymbol{r},t) = 0. \tag{9.1}$$

In terms of the inverse problem we are lead to similar difficulties with regard to non-uniqueness as occur in impedance imaging. In particular a diffeomorphism of Ω that preserves the Robin-to-Neumann map will be equivalent to a change in the local orientation of the eigenvectors of K. As a mechanism for compensating for such non-uniqueness, the modelling error approach (see section 7.4) has been employed, [167, 168].

9.2. Refractive index variation

All the inverse problems we have discussed in this article were based on the transport equation or diffusion approximation with constant speed (i.e. constant refractive index). However, as discussed in section 3, there exist a number of models for variable refractive index including (3.27) and (3.79). In addition there have been several treatments of the radiative transport equation in piecewise homogeneous (i.e. layered) domains with constant absorption and scattering but differing refractive index [169–172] wherein semi-analytic methods can be derived in terms of Fresnel boundary conditions at the interfaces between sub-domains.

With regard to the corresponding terms of an inverse problem, references [173, 174] have considered the diffusion approximation version of the problem as in (3.79) as the basis for an inversion to recover refractive index changes.

9.3. Time-varying optical tomography

In real biological applications, the subject being studied is not static but time varying. In this case we have the choice to reconstruct time-point by time-point or to consider a spatial temporal model.There are two main considerations. Firstly the time scale on which x is changing may be comparable to the time scale on which the experiments are performed. In other words the data acquisition process may be corrupted by movement. In this case there are methods from time-series analysis in which the dynamic changes in x are included in the reconstruction model. A well-known example is the Kalman filter method which assumes a random walk model in some state space variables whose statistics is learnt over the course of time-series experiment. This idea was explored in diffuse optical tomography in [175]. In [176] the Kalman filter approach was extended to utilize a state space model in which the underlying temporal variation was taken to be a mixture of pseudo-sinusoidal signals with random amplitude and phase, which allowed separable reconstruction of functional components in brain images. In [177] a fully space-time model was used with data reduction achieved via a Kronecker product representation of forward operators.

The second consideration it that physiological models may be used to predict temporal variation which may be used to constrain the reconstruction. These methods used coupled

systems of nonlinear equations which are much more complex than those used in a state–space model [178]. A review can be found in [179]

9.4. Multimodality Imaging

By multimodality imaging we mean the combination of the varying capacities of two or more medical imaging technologies in extracting physiological and anatomical information of organisms in a complementary fashion, in order to enable the consistent retrieval of accurate and content rich biological information [180–190]. In optical tomography, the aim is to enhance the low-resolution functional image information with high-resolution complementary structural information using for example ultrasound [191, 192], MRI [193, 194] or x-rays [195, 196].

One method for utilizing the prior information is to construct priors that constrain the reconstruction of the optical information to be commensurate with the auxiliary information in a well-defined way. One approach is the use of structural priors which smooth the solution within regions but not across boundaries of regions inferred from the auxiliary modality [197–199]. An alternative is to define an information theoretic measure of similarity between the reconstructed image and the prior image [200].

The topic of multimodality imaging and the appropriate use of cross-information is rapidly expanding within optical tomography in particular and medical imaging in general.

Acknowledgments

SRA wishes to acknowledge the inspirational and pioneering work of David Delpy in the field of biomedical optics and to his many collaborators within the Biomedical Optics Research Laboratory (BORL) and the Centre for Medical Image Computing (CMIC) at UCL: in particular, Clare Elwell, Adam Gibson, Jem Hebden, Simon Prince, Martin Schweiger, Vadim Soloviev and Athanasios Zacharopoulos. He wishes to acknowledge the very fruitful and longstanding collaborations with David Boas, Rinaldo Cubeddu, Oliver Dorn, Paul French, Juha Heiskala, Jenni Heino, Jari Kaipio, Ville Kolehmainen, Ilkka Nissilä, Jorge Ripoll, Erkki Somersalo, Tanja Tarvainen, and Arjun Yodh amongst others. JCS is indebted to Britton Chance and John S. Leigh for their encouragement and generous support over the years. He would also like to express his gratitude to Vadim Markel, with whom he has had a long-standing and fruitful collaboration. In addition, he would like to thank Guillaume Bal, Scott Carney, Masaru Ishii, Manabu Machida, Shari Moskow, George Panasyuk, Soren Konecky and Zhengmin Wang who have helped to shape his understanding of optical tomography. Finally, he is grateful for the kind hospitality of Remi Carminati and Mathias Fink of Ecole Supérieure de Physique et Chimie Industrielles, where much of this review was written. Both authors would like to acknowledge that this work would not have been possible without the intervention of Scott Carney, who obediently locked us in the attic of his house in Leiden until a plan for this review was conceived. SRA acknowledges support that has been received from EPSRC, the Wellcome Trust, the British Council and the European Community framework programmes. JCS acknowledges support from NSF grant DMS-0554100 and NIH grant R01-EB-004832.

References

[1] Arridge S R 1999 Optical tomography in medical imaging *Inverse Problems* **15** R41–93
[2] Yodh A and Chance B 1995 Spectroscopy and imaging with diffusing light *Phys. Today* **48** 38–40

[3] Hebden J C, Arridge S R and Delpy D T 1997 Optical imaging in medicine: I. Experimental techniques *Phys. Med. Biol.* **42** 825–40

[4] Arridge S R and Hebden J C 1997 Optical imaging in medicine: II. Modelling and reconstruction *Phys. Med. Biol.* **42** 841–53

[5] Hawysz D and Sevick-Muraca E M 2000 Developments towards diagnostic breast cancer imaging using near-infrared optical measurements and flourescent contrast agents *Neoplasia* **2** 388–417

[6] Boas D A, Brooks D H, Miller E L, DiMarzio C A, Kilmer M, Gaudette R J and Zhang Q 2001 Imaging the body with diffuse optical tomography *IEEE Sig. Proc. Mag.* **18** 57–75

[7] Gibson A, Hebden J C and Arridge S R 2005 Recent advances in diffuse optical tomography *Phys. Med. Biol.* **50** R1–43

[8] Leff D R, Warren O J, Enfield L C, Gibson A, Athanasiou T, Patten D K, Hebden J, Yang G Z and Darzi A 2008 Diffuse optical imaging of the healthy and diseased breast: a systematic review *Breast Cancer Res. Treat.* **108** 9–22

[9] Ripoll J and Ntziachristos V 2004 Imaging scattering media from a distance: theory and applications of noncontact optical tomography *Mod. Phys. Lett.* B **18** 1403–31

[10] Ripoll J, Schulz R B and Ntziachristos V 2003 Free-space propagation of diffuse light: theory and experiments *Phys. Rev. Lett.* **91** 103901

[11] Cuccia D J, Bevilacqua F, Durkin A J and Tromberg B J 2005 Modulated imaging: quantitative analysis and tomography of turbid media in the spatial-frequency domain *Opt. Lett.* **30** 1354–6

[12] Hillman E M C and Burgess S A 2009 Sub-millimeter resolution 3D optical imaging of living tissue using laminar optical tomography *Laser Photonics Rev.* **3** 159–79

[13] van Rossum M W and Nieuwenhuizen T W 1999 Multiple scattering of classical waves: microscopy, mesoscopy and diffusion *Rev. Mod. Phys.* **71** 313–71

[14] Ishimaru A 1997 *Wave Propagation and Scattering in Random Media* (Piscataway, NJ: IEEE Press)

[15] Case K M and Zweifel P F 1967 *Linear Transport Theory* (Reading, MA: Addison-Wesley)

[16] Leakeas C L and Larsen E W 2001 Generalized Fokker–Planck approximations of particle transport with highly forward-peaked scattering *Nucl. Sci. Eng.* **137** 236–50

[17] Gonzalez-Rodriguez P and Kim A D 2008 Light propagation in tissues with forward-peaked and large-angle scattering *Appl. Opt.* **47** 2599–609

[18] Ryzhik L, Papanicolaou G and Keller J B 1996 Transport equations for elastic and other waves in random media *Wave Motion* **24** 327–70

[19] Bal G 2006 Radiative transfer equations with varying refractive index: a mathematical perspective *J. Opt. Soc. Am.* A **23** 1639–44

[20] Clark J, Gonzalez-Rodriguez P and Kim A D 2009 Using polarization to find a source in a turbid medium *J. Opt. Soc. Am.* A **26** 1129–38

[21] Sharpe J *et al* 2002 Optical projection tomography as a tool for 3D microscopy and gene expression studies *Science* **296** 541–5

[22] Vinegoni C *et al* 2008 In vivo imaging of Drosophila melanogaster pupae with mesoscopic fluorescence tomography *Nat. Methods* **5** 45–7

[23] Bal G 2009 Inverse transport theory and applications *Inverse Problems* **25** 053001

[24] Bal G, Langmore I and Monard F 2008 Inverse transport with isotropic sources and angularly averaged measurements *Inverse Problem Imaging* **2** 23–42

[25] Panasyuk G, Schotland J C and Markel V A 2006 Radiative transport equation in rotated reference frames *J. Phys. A: Math. Gen.* **39** 115–37

[26] Kim A D and Keller J B 2003 Light propagation in biological tissue *J. Opt. Soc. Am.* A **20** 92–8

[27] Kim A D 2004 Transport theory for light propagation in biological tissue *J. Opt. Soc. Am.* A **21** 820–7

[28] Kim A D and Schotland J C 2006 Self-consistent scattering theory for the radiative transport equation *J. Opt. Soc. Am.* A **23** 596–602

[29] Larsen E W and Keller J B 1974 Asymptotic solution of neutron-transport problems for small mean free paths *J. Math. Phys.* **15** 75–81

[30] Arridge S R and Lionheart W R B 1998 Non-uniqueness in diffusion-based optical tomography *Opt. Lett.* **23** 882–4

[31] Harrach B 2009 On uniqueness in diffuse optical tomography *Inverse Problems* **25** 055010

[32] Soloviev V Y, Tahir K B, McGinty J, Elson D S, Neil M A A, French P M W and Arridge S R 2007 Fluorescence lifetime imaging by using time gated data acquisition *Appl. Opt.* **46** 7384–91

[33] Milstein A B, S Oh, Reynolds J S, Webb K J, Bouman C A and Millane R P 2002 Three dimensional Bayesian optical diffusion tomography with experimental data *Opt. Lett.* **27** 95–7

[34] Joshi A, Bangerth W and Sevick-Muraca E M 2006 Non-contact fluorescence optical tomography with scanning patterned illumination *Opt. Express* **14** 6516–34

[35] Kumar A T N, Skoch J, Bacskai B J, Boas D A and Dunn A K 2005 Fluorescence-lifetime-based tomography for turbid media *Opt. Lett.* **30** 3357–49
[36] Kumar A T N, Raymond S B, Bacskai B J and Boas D A 2008 Comparison of frequency-domain and time-domain fluorescence lifetime tomography *Opt. Lett.* **33** 470–2
[37] Soloviev V Y, D'Andrea C, Valentini G, Cubeddu R and Arridge S R 2008 Simultaneous reconstruction of optical parameters, quantum yield and lifetime in turbid media using time gated data acquisition technique *Appl. Opt.* **48** 28–36
[38] Soloviev V Y, D'Andrea C, Brambilla M, Valentini G, Schulz R B, Cubeddu R and Arridge S R 2008 Adjoint time domain method for fluorescent imaging in turbid media *Appl. Opt.* **47** 2303–11
[39] Corlu A, Durduran T, Choe R, Schweiger M, Hillman E M C, Arridge S R and Yodh A G 2003 Uniqueness and wavelength optimization in continuous-wave multispectral diffuse optical tomography *Opt. Lett.* **28** 23
[40] Corlu A, Choe R, Durduran T, Lee K, Schweiger M, Arridge S R, Hillman E M C and Yodh A G 2005 Diffuse optical tomography with spectral constraints and wavelength optimisation *Appl. Opt.* **44** 2082–93
[41] Zacharopoulos A D, Svenmarker P, Axelsson J, Schweiger M, Arridge S R and Andersson-Engels S 2009 A matrix-free algorithm for multiple wavelength fluorescence tomography *Opt. Express* **17** 3042–51
[42] Kuo C, Coquoza O, Troy T, Zwang D and Rice B 2005 Bioluminescent tomography for *in vivo* localization and quantification of luminescence sources from a multiple-view imaging system *Mol. Imaging* **4** 370
[43] X Gu, Zhang Q, Larcom L and Jiang H 2004 Three-dimensional bioluminescence tomography with model-based reconstruction *Opt. Express* **12** 3996–4000
[44] Wang G, Y Li and Jiang M 2004 Uniqueness theorems in bioluminescence tomography *Med. Phys.* **31** 2289–99
[45] Cong W *et al* 2005 Practical reconstruction methods for bioluminescence tomography *Opt. Express* **13** 6756–71
[46] Chaudhari A J, Darvas F, Bading J R, Moasts R A, Conti P S, Smith D J, Cherry S R and Leahy R M 2005 Hyperspectral and multispectral bioluminescence optical tomography for small animal imaging *Phys. Med. Biol.* **50** 5421–41
[47] Alexandraxis G, Rannou F R and Chatziioannou A F 2005 Tomographic bioluminescence imaging by use of a combined optical-PET (OPET) system: a computer feasibility study *Phys. Med. Biol.* **50** 4225–41
[48] Abdoulaev G S and Hielscher A H 2003 Three-dimensional optical tomography with the equation of radiative transfer *J. Electron. Imaging* **12** 594–601
[49] Aydin E D, de Oliveira C R E and Goddard A J H 2002 A comparison between transport and diffusion calculations using a finite element-spherical harmonics radiation transport method *Med. Phys.* **2** 2013–23
[50] Richling S, E Meinköhn E, Kryzhevoi M and Kanschat G 2001 Radiative transfer with finite elements I Basic method and tests *Astron. Astrophys.* **380** 776–88
[51] Tarvainen T, Vauhkonen M, Kolehmainen V and Kaipio J P 2005 A hybrid radiative transfer-diffusion model for optical tomography *Appl. Opt.* **44** 876–86
[52] Tarvainen T, Vauhkonen M, Kolehmainen V and Kaipio J P 2006 Finite element model for the coupled radiative transfer equation and diffusion approximation *Int. J. Numer. Methods Eng.* **65** 383–405
[53] Buchan A G, Pain C C, Eaton M D, Smedley-Stevenson R P and Goddard A J H 2005 Linear and quadratic octahedral wavelets on the sphere for angular discretisations of the Boltzmann transport equation *Ann. Nucl. Energy* **32** 1224–73
[54] Dorn O 1998 A transport–backtransport method for optical tomography *Inverse Problems* **14** 1107–30
[55] Klose A D and Hielscher A H 1999 Iterative reconstruction scheme for optical tomography based on the equation of radiative transfer *Med. Phys.* **26** 1698–707
[56] Klose A D and Hielscher A H 2002 Optical tomography using the time-independent equation of radiative transfer: Part 2. Inverse model *J. Quant. Spectrosc. Radiat. Transfer* **72** 715–32
[57] Abdoulaev G S, Ren K and Hielscher A H 2005 Optical tomography as a PDE-constrained optimisation problem *Inverse Problems* **21** 1507–30
[58] Ren K, Bal G and Hielscher A 2006 Frequency domain optical tomography based on the equation of radiative transfer *SIAM J. Sci. Comput.* **28** 1463–89
[59] Lathrop K D 1968 Ray effects in discrete ordinates equations *Nucl. Sci. Eng.* **32** 357–69
[60] Lathrop K D 1971 Remedies for ray effects *Nucl. Sci. Eng.* **45** 255–68
[61] Tarvainen T, Vauhkonen M, Kolehmainen V, Arridge S R and Kaipio J P 2005 Coupled radiative transfer equation and diffusion approximation model for photon migration in turbid medium with low-scattering and non-scattering regions *Phys. Med. Biol.* **50** 4913–30
[62] Kanschat G 1998 A robust finite element discretization for radiative transfer problems with scattering *East-West J. Numer. Math* **6** 265–72
[63] Soloviev V Y and Krasnosselskaia L V 2006 Dynamically adaptive mesh refinement technique for image reconstruction in optical tomography *Appl. Opt.* **45** 2828–37
[64] Soloviev V Y 2006 Mesh adaptation technique for Fourier-domain fluorescence lifetime imaging *Med. Phys.* **33** 4176–83

[65] Kwon K, Yazici B and Guven M 2006 Two-level domain decomposition methods for diffuse optical tomography *Inverse Problems* **22** 1533–59
[66] Guven M, Yazici B, Kwon K, Giladi E and Intes X 2007 Effect of discretization error and adaptive mesh generation in diffuse optical absorption imaging: I *Inverse Problems* **23** 1115–33
[67] Guven M, Yazici B, Kwon K, Giladi E and Intes X 2007 Effect of discretization error and adaptive mesh generation in diffuse optical absorption imaging: II *Inverse Problems* **23** 1135–50
[68] Joshi A, Bangerth W, Hwan K, Rasmussen J C and Sevick-Muraca E M 2006 Fully adaptive fem based fluorescence tomography from time-dependant measurements with area illumination and detection *Med. Phys.* **33** 1299–310
[69] Lee J H, Joshi A and Sevick-Muraca E M 2007 Fully adaptive finite element based tomography using tetrahedral dual-meshing for fluorescence enhanced optical imaging in tissue *Opt. Express* **15** 6955–75
[70] Qin C, Tian J, Yang X, Liu K, Yan G, Fengand J, Y Lv and M Xu 2008 Galerkin-based meshless methods for photon transport in the biological tissue *Opt. Express* **16** 20317–33
[71] Sikora J, Zacharopoulos A, Douiri A, Schweiger M, Horesh L, Arridge S R and Ripoll J 2006 Diffuse photon propagation in multilayered geometries *Phys. Med. Biol.* **51** 497–516
[72] Aliabadi M H 2002 *The Boundary Element Method* (New York: Wiley)
[73] Becker A A 1992 *The Boundary Element Method in Engineering* (Maidenhead: McGraw-Hill)
[74] Beer G 2001 *Programming the Boundary Element Method. An Introduction for Engineers* (New York: Wiley)
[75] Sikora J, Riley J, Arridge S R, Zacharopoulos A D and Ripoll J 2004 Light propagation in diffusive media with non-scattering regions using 3D BEM *Proc. of 3rd Int. Conf. on Boundary Integral Methods: Theory and Applications* ed S C Wilde
[76] Sikora J and Arridge S R 2002 Some numerical aspects of 3D BEM application to optical tomography *4th Int. Workshop Computational Problems of Electrical Engineering* pp 59–62
[77] Bonnet M 1995 BIE and material differentiation applied to the formulation of obstacle inverse problems *Eng. Anal. Bound. Elem.* **15** 121–36
[78] Ripoll J, Ntziachristos V and Nieto-Vesperinas M 2001 The Kirchhoff approximation for diffusive waves *Phys. Rev.* E **64** 1–8
[79] Ripoll J and Ntziachristos V 2006 From finite to infinite volumes: removal of boundaries in diffuse wave imaging *Phys. Rev. Lett.* **96** 173903
[80] Prahl S A, Keijzer M, Jacques S L and Welch A J 1989 A Monte Carlo model of light propagation in tissue *Dosimetry of Laser Radiation in Medicine and Biology* vol 5, ed G J Müller and D H Sliney pp 102–11 (Bellingham, WA: SPIE)
[81] Wang L, Jacques S L and Zheng L 1995 MCML-Monte Carlo modeling of light transport in multi-layered tissues *Comput. Methods Programs Biomed.* **47** 131–46
[82] Boas D A, Culver J P, Stott J J and Dunn A K 2002 Three dimensional Monte Carlo code for photon migration through complex heterogenous media including the adult human head *Opt. Express* **10** 159–70
[83] Heiskala J, Hiltunen P and Nissila I 2009 Significance of background optical properties, time-resolved information and optode arrangement in diffuse optical imaging of term neonates *Phys. Med. Biol.* **54** 535–54
[84] Heiskala J, Nissilä I, Neuvonen T, Järvenpää S and Somersalo E 2005 Modeling anisotropic light propagation in a realistic model of the human head *Appl. Opt.* **44** 2049–57
[85] Heiskala J 2009 Accurate modeling of tissue properties in diffuse optical imaging of the human brain *PhD Thesis* Helsinki University of Technology
[86] Steinbrink J, Wabnitz H, Obrig H, Villringer A and Rinneberg H 2001 Determining changes in NIR absorption using a layered model of the human head *Phys. Med. Biol.* **46** 879–96
[87] Alerstam E, Svensson T and Andersson-Engels S 2008 Parallel computing with graphics processing units for high-speed Monte Carlo simulation of photon migration *J. Biomed. Opt.* **13** 060504
[88] Alerstam E, Andersson-Engels S and Svensson T 2008 White Monte Carlo for time-resolved photon migration *J. Biomed. Opt.* **13** 041304
[89] Wang L and Jacques S L 1993 Hybrid model of Monte Carlo simulation diffusion theory for light reflectance by turbid media *J. Opt. Soc. Am.* A **10** 1746–52
[90] Alexandrakis G, Farrell T and Patterson M 2000 Monte Carlo diffusion hybrid model for photon migration in a two-layer medium in the frequency *Appl. Opt.* **39** 2235–44
[91] Hayashi T, Kashio Y and Okada E 2003 Hybrid Monte Carlo-diffusion method for light propagation in tissue with a low-scattering region *Appl. Opt.* **42** 2888–96
[92] Kim A D and Keller J B 2003 Light propagation in biological tissue *J. Opt. Soc. Am.* A **20** 92–8
[93] Bal G and Maday Y 2002 Coupling of transport and diffusion models in linear transport theory *Math. Model. Numer. Anal.* **36** 69–86

[94] Custo A, W M Wells III, Barnett A H, Hillman E M C and Boas D A 2006 Effective scattering coefficient of the cerebral spinal fluid in adult head models for diffuse optical imaging *Appl. Opt.* **45** 4747–55
[95] Ackroyd R T, Issa J G and Riyait N S 1986 Treatment of voids in finite element transport methods *Prog. Nucl. Energy* **18** 85–9
[96] Ackroyd R T and Riyait N S 1989 Iteration and extrapolation methods for the approximate solution of the even-parity transport equation for systems with voids *Ann. Nucl. Energy* **16** 1–32
[97] Firbank M, Arridge S R, Schweiger M and Delpy D T 1996 An investigation of light transport through scattering bodies with non-scattering regions *Phys. Med. Biol.* **41** 767–83
[98] Hielscher A H, Alcouffe R E and Barbour R L 1998 Comparison of finite-difference transport and diffusion calculations for photon migration in homogeneous and heterogeneous tissue *Phys. Med. Biol.* **43** 1285–302
[99] Arridge S R, Dehghani H, Schweiger M and Okada E 2000 The finite element model for the propagation of light in scattering media: a direct method for domains with non-scattering regions *Med. Phys.* **27** 252–64
[100] Dehghani H, Arridge S R, Schweiger M and Delpy D T 2000 Optical tomography in the presence of void regions *J. Opt. Soc. Am.* A **17** 1659–70
[101] Bal G and Ren K 2003 Generalised diffusion model in optical tomography with clear layers *J. Opt. Soc. Am.* A **20** 2355–64
[102] Okada E, Firbank M, Schweiger M, Arridge S R, Cope M and Delpy D T 1997 Theoretical and experimental investigation of near-infrared light propagation in a model of the adult head *Appl. Opt.* **36** 21–31
[103] Ripoll J 2000 Difusión de Luz en Medios Turbos con Aplicación Biomédica *PhD Thesis* University Autónoma of Madrid
[104] Ripoll J, Arridge S R, Dehghani H and Nieto-Vesperinas M 2000 Boundary conditions for light propagation in diffusive media with nonscattering regions *J. Opt. Soc. Am.* A **17** 1671–81
[105] Hyvönen N 2002 Analysis of optical tomography with non-scattering regions *Proc. Edinburgh Math. Soc.* **35** 257–76
[106] Hyvönen N 2004 Diffusive tomography methods: special boundary conditions and characterization of inclusions *PhD thesis* Dept. Engineering Physics and Mathematics, Helsinki University of Technology
[107] Hyvönen N 2007 Frechet derivative with respect to the shape of a strongly convex nonscattering region in optical tomography *Inverse Problems* **23** 2249–70
[108] Hyvönen N 2007 Locating transparent regions in optical absorption and scattering tomography *SIAM J. Appl. Math.* **67** 1101–23
[109] Bal G and Ren K 2006 Reconstruction of singular surfaces by shape sensitivity analysis and level set method *Math. Models Methods Appl. Sci.* **16** 1347–73
[110] Epstein C L and Schotland J C 2008 The bad truth about Laplace's transform *SIAM Rev.* **50** 504–20
[111] Schotland J C 1997 Continuous-wave diffusion imaging *J. Opt. Soc. Am.* A **14** 275–9
[112] Schotland J C and Markel V A 2001 Inverse scattering with diffusing waves *J. Opt. Soc. Am.* A **18** 2767–77
[113] Markel V A and Schotland J C 2004 Symmetries, inversion formulas and image reconstruction for optical tomography *Phys. Rev.* E **70** 056616
[114] Markel V A and Schotland J C 2002 Effects of sampling and limited data in optical tomography *App. Phys. Lett.* **81** 1180–2
[115] Wang Z M, Panasyuk G, Markel V A and Schotland J C 2005 Experimental demonstration of an analytic method for image reconstruction in optical diffusion tomography with large data sets *Opt. Lett.* **30** 3338–40
[116] Konecky S *et al* 2008 Imaging complex structures with diffuse light *Opt. Express* **16** 5048–60
[117] Moses H E 1956 Calculation of the scattering potential from reflection coefficients *Phys. Rev.* **102** 550–67
[118] Prosser R 1969 Formal solutions of the inverse scattering problem *J. Math. Phys.* **10** 1819–22
[119] Markel V A, O'Sullivan J A and Schotland J C 2003 Inverse problem in optical diffusion tomography. IV. Nonlinear inversion formulas *J. Opt. Soc. Am.* A **20** 903–12
[120] Moskow S and Schotland J C 2008 Convergence and stability of the inverse scattering series for diffuse waves *Inverse Problems* **24** 065005
[121] Moskow S and Schotland J C 2009 Numerical studies of the inverse Born series for diffuse waves *Inverse Problems* **25** 095007
[122] Schotland J C and Markel V A 2007 Fourier–Laplace structure of the inverse scattering problem for the radiative transport equation *Inverse Problems Imaging* **1** 181–8
[123] Markel V A, Florescu L and Schotland J C 2009 Single-scattering optical tomography *Phys. Rev.* E **79** 036607
[124] Calvetti D and Somersalo E 2007 *Introduction to Bayesian Scientific Computing* (New York: Springer)
[125] Bertero M and Boccacci M 1998 *Introduction to Inverse Problems in Imaging* (Bristol: Institute of Physics Publishing)
[126] Golub G H and van Loan C F 1983 *Matrix Computations* (Baltimore, MD: John Hopkins University)

[127] Kaczmarz S 1937 Angenäherte auflösung von systemen linearer gleichungen *Bull. Int. Acad. Polon. Sci. Lett.* A **35** 335–57
[128] Strohmer T and Vershynin R 2006 A randomized solver for linear systems with exponential convergence *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (Lecture Notes in Computer Science,* vol 4110*)* (Berlin/Heidelberg: Springer) pp 499–507
[129] Schweiger M, Arridge S R and Nissilä I 2005 Gauss-Newton method for image reconstruction in diffuse optical tomography *Phys. Med. Biol.* **50** 2365–86
[130] Tarvainen T, Vauhkonen M and Arridge S R 2008 Image reconstruction in optical tomography using the finite element solution of the frequency domain radiative transfer equation *J. Qaunt. Spect. Rad. Transfer* **109** 2767–78
[131] Arridge S R and Schweiger M 1998 A gradient-based optimisation scheme for optical tomography *Opt. Express* **2** 213–26
[132] Kim H K and Charette A 2007 A sensitivity function-based conjugate gradient method for optical tomography with the frequency-domain equation of radiative transfer *J. Quant. Spectrosc. Radiat. Transf.* **104** 24–39
[133] Nocedal J and Wright S J 1999 *Numerical Optimization* (New York: Springer)
[134] Liu D C and Nocedal J 1989 On the limited memory BFGS method for large scale optimization *Math. Prog.* B **45** 503–28
[135] Natterer F and Wübbeling F 2001 *Mathematical Methods in Image Reconstruction* (Philadelphia: SIAM)
[136] Gonzalez-Rodriguez P and Kim A D 2009 Comparison of light scattering models for diffuse optical tomography *Opt. Express* **17** 8756–74
[137] Fessler J A 1994 Penalized weighted least-squares image reconstruction for positron emission tomography *IEEE Trans. Med. Imaging* **13** 290–300
[138] J C Ye, Webb K J, Bouman C A and Millane R P 1999 Optical diffusion tomography by iterative coordinate-descent optimization in a Bayesian framework *J. Opt. Soc. Am.* A **16** 2400–12
[139] Milstein A B, S Oh, Webb K J, Bouman C A, Zhanga Q, Boas D A and Millane R P 2003 Fluorescence optical diffusion tomography *Appl. Opt.* **42** 3081–94
[140] Ye J C, Bouman C A, Webb K J and Millane R P 2001 Nonlinear multigrid algorithms for Bayesian optical diffusion tomography *IEEE Trans. Imaging Proc.* **10** 909–22
[141] Bangerth W and Joshi A 2008 Adaptive finite element methods for the solution of inverse problems in optical tomography *Inverse Problems* **24** 034011
[142] Kim H K and Hielscher A H 2009 A PDE-constrained SQP algorithm for optical tomography based on the frequency-domain equation of radiative transfer *Inverse Problems* **25** 015010
[143] Kaipio J and Somersalo E 2005 *Statistical and Computational Inverse Problems* (New York: Springer)
[144] Arridge S R, Kaipio J P, Kolehmainen V, Schweiger M, Somersalo E, Tarvainen T and Vauhkonen M 2006 Approximation errors and model reduction with an application in optical diffusion tomography *Inverse Problems* **22** 175–96
[145] Bouman C A and Sauer K 1993 A generalised Gaussian image model for edge-preserving MAP estimation *IEEE Trans. Image Process.* **2** 296–310
[146] Douiri A, Schweiger M, Riley J and Arridge S R 2005 Local diffusion regularization method for optical tomography reconstruction by using robust statistics *Opt. Lett.* **30** 2439–41
[147] Pineda A R, Schweiger M, Arridge S R and Barrett H H 2006 Information content of data types in time-domain optical tomography *J. Opt. Soc. Am.* A **12** 2989–96
[148] Acar R and Vogel C R 1994 Analysis of bounded variation penalty methods for ill-posed problems *Inverse Problems* **10** 1217–29
[149] Geman S and Geman D 1984 Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images *IEEE Trans. Patt. Anal. Mach. Intell.* **6** 721–41
[150] Rudin L I, Osher S and Fatemi E 1992 Nonlinear total variation based noise removal algorithm *Physica* D **60** 259–68
[151] Vogel C R and Oman M E 1996 Iterative methods for total variation denoising *SIAM J. Sci. Comput.* **17** 227–38
[152] Kolehmainen V, Arridge S R, Lionheart W R B, Vauhkonen M and Kaipio J P 1999 Recovery of region boundaries of piecewise constant coefficients of an elliptic PDE from boundary data *Inverse Problems* **15** 1375–91
[153] Kolehmainen V, Arridge S R, Vauhkonen M and Kaipio J P 2000 Simultaneous reconstruction of internal tissue region boundaries and coefficients in optical diffusion tomography *Phys. Med. Biol.* **45** 3267–83
[154] Kolehmainen V, Vauhkonen M, Kaipio J P and Arridge S R 2000 Recovery of piecewise constant coefficients in optical diffusion tomography *Opt. Express* **7** 468–80
[155] Kilmer M E, Miller E L, Boas D and Brooks D 2000 A shape-based reconstruction technique for DPDW data *Opt. Express* **7** 481–91

[156] Kilmer M E, Miller E L, Barbaro A and Boas D 2003 Three-dimensional shape-based imaging of absorption perturbation for diffuse optical tomography *Appl. Opt.* **42** 3129–44
[157] Dorn O 2004 Shape reconstruction in scattering media with voids using a transport model and level sets *Canad. Appl. Math. Q.* **10** 239–75
[158] Dorn O 2009 Shape reconstruction for an inverse radiative transfer problem arising in medical imaging *Springer Series Computational Science and Engineering: Numerical Methods for Multidimensional Radiative Transfer Problems* ed R Rannacher (Berlin: Springer) pp 299–309
[159] Schweiger M, Arridge S R, Dorn O, Zacharopoulos A and Kolehmainen V 2006 Reconstructing absorption and diffusion shape profiles in optical tomography using a level set technique *Opt. Lett.* **31** 471–3
[160] Zacharopoulos A, Arridge S R, Dorn O, Kolehmainen V and Sikora J 2006 3D shape reconstruction in optical tomography using spherical harmonics and BEM *J. Electromagn. Waves Appl.* **20** 1827–36
[161] Arridge S R, Dorn O, Kaipio J P, Kolehmainen V, Schweiger M, Tarvainen T, Vauhkonen M and Zacharopoulos A 2006 Reconstruction of subdomain boundaries of piecewise constant coefficients of the radiative transport equation from optical tomography data *Inverse Problems* **22** 2175–98
[162] Kolehmainen V, Lassa M and Ola P 2007 The inverse conductivity problem with an imperfectly known boundary in three dimensions *SIAM J. Appl. Math.* **67** 1440–52
[163] Kolehmainen V, Lassa M and Ola P 2008 Electrical impedance tomography problem with inaccurately known boundary and contact impedances *IEEE Trans. Med. Imaging* **27** 1404–14
[164] Schweiger M, Dorn O and Arridge S R 2008 3D shape and contrast reconstruction in optical tomography with level sets *Proc. 4th AIP Int. Conf. and the 1st Congr. of the IPIA, J. Phys.: Conf. Ser.* **124** 012043
[165] Heino J and Somersalo E 2002 Estimation of optical absorption in anisotropic background *Inverse Problems* **18** 559–57
[166] Heino J, Arridge S R, Sikora J and Somersalo E 2003 Anisotropic effects in highly scattering media *Phys. Rev.* E **68** 31908
[167] Heino J and Somersalo E 2004 A modelling error approach for the estimation of optical absorption in the presence of anisotropies *Phys. Med. Biol.* **49** 4785–98
[168] Heino J, Somersalo E and Kaipio J P 2005 Compensating for geometrix mismodelling by anisotropies in optical tomography *Opt. Express* **13** 296–308
[169] Cassell J S and Williams M M R 2006 Radiation transport and internal reflection in a sphere *J. Quant. Spectrosc. Radiat. Transf.* **101** 16–28
[170] Cassell J S and Williams M M R 2007 Radiation transport and internal reflection in a two region, turbid sphere *J. Quant. Spectrosc. Radiat. Transf.* **104** 400–27
[171] Garcia R D M, Siewert C E and Yacout A M 2008 On the use of Fresnel boundary and interface conditions in radiative transfer calculations for multilayered media *J. Quant. Spectrosc. Radiat. Transf.* **109** 752–69
[172] Elaloufi R, Arridge S R, Pierrat R and Carminati R 2007 Light propagation in multilayered scattering media beyond the diffusive regime *Appl. Opt.* **46** 3628–38
[173] Khan T and Thomas A 2006 Comparison of p_n or spherical harmonics approximation for scattering media with spatially varying and spatially constant refractive indices *Opt. Commun.* **1** 130–66
[174] Khan T and Thomas A 2006 Inverse problem in refractive index based optical tomography *Opt. Commun.* **22** 1121–37
[175] Kolehmainen V, Prince S, Arridge S R and Kaipio J P 2000 A state estimation approach to non-stationary optical tomography problem *J. Opt. Soc. Am.* A **20** 876–84
[176] Prince S, Kolehmainen V, Kaipio J P, Franceschini M A, Boas D and Arridge S R 2003 Time series estimation of biological factors in optical diffusion tomography *Phys. Med. Biol.* **48** 1491–504
[177] Zhang Q, Brooks D H and Boas D A 2005 A haemodynamic response function model in spatio-temporal diffuse optical tomography *Phys. Med. Biol.* **50** 4625–44
[178] Diamond S G, Huppert T J, Kolehmainen V, Franceschini M A, Kaipio J P, Arridge S R and Boas D A 2006 Dynamic physiological modeling for functional diffuse optical tomography *Neuroimage* **30** 88–101
[179] Huppert T J, Diamond S G, Franceschini M A and Boas D A 2009 HomER: A review of time-series analysis methods for near-infrared spectroscopy of the brain *Appl. Opt.* **48** 280–98
[180] Ardekani B A, Braun M, Hutton B F, Kanno I and Iida H 1996 Minimum cross-entropy reconstruction of PET images using prior anatomical information *Phys. Med. Biol.* **41** 2497–517
[181] Comtat C, Kinahan P E, Fessler J A, Beyer T, Townsend D W, Defrise M and Michel C 2002 Clinically feasible reconstruction of 3D whole body PET/CT data using blurred anatomical labels *Phys. Med. Biol.* **47** 1–20
[182] Gindi G, Lee M, Rangarajan A and Zubal I G 1993 Bayesian reconstruction of functional images using anatomical information as priors *IEEE Trans. Med. Imaging* **12** 670–80
[183] H H Lu, Chen C M and Yang I H 1998 Cross-reference weighted least squares estimates for positron emission tomography *IEEE Trans. Med. Imaging* **17** 1–8

[184] Baillet S, Garnero L, Gildas M and Hugonin J-P 1999 Combined MEG and EEG source imaging by minimization of mutual information *IEEE Transactions on Biomedical Engineering* **46** 522–34
[185] Rangarajan A, Hsiao I T and Gindi G 2000 A Bayesian joint mixture framework for the integration of anatomical information in functional image reconstruction *J. Math. Imaging Vis.* **12** 119–217
[186] Sastry S and Carson R E 1997 Multimodality Bayesian algorithm for image reconstruction in positron emission tomography: a tissue composition model *IEEE Trans. Med. Imaging* **16** 750–61
[187] Som S, Hutton B H and Braun M 1998 Properties of minimum cross-entropy reconstruction of emission tomography with anatomically based prior *IEEE Trans. Nucl. Sci.* **46** 3014–21
[188] Zaidi H, Montandon M L and Slosman D O 2003 Magnetic resonance image-guided attenuation correction in 3D brain positron emission tomography *Med. Phys.* **30** 937–48
[189] Ahlfors S P and Simpson G V 2004 Geometrical interpretation of fMRI-guided MEG/EEG inverse estimates *NeuroImage* **22** 323–32
[190] Babiloni F *et al* 2005 Estimation of the cortical functional connectivity with the multimodal integration of high-resolution EEG and fMRI data by directed transfer function *Neuroimage* **24** 118–31
[191] Zhu Q, Chen N and Kurtzman S H 2003 Imaging tumour angiogenesis by use of combined near-infrared diffusive light and ultrasound *Opt. Lett.* **25** 337–9
[192] Zhu Q, Kurtzma S H, Hegde P, Tannenbaum S, Kane M, Huang M, Chen N G, Jagjivan B and Zarfos K 2005 Utilizing optical tomography with ultrasound localization to image heterogeneous hemoglobin distribution in large breast cancers *Neoplasia* **7** 263–70
[193] Ntziachristos V, Yodh A G, Schnall M D and Chance B 2002 MRI-guided diffuse optical spectroscopy of malignant and benign breast lesions *Neoplasia* **4** 347–54
[194] Brooksby B A, Dehghani H, Pogue B W and Paulsen K D 2003 Near-infrared (NIR) tomography breast image reconstruction with *a priori* structural information from MRI: algorithm development for reconstructing heterogeneities *IEEE J. Quantum Electron.* **9** 199–209
[195] Li A *et al* 2003 Tomographic optical breast imaging guided by three-dimensional mammography *J. Appl. Opt.* **42** 5181–90
[196] Zhang Q *et al* 2005 Coregistered tomographic X-ray and optical breast imaging: Initial results *J. Biomed. Opt.* **10** 024033 (1–9)
[197] Kaipio J P, Kolehmainen V, Vauhkonen M and Somersalo E 1999 Inverse problems with structural prior information *Inverse Problems* **15** 713–29
[198] Brooksby B, Jiang S, Dehghani H, Pogue B W and Paulsen K D 2005 Combining near infrared tomography and magnetic resonance imaging to study in vivo breast tissue: implementation of a Laplacian-type regularization to incorporate MR structure *J. Biomed. Opt.* **10** 0515041–10
[199] Douiri A, Schweiger M, Riley J and Arridge S R 2007 Anisotropic diffusion regularisation methods for diffuse optical tomography using edge prior information *Meas. Sci. Technol.* **18** 87–95
[200] Panagiotou C, Somayajula S, Gibson A P, Schweiger M, Leahy R M and Arridge S R 2009 Information theoretic regularization in diffuse optical tomography *J. Opt. Soc. Am.* A **26** 1277–90