Mathematical and numerical challenges in Diffuse Optical Tomography inverse problems

Andrea Aspri[1], **Alessandro Benfenati**[2], **Paola Causin**[1], **Cecilia Cavaterra**[1,3], **Giovanni Naldi**[2]

[1] Department of Mathematics, University of Milan, via Saldini 50, 20133 Milan, Italy
[2] Department of Environmental Science and Policy, University of Milan, via Celoria 2, 20133 Milan, Italy
[3] IMATI, CNR, Via Ferrata 1, 27100 Pavia, Italy

E-mail: andrea.aspri@unimi.it, alessandro.benfenati@unimi.it, paola.causin@unimi.it, cecilia.cavaterra@unimi.it, giovanni.naldi@unimi.it

Dedicated to Pierluigi Colli on the occasion of his 65th birthday

Abstract. Computed Tomography (CT) is an essential imaging tool for medical inspection, diagnosis and prevention. While X-rays CT is a consolidated technology, there is nowadays a strong drive for innovation in this field. Between the emerging topics, Diffuse Optical Tomography (DOT) is an instance of Diffuse Optical Imaging which uses non-ionizing light in the near-infrared (NIR) band as investigating signal. Non-trivial challenges accompany DOT reconstruction, which is a severely ill-conditioned inverse problem due to the highly scattering nature of the propagation of light in biological tissues. Correspondingly, the solution of this problem is far from being trivial. In this review paper, we first recall the theoretical basis of NIR light propagation, the relevant mathematical models with their derivation in the perspective of a hierarchy of modeling approaches and the analytical results on the uniqueness issue and stability estimates. Then we describe the state-of-the-art in analytic theory and in computational and algorithmic methods. We present a survey of the few contributions regarding DOT reconstruction aided by machine learning approaches and we conclude providing perspectives in the mathematical treatment of this highly challenging problem.

Keywords: Diffuse Optical Tomography, Diffuse Optical Imaging, Regularization of inverse problems, CT reconstruction, Deep Learning

Mathematics Subject Classification (2020): 35R30, 68T05, 68T07, 92C55, 65M32

Contents

1. Introduction

Imaging has become vital to many aspects of medicine, including detection and diagnosis of disease, treatment planning, and monitoring of response to therapy. Computerized Tomography (CT) is a central pillar in this field and has offered a huge beneficial impact in clinical practice. Based on X-rays as investigating signals, CT was developed all along the 20th century with an exceptional theoretical and technological effort: we refer to [126] for a historical review written in occasion of the 50th anniversary of the first CT scan of a live patient. Nowadays, advances in CT imaging technology explore different paths aimed at overcoming limitations of the traditional approach, in first place to reduce patients' exposure to radiations [2]. A promising class of approaches uses different non–ionizing investigating signals (light, sound, electric currents). These modalities are intended as alternatives to X-ray CT but also as complementary techniques able to provide additional information about the targeted organ, for example related to its metabolic or haemodynamic functional status. In this paper we focus on the so–called Diffuse Optical Tomography (DOT), an instance of Diffuse Optical Imaging

(DOI). This functional imaging tool relies on the relatively low attenuation of near-infrared (NIR) light to probe the internal optical properties of biological tissues. DOT imaging has been explored in different medical fields, including brain [128], thyroid and breast cancer imaging [49, 129, 105] and, more in general, for detecting and monitoring functional changes related to blood flow. In DOT, one reconstructs images of tissue spatial optical properties from boundary measurements of the light propagated through the tissue itself [18]. Photons emerging from the biological samples (classified as turbid media) consist of a mixture of very few coherent and quasi–coherent photons, with a predominant component of incoherent photons which experienced multiple scattering events. Due to these latter phenomena and to measurement noise, DOT reconstruction is a mathematically severely ill-posed inverse problem: in this framework, standard backprojection algorithms, as routinely employed in X-ray CT, have indeed limited applicability, and more complex image-reconstruction algorithms must be adopted [21]. Model-based image reconstruction algorithms have been for several years the standard approach in this field; they usually consist of three components [89]: i) a forward model, often based on partial differential equations, that provides a prediction of the measurements based on a guess of the system parameters; ii) an objective functional that compares the predicted data with the measured data; iii) an efficient way of updating the system parameters of the forward model, which in turn furnishes a new set of predicted data [88]. This latter point has been approached both by linearization (Born or Rytov approximations developed from an analytical solution to the forward model) and nonlinear iterations formulated as optimization problems. A more recent trend is machine learning-supported DOT image reconstruction, which is motivated by the outstanding performance of deep learning on computer vision problem tasks. Very few examples exist at present of this approach.

Aim of this paper is to review from a mathematical viewpoint, both in its theoretical analysis and computational aspects, the DOT reconstruction problem. Namely, we consider the inverse problem of reconstructing internal optical properties of a body/sample by measurements which are typically available on the boundary of the investigated body [13]. In practice, the measurements do not provide a full knowledge of the data. As a matter of fact, even if one performs a set of different experiments, with various excitation patterns (*i.e.*, with different influx and/or source terms) measurements can be collected only at discrete locations along the boundary of the domain. This implies that one has only partial knowledge of the induced response. Due to relevant technological differences, which directly impact on the reconstruction problem, we specifically focus on DOT based on Steady-State-Domain (SSD) systems, which are among the most widely used in clinical settings for breast cancer screening, a main field of application of DOT [104]. In this technology, the light source continuously emits light into the tissue at a single frequency (continuous wave modality, CW) and the propagated light fluence is measured on the tissue boundary. In some "non-contact" applications, the situation may also be more involved, since free-space propagation from

the sample boundary to a distant detector, for example a CCD sensor, through another medium (air, liquid) should also be considered. We will however not enter into details of this latter aspect.

We start discussing in Section 2 the derivation of the optical coefficients - which are the target quantities in the reconstruction - and we proceed to show model-driven approaches, being the most established and historical resolution methods. To do this, first we present in Section 3 the staple Radiative Transfer Model and then we derive the so-called "Diffuse Approximation", based on a Schrödinger type equation. This is the most known and applied approximation used in the applications thanks to the description of light propagation with a reasonable accuracy and to the possibility of computing its numerical solution with efficient and fast algorithms. We then recall the main analytical results concerning the inverse problem of reconstructing both the diffusion and the absorption coefficients in the diffusion approximation equation, especially the well-known non-uniqueness result due to Arridge and Lionheart [20]. However, imposing some a priori assumptions on the coefficients, a uniqueness result can be proven and stability estimates provided. Numerical algorithms are then discussed in Section 3 to tackle the computational solution of the problem, with a specific accent on regularization procedures. We also outline the concept of Bayesian approach and implementations in the DOT framework. Eventually, since recent progress in computational technology, including artificial intelligence (AI) and high-performance supercomputers, are expected to have a deep influence also on reconstruction in DOT, we address this point in Section 4 by presenting the available data-enhanced reconstruction approaches. Finally, in Section 5 we draw the conclusions of the paper and discuss some perspectives in this innovative and complex field.

For ease of reading, Tab.1 summarizes the main symbols used in the text, classified accordingly to the area they pertain. It is our convention to use bold letters for vectors as well as for discretized functions and matrices.

2. Physical processes in light–tissue interaction and optical coefficients

In this section we provide a concise discussion about the main physical phenomena occurring in light propagation in highly scattering media as biological tissues. Then, we present a derivation of the optical coefficients that mathematically describe the above phenomena and are the object of the DOT reconstruction.

The optical properties of biological tissues are the result of the internal tissue microstructure and refractive index distribution. These correlations are highly complex since due to structural complexity and diversity, with typical sizes ranging from tenths of nanometers to hundreds of micrometers. Figure 1a) shows the principal phenomena occurring to a light beam incident on a slice of biological matter and, namely, reflection, absorption and (multi)scattering. Such physical processes can be quantified by a series

Nomenclature	Description
	Geometrical quantities
d	spatial dimension (=2,3)
$\mathbf{r}$	position vector
$\widehat{\mathbf{s}}', \widehat{\mathbf{s}}$	in, out coming light directions
$d\widehat{\mathbf{s}}$	solid angle around direction $\widehat{\mathbf{s}}$
S^{d-1}	unit sphere in $\mathbb{R}^d$
	Physical model parameters
ν	refraction coefficient
μ_s, μ_s'	scattering coefficient, reduced scattering coefficient
D, μ_a	diffusion and absorption coefficient
L, u	light radiance (RTE model), fluence (DA model)
$p_s(\widehat{\mathbf{s}}, \widehat{\mathbf{s}}')$, g	normalized scattering phase function, avg cosine of scatter
δ	noise level
	Functional spaces and operators
Q, U, Y	optical properties, optical field and data Banach spaces
Θ	generic parameter space
$\mathcal{M} : D(\mathcal{M}) \subset (Q \times U) \to Y$	measurement map
$\mathcal{F} : \Theta \to Y$	forward map
$\mathcal{D} : Y \to [0, \infty]$	discrepancy measure
$\mathcal{R} : \Theta \to [0, \infty]$	regularization functional (with parameter λ)
	Other mathematical symbols
Ω, $\partial\Omega$	computational domain, boundary of the domain
$\mathbf{n}$	unit normal vector on $\partial\Omega$
	Discretization parameters
h	mesh size

Table 1. Symbols (organized into macro-areas) and relative significance as used in the text.

of parameters, which are presented as coefficients and include the index of refraction (ν) and the absorption (μ_a) and scattering (μ_s) coefficients. Notice that the definition of such coefficients implicitly impinges on the hypothesis that the tissue is as an isotropic

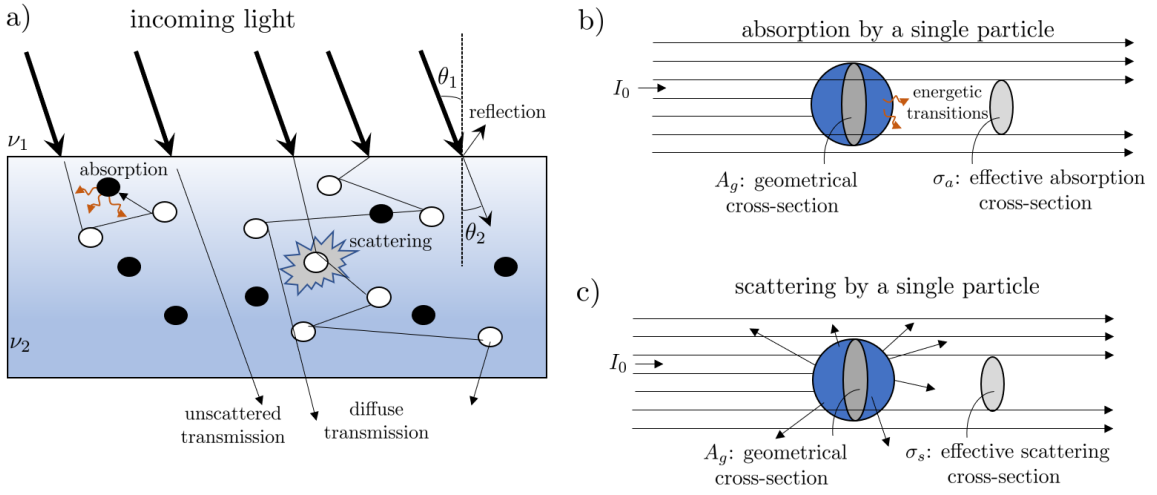

Figure 1. a) Main physical phenomena occurring in light propagation in biological tissues, turbid media characterized by a highly complex structure; b) absorption by a single particle; c) scattering by a single particle.

medium. While the isotropy assumption is reasonable for "randomly" organized soft tissues as the breast, it may not be true for highly structured tissues such as the muscle, where a fully anisotropic description would be more appropriate.

2.1. Refraction

The index of refraction is a measure of the bending of a ray of light when passing from one medium into another. The (real part of the) refractive index at wavelength λ is defined in terms of the phase velocity of light in the medium $c_m = c_m(\lambda)$ with respect to the velocity of light in vacuum c as $\nu(\lambda) = c/c_m$. When a light wave propagating in a material with a given refractive index encounters a boundary with a medium with a mismatch in the refraction index, the path of the light is redirected and reflection and refractive transmission result, see Figure 1a), where ν_1 and ν_2 indicate the refraction indices of the air and tissue, respectively. The amount of light reflected by and transmitted through the boundary depends on the refractive indices of the two materials, the angle of incidence (via the Snell's law $\nu_1 \sin\theta_1 = \nu_2 \sin\theta_2$, see Figure 1a), and the polarization of the incoming wave [109].

2.2. Absorption

Absorption is a process involving the extraction of energy from light by a molecular species -and in particular by some of its parts, known as chromophores - which undergo a transition process (electronic transition or vibrational state transition). The probability of transition between different states or energy levels is governed by complex quantum mechanical rules that depend on the chemical structure, size, and symmetry of the molecules. The level of absorption of incident light by tissue chromophores

provides a measure of water content, hemoglobin saturation with oxygen, and possible concentration of pharmaceutical products in blood and tissues. These characteristics can be used in the diagnosis of various diseases as for example cancer or stroke, which are known to alter such physiological values [123]. Since living tissues do not contain strong intrinsic chromophores that absorb radiation in the NIR band [144], DOI applications take advantage of this wavelength window which allows deeper penetration. For a localized absorber, the absorption cross section σ_a (cm^2) - independent of the relative orientation of the impinging light and the absorber- can be defined as $\sigma_a = Q_a A_g$, where A_g is the geometrical cross section of the particle and Q_a an efficiency correction factor. One can imagine that the localized absorber "blocks incident light and casts a shadow", which constitutes absorption, see Figure 1b). The absorption cross section corresponds then to the ratio

$$\sigma_a = \frac{P_{abs}}{I_0}, \tag{1}$$

where P_{abs} (W) is the amount of power absorbed out of an initially uniform plane wave of intensity I_0 ($\mathrm{W/cm}^2$) which invests the particle. A medium with a distribution of identical absorbing particles with density $\rho_a = \rho_a(\mathbf{r})$ ($\#/\mathrm{cm}^3$) can be characterized by the absorption coefficient μ_a (cm^{-1}) given by

$$\mu_a(\mathbf{r}) = \rho_a(\mathbf{r})\sigma_a. \tag{2}$$

Notice that the probability per infinitesimal path length that a photon will be absorbed by the tissue is given by $1/\mu_a$. A transmission measurement through a pathlength of an absorbing medium of thickness L_t (cm) with light intensity I_0 entering the sample and light intensity I leaving the sample is modeled by the Beer-Lambert law, which linearly correlates absorbance and concentration of an absorbing species as

$$I = I_0 \mathrm{e}^{-\mu_a L_t}. \tag{3}$$

Historically, the definition of μ_a was based on spectrometers that reported transmission $T = I/I_0 = 10^{-\varepsilon_\lambda C L_t}$, where ε_λ ($\mathrm{cm}^2\ \mathrm{mol}^{-2}$) is the extinction coefficient, which is a measure of the "absorbing power" of the species at wavelength λ, and C is the molar concentration ($\mathrm{mol\ cm}^{-3}$) of the absorption species. Therefore, $\mu_a = \varepsilon C \log(10)$. In a complex medium, where several chromophores contribute to absorption, the absorption coefficient can be expressed as

$$\mu_a = \sum_i \varepsilon_{i,\lambda} C_i, \tag{4}$$

where $\varepsilon_{i,\lambda}$ and C_i are the extinction coefficient and concentration of the i-th chromophore, respectively. At NIR wavelengths the main absorbing chromophores in biological tissues are haemoglobin (in its oxy–genated or deoxy–genated forms), and to a lesser extent water, lipids, melanin, myoglobin, and cytochromes. In this setting, one can write

$$\mu_a = B(Sat \cdot \mu_{a,oxy} + (1 - Sat) \cdot \mu_{a,deoxy}) + W\mu_{a,water} + \sum_i \varepsilon_{i,NIR} C_i, \tag{5}$$

where Sat is blood oxygen saturation, B is blood volume fraction in the tissue, W water content in the tissue and the sum accounts for other minor absorbers.

2.3. Scattering

Scattering originates from the interaction of photons with structural heterogeneities present inside the material, mainly cellular organelles such as mitochondria, thin fibrillar structures of connective tissues, melanin granules, and red blood cells and from the non-uniform temporal/spatial distribution of refractive index in the medium. The scattering interaction between a photon and a molecule results in a photon moving in a different direction. When scattering is elastic, the photon emerges at the same frequency of the incident beam and without extraction of energy. Elastic scattering represents the most significant scattering events in the NIR windows, unelastic scattering such as fluorescence, phosphorescence, and Raman scattering being negligible. Particles with a characteristic dimension similar to or greater than the incident wavelength, like cell nuclei, mainly scatter in a forward direction (Mie scattering), while structures with a characteristic dimension smaller than the incident wavelength, like the cytoskeleton and mitochondrial membranes, have a more isotropic scattering pattern (Rayleigh scattering or also small-scale limit of Mie scattering). The presence in biological tissues of both large and small structures causes both Mie and Rayleigh scattering events. Moreover, the presence of non–spherical scattering bodies further complicates the modeling problem. Analogously to (1) - but in a merely a convenient way to quantify the scattering strength without a geometrical significance - one can define a scattering cross section σ_s (cm^2) as $\sigma_s = Q_s A_g$, where Q_s is the scattering efficiency and A_g the geometrical area of the scatterer, see Figure 1c). The scattering cross section corresponds then to the ratio

$$\sigma_s(\widehat{\mathbf{s}}) = \frac{P_{scatt}}{I_0}, \tag{6}$$

where P_{scatt} (W) is the amount of power which is spatially redirected (scattered) along the outgoing direction $\widehat{\mathbf{s}}$. The angular distribution of the scattered radiation is given by the differential cross section $\frac{d\sigma_s}{d\widehat{\mathbf{s}}}(\widehat{\mathbf{s}}', \widehat{\mathbf{s}})$, where the outgoing direction $\widehat{\mathbf{s}}$ defines the axis of a cone of solid angle $d\widehat{\mathbf{s}}$ originating at the scatterer, see Figure 2. Notice that in the definition (6) we have implicitly assumed that the scattering cross section is independent of the relative orientation $\widehat{\mathbf{s}}'$ of the incident light and the scatterer. A medium containing a uniform distribution of identical scatterers is characterized by the scattering coefficient

$$\mu_s(\mathbf{r}) = \rho_s(\mathbf{r})\sigma_s, \tag{7}$$

where ρ_s ($\#/\mathrm{cm}^3$) is the number density of scatterers. The scattering mean free path $l_s = 1/\mu_s$ represents the average distance a photon travels between consecutive scattering events.

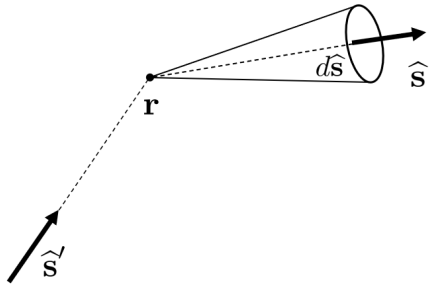

Figure 2. Notation for a single scattering event: light incides at position $\mathbf{r}$ from direction $\widehat{\mathbf{s}}'$ and is scattered in direction $\widehat{\mathbf{s}}$ around a cone of angle $d\widehat{\mathbf{s}}'$.

3. Physics–driven approaches

In this section we deal with approaches known as physics-driven (or *knowledge-driven*) modeling because the reconstruction procedure utilizes implicitly or explicitly the knowledge of mathematical models of the implied physical phenomena. Optical imaging modalities involve physical phenomena which can be modeled by partial differential equations (PDEs), whose solutions depend on unknown parameters representing the coefficients of the PDEs themselves. The crucial step in the reconstruction of the unknowns is thus generally performed by a PDE-based map and the related optimization problem combines the boundary measurements with the solution of the PDE modeling the problem. Let Ω be the bounded domain in $\mathbb{R}^d$, $d = 2, 3$, with boundary $\partial\Omega$, representing the sample under investigation. We denote by $q \in Q$ the set of spatially dependent optical parameters and by $u \in U$ the state variable corresponding to the light distribution field in the body upon application of a light source. Here Q and U are generic Banach spaces. We consider the PDE problem

$$\begin{aligned} &\mathcal{L}(u, q) = f \quad \text{in } \Omega, \\ &\mathcal{B}(u) = r \qquad \text{on } \partial\Omega, \end{aligned} \tag{8}$$

where $\mathcal{L} : D(\mathcal{L}) \subseteq U \times Q \to Z$ is a differential operator expressed by a polynomial in the derivatives of the state variable $u \in U$ and depending on parameters $q \in Q$, and f is a given source term. The differential operator is augmented by generic (Dirichlet, Neumann or Robin) boundary conditions, represented by the operator $\mathcal{B}$, in order to guarantee the well-posedness of problem (8) in appropriate functional spaces. Under suitable assumptions on the Fréchet differentiability of the operator $\mathcal{L}$, see for example [99], one can introduce the parameter-to-state map $S : D(S) \subseteq Q \to U$, *i.e.* $S(q) = u$, which is a solution to the boundary value problem (8). We then define the forward map as $\mathcal{F} = P \circ S$, P being the observation operator, providing the measured data on the boundary of the domain. Hence, the inverse problem of the identification of the set of parameters q, subject to (8), can be equivalently rewritten as the operator equation

$$\mathcal{F}(q) = y_q \approx y_\delta, \tag{9}$$

where y_δ is the noisy measurement vector for some noise level $\delta \geq 0$ obtained via the measurement map

$$y_\delta = \mathcal{M}(q; u). \tag{10}$$

Here $\mathcal{M} : D(\mathcal{M}) \subset (Q \times U) \to Y$ is a possibly nonlinear map and Y is a Banach space.

In the following of this section we provide a brief review of the derivation of models of type (8) for DOI applications.

3.1. Mathematical models of light-tissue interaction: Radiative Transfer Model

The knowledge of light distribution during irradiation in a tissue with known optical properties is a fundamental point in DOI applications. Classical Maxwell theory considers light to be an oscillating electromagnetic field. However, the multiple scattering events and associated decoherence effects in turbid media effectively suppress the wave nature of light; therefore, instead of tracking light waves, one can track only the average energy they contain. The flow of light energy through the medium is described by radiation transport (RT) theory, a continuum–based model which explicitly disregards wave interference effect [92]. Let $N(\mathbf{r}, \widehat{\mathbf{s}}, t)$ ($\mathrm{cm}^{-3}\mathrm{sr}^{-1}$) be the photon density, that is the number of photons per unit volume at a position $\mathbf{r}$, propagating in direction $\widehat{\mathbf{s}}$ within a solid angle $d\widehat{\mathbf{s}}$ at time t. The photon density can be connected to the scalar field energy radiance $L(\mathbf{r}, \widehat{\mathbf{s}}, t)$ as $L = E_{phot} c N$ ($\mathrm{Wcm}^{-2}\mathrm{sr}^{-1}$), E_{phot} being the quantized photon energy. The RTE is an energy balance equation which considers a small volume of tissue V with absorbing and scattering centers uniformly distributed inside. The change in energy in the volume within the solid angle element per unit time is given by

$$\frac{1}{c}\frac{\partial}{\partial t}\int_V L\, dV = \underbrace{-\int_V \nabla L \cdot \widehat{\mathbf{s}}\, dV}_{\text{transport term}} \underbrace{-\int_V (\mu_a + \mu_s) L\, dV}_{\text{extinction}} + \underbrace{\int_V \mu_s \int_{S^{d-1}} p_s(\widehat{\mathbf{s}}, \widehat{\mathbf{s}}\,') L d\widehat{\mathbf{s}}\,' dV}_{\text{cross-scattering}} + \underbrace{\int_V f\, dV}_{\text{internal source}}, \tag{11}$$

where μ_a and μ_s are the optical coefficients introduced in Section 2 and p_s [sr^{-1}] is the scattering phase function (see below). Moreover, "extinction" represents energy loss due to absorption and scattering from propagation direction $\widehat{\mathbf{s}}$ to direction $\widehat{\mathbf{s}}\,'$, and cross-scattering describes the energy gain due to scattering from any direction $\widehat{\mathbf{s}}\,'$ and scattered into $d\widehat{\mathbf{s}}$ around direction $\widehat{\mathbf{s}}$ per unit time. Function f stands for a source term in the medium. Relation (11) is an equation of integro-differential type.

3.1.1. Scattering phase function The normalized scattering phase function $p_s(\widehat{\mathbf{s}}, \widehat{\mathbf{s}}\,')$ represents the probability density function for scattering from an incident direction $\widehat{\mathbf{s}}'$ into a new direction $\widehat{\mathbf{s}}$ within the unit solid angle $d\widehat{\mathbf{s}}$ (see Figure 2). Hence

$$p_s(\widehat{\mathbf{s}}, \widehat{\mathbf{s}}\,') \geq 0, \qquad \forall\, \widehat{\mathbf{s}}, \widehat{\mathbf{s}}\,' \in S^{d-1},$$

where $S^{d-1} = \{s : \ s = (s_1, \ldots, s_d), \ |s| = 1\}$ is the unit sphere in $\mathbb{R}^d$. The phase function has to satisfy some mathematical properties, reflecting the physical properties of the light and the scattering process:

- it is standard the assumption that the probability of scattering over S^{d-1} is equal to one, regardless of the incoming radiation direction, that is

$$\int_{S^{d-1}} p_s(\widehat{\mathbf{s}}, \widehat{\mathbf{s}}\,') \, d\widehat{\mathbf{s}}' = 1, \qquad \forall \widehat{\mathbf{s}} \in S^{d-1}, \tag{12}$$

- the phase function satisfies the so-called "reciprocity of light propagation" property which corresponds mathematically to require that

$$p_s(\widehat{\mathbf{s}}, \widehat{\mathbf{s}}\,') = p_s(-\widehat{\mathbf{s}}', -\widehat{\mathbf{s}}).$$

In case when the scattering is independent on the absolute direction of the incident light, the phase function fulfills the further relation

$$p_s(\widehat{\mathbf{s}}, \widehat{\mathbf{s}}\,') = p_s(-\widehat{\mathbf{s}}, -\widehat{\mathbf{s}}\,'),$$

which corresponds to an isotropy condition. More precisely, the scattering phase function depends only on the angle between the incident $\widehat{\mathbf{s}}$ and scattered directions $\widehat{\mathbf{s}}\,'$, hence is a function of only $\widehat{\mathbf{s}}' \cdot \widehat{\mathbf{s}}$. In this context, a constant of interest is the cosine-weighted average of the scattering (expectation value of the cosine of the scattering angle or average cosine of scatter), that is

$$g = \int_{S^{d-1}} p_s(\widehat{\mathbf{s}} \cdot \widehat{\mathbf{s}}')\widehat{\mathbf{s}} \cdot \widehat{\mathbf{s}}' \, d\widehat{\mathbf{s}}', \qquad g \in [-1, 1]. \tag{13}$$

This is a measure of the forward scattering bias and quantifies how efficiently photons keep propagating in the forward direction despite scatter [85]. When $g = 0$ one has fully isotropic scattering, while when $g \to 1$ the propagation is strongly forward biased, that is scattering vanishes, and for $g \to -1$ scattering becomes completely backward directed. For in vitro tissues at the visible and NIR wavelengths one has typically $0.65 < g < 0.95$ [93].

A convenient approximate model for $p_s(\widehat{\mathbf{s}} \cdot \widehat{\mathbf{s}}')$ is derived from the so-called *Henyey-Greenstein* function $H_d(\widehat{\mathbf{s}} \cdot \widehat{\mathbf{s}}'; g)$, which has different expressions depending on the space dimension d. For example, when $d = 3$, we have that

$$H_3(\widehat{\mathbf{s}} \cdot \widehat{\mathbf{s}}'; g) = \frac{1}{4\pi} \frac{1 - g^2}{(1 + g^2 - 2g\widehat{\mathbf{s}} \cdot \widehat{\mathbf{s}}\,')^{3/2}}.$$

Making use of Legendre polynomials $P_l, l = 1, \ldots, \infty$, we can represent the function H_3 as a series, i.e.,

$$H_3(\widehat{\mathbf{s}} \cdot \widehat{\mathbf{s}}'; g) = \sum_{l=0}^{\infty} \frac{2l+1}{4\pi} h_l P_l(\widehat{\mathbf{s}} \cdot \widehat{\mathbf{s}}'), \tag{14}$$

where the coefficient $(2l + 1)/4\pi$ is added only for numerical convenience. By straightforward calculations one finds the useful property that the coefficients of the expansion (14) are the powers of the parameter g, that is

$$h_l = \int_{S^2} P_l(\tau) H_3(\tau) \, d\tau = g^l, \qquad \text{for } \ l \geq 0.$$

In the two-dimensional setting, given that $\widehat{\mathbf{s}} = \varphi$ and $\widehat{\mathbf{s}}' = \varphi'$, where φ, φ' are two arbitrary angles, the Henyey-Greenstein function has the following expression

$$H_2(\varphi, \varphi'; g) = \frac{1}{2\pi} \frac{1 - g^2}{1 + g^2 - 2g\cos(\varphi - \varphi')},$$

which can be represented as a Fourier series (similarly to (14)), see [85].

We mention that the Henyey-Greenstein function is not only the possible choice for the phase function. In the literature, in fact, the modified Henyey-Greenstein and delta-Eddington functions have also been postulated to represent single scattering phase functions for tissue. We refer the reader to [87] for more details.

Assuming sufficient regularity one can write the differential formulation of (11) as

$$\left(\frac{1}{c}\frac{\partial}{\partial t} + \widehat{\mathbf{s}} \cdot \nabla + (\mu_a + \mu_s)(\mathbf{r})\right) L(\mathbf{r}, \widehat{\mathbf{s}}, t) - Q(L)(\mathbf{r}, \widehat{\mathbf{s}}, t) = 0 \quad \text{in}\, \Omega \times S^{d-1} \times [0, \infty), \quad (15)$$

where

$$Q(L)(\mathbf{r}, \widehat{\mathbf{s}}, t) = -\mu_s(\mathbf{r}) \int_{S^{d-1}} p_s(\widehat{\mathbf{s}}' \cdot \widehat{\mathbf{s}}) L(\mathbf{r}, \widehat{\mathbf{s}}, t)\, d\widehat{\mathbf{s}}' - f(\mathbf{r}, \widehat{\mathbf{s}}, t). \quad (16)$$

Applying the Fourier transform with respect to the time variable in (15)-(16), the previous equation is written in the frequency domain as

$$\left(\frac{i\omega}{c} + \widehat{\mathbf{s}} \cdot \nabla + (\mu_a + \mu_s)(\mathbf{r})\right) L(\mathbf{r}, \widehat{\mathbf{s}}) - Q(L)(\mathbf{r}, \widehat{\mathbf{s}}) = 0, \quad \text{in} \;\; \Omega \times S^{d-1}, \quad (17)$$

where for simplicity we omit the dependence of $L(\mathbf{r}, \widehat{\mathbf{s}})$ from $\omega \in [0, +\infty)$. We define the boundary sets

$$\Gamma_{\pm} = \{(\mathbf{r}, \widehat{\mathbf{s}}) \in \partial\Omega \times S^{d-1} \text{ such that } \quad \pm\, \widehat{\mathbf{s}} \cdot \mathbf{n} > 0\},$$

where $\mathbf{n}$ is the outer unit normal vector to the domain at $\mathbf{r} \in \partial\Omega$. Adding to (17) boundary conditions on Γ_-, we get the following boundary value problem

$$\left(\frac{i\omega}{c} + \widehat{\mathbf{s}} \cdot \nabla + (\mu_a + \mu_s)(\mathbf{r})\right) L(\mathbf{r}, \widehat{\mathbf{s}}) - Q(L)(\mathbf{r}, \widehat{\mathbf{s}}) = 0, \qquad \text{in} \;\; \Omega \times S^{d-1}, \quad (18)$$

$$L(\mathbf{r}, \widehat{\mathbf{s}}) = g_L(\mathbf{r}, \widehat{\mathbf{s}}), \quad \text{on} \;\; \Gamma_-, \quad (19)$$

where g_L represents the flux of particles incident on $\partial\Omega$ at the point $\mathbf{r} \in \partial\Omega$. In order to provide a mathematical characterization, we set

$$\mathcal{Q}_{ad} = \{(\mu_a, \mu_s) \,:\, \mu_a, \mu_s \geq 0,\ (\mu_a, \mu_s) \in L^{\infty}(\Omega) \times L^{\infty}(\Omega)\}\,, \quad (20)$$

which states that the absorption and the scattering coefficients μ_a, μ_s are nonnegative and bounded functions. For the phase function, we require that $p_s \in L^1(S^{d-1})$. We also introduce the functional spaces

$$L^2_{\widehat{\mathbf{s}}\cdot\mathbf{n}}(\Gamma_{\pm}) := \left\{\eta(\mathbf{r}, \widehat{\mathbf{s}}) \;:\; \int_{\Gamma_{\pm}} \eta^2(\mathbf{r}, \widehat{\mathbf{s}}) |\widehat{\mathbf{s}} \cdot \mathbf{n}(\mathbf{r})|\, d\sigma(\mathbf{r}) d\widehat{\mathbf{s}} < +\infty\right\},$$

$$\mathcal{W}^2(\Omega \times S^{d-1}) := \left\{\eta(\mathbf{r}, \widehat{\mathbf{s}}) \;:\; \eta \in L^2(\Omega \times S^{d-1}),\ \ \widehat{\mathbf{s}} \cdot \nabla\eta \in L^2(\Omega \times S^{d-1})\right\}.$$

In this functional setting, the well-posedness of the boundary value problem (18)-(19) has been proved, see for example [4, 121].

Proposition 3.1 *Assume that* $(\mu_a, \mu_s) \in \mathcal{Q}_{ad}$, ω *is finite,* $g_L \in L^2_{\widehat{\mathbf{s}}\cdot\mathbf{n}}(\Gamma_-)$, *and* $f(\mathbf{r}, \widehat{\mathbf{s}}) \in L^2(\Omega \times S^{d-1})$. *Then the forward problem (18)-(19) is well-posed and admits a unique solution* $L(\mathbf{r}, \widehat{\mathbf{s}}) \in \mathcal{W}^2(\Omega \times S^{d-1})$.

As in [121], we can then define the following operator, called the albedo operator, which maps the incoming flux on the boundary Γ_- into the outgoing flux on Γ_+,

$$M : g_L \in L^2_{\widehat{\mathbf{s}}\cdot\mathbf{n}}(\Gamma_-) \rightarrow L|_{\Gamma_+} \in L^2_{\widehat{\mathbf{s}}\cdot\mathbf{n}}(\Gamma_+),$$

which depends on the optical parameters μ_a and μ_s. We point out that in practice only measurements of the outgoing currents, which are averages of the outgoing flux, are available [53, 121].

Diffusion approximation. The RTE model is of integro-differential type and requires heavy computations to get accurate solutions. To overcome this issue, typically, an approximation of the RTE equation is considered, based on the assumption that the photons propagate diffusively through the tissue [95]. A standard approximation method for the RTE is the P_N approximation, see e.g. [46], which consists of expanding the radiance, the phase function in (16), and the source term via spherical harmonics, when $d = 3$, and via Fourier series, in the case $d = 2$ (see for example [18, 85]). If the series is truncated at order N, we get the so-called P_N approximation. The P_1 case corresponds to the Diffusion Approximation (DA), as explained, for example, in [18, 21, 19]. Since the P_N approximation is well-known in the literature [18] and in the sequel we work with the P_1 approximation only, we summarize here the principal step for getting the DA framework by a projection argument, as in [86]. With this target, given $\widehat{\mathbf{s}} = (\widehat{s}_1, \ldots, \widehat{s}_d)$, we introduce the spanning set

$$A_1 = \text{span}\{1, x_i, \quad 1 \leq i \leq d\} \subset L^2(S^{d-1}),$$

which is composed of polynomial of degree less or equal than one. Defining the projection operator $\Pi : L^2(S^{d-1}) \rightarrow A_1$, and the *photon fluence* [W cm^{-2}] and the *photon current* [W cm^{-1}] which are given by

$$u(\mathbf{r}, t) = \int_{S^{d-1}} L(\mathbf{r}, \widehat{\mathbf{s}}', t) d\widehat{\mathbf{s}}', \tag{21}$$

$$\mathbf{F}(\mathbf{r}, t) = \int_{S^{d-1}} \widehat{\mathbf{s}}' L(\mathbf{r}, \widehat{\mathbf{s}}', t) d\widehat{\mathbf{s}}', \tag{22}$$

then the P_1-approximation of the radiance corresponds to the projection of the radiance in A_1, that is,

$$L(\mathbf{r}, \widehat{\mathbf{s}}, t) \approx \Pi L(\mathbf{r}, \widehat{\mathbf{s}}, t) = \frac{1}{|S^{d-1}|} u(\mathbf{r}, t) + \frac{d}{|S^{d-1}|} \mathbf{F}(\mathbf{r}, t) \cdot \widehat{\mathbf{s}}, \tag{23}$$

thanks to variational arguments. DA is obtained by integrating over S^{d-1} the RTE (15) and the RTE (15) multiplied by $\widehat{\mathbf{s}}$, and then using (12) and the first order approximations for the radiance (23), both for the phase function p_s and for the source term f. After

lengthy calculations, we get two coupled equations in terms of the photon density and current

$$\frac{1}{c}\frac{\partial u}{\partial t} + \mu_a u + \nabla \cdot \mathbf{F} = f_0, \tag{24}$$

$$\frac{1}{c}\frac{\partial \mathbf{F}}{\partial t} + (\mu_a + \mu_s')\mathbf{F} + \frac{1}{d}\nabla u = \mathbf{f}_1, \tag{25}$$

where μ_s' is the so-called *effective or reduced scattering coefficient* defined as

$$\mu_s' = \mu_s(1-g), \tag{26}$$

and

$$f_0(\mathbf{r},t) = \int_{S^{d-1}} f(\mathbf{r},\widehat{\mathbf{s}},t)\, d\widehat{\mathbf{s}}, \qquad \mathbf{f}_1(\mathbf{r},t) = \int_{S^{d-1}} \widehat{\mathbf{s}}\, f(\mathbf{r},\widehat{\mathbf{s}},t)\, d\widehat{\mathbf{s}}.$$

In order to get a single second order partial differential equation, we consider other two possible simplifications, assuming that

(i) only isotropic sources exist, i.e. the linearly anisotropic source term $\mathbf{f}_1 = \mathbf{0}$;

(ii) the change with respect to the time of the flux is assumed to be negligible (or, alternatively, proportional to $\mathbf{F}$), $\partial\mathbf{F}/\partial t \approx 0$.

The previous two assumptions hold when the medium is scattering dominated, which corresponds to the physical condition $\mu_s \gg \mu_a$, that is the diffusion approximation (DA) to the RTE is valid in the regime where the scattering length is small compared to the distance of propagation. Moreover, the phase function is independent of the absolute angle and the photon flux changes slowly and that all sources are isotropic. A detailed description of reduction process is given in [21]. With the two assumptions stated above, equation (25) reduces to Fick's law

$$\mathbf{F} = \frac{1}{d(\mu_a + \mu_s')}\nabla u = -D\nabla u, \tag{27}$$

where

$$D = \frac{1}{d(\mu_a + \mu_s')} \sim \frac{1}{d\mu_s'} \tag{28}$$

is the diffusion coefficient. Inserting the expression (27) for $\mathbf{F}$ into the equation (24), we finally get the diffusion approximation of the RTE, that is

$$\frac{1}{c}\frac{\partial}{\partial t}u(\mathbf{r},t) - \nabla \cdot (D(\mathbf{r})\nabla u(\mathbf{r},t)) + \mu_a(\mathbf{r})u(\mathbf{r},t) = f_0(\mathbf{r},t). \tag{29}$$

In applications like the SSD-DOT, one can assume that the distribution of light inside the domain of interest is instantaneous. Mathematically, this corresponds to reducing the problem to a stationary equation, hence assuming that $\frac{1}{c}\frac{\partial}{\partial t}u(\mathbf{r},t) \simeq 0$. Then, with this further simplification, (29) reduces to following elliptic equation

$$-\nabla \cdot (D(\mathbf{r})\nabla u(\mathbf{r})) + \mu_a(\mathbf{r})u(\mathbf{r}) = f_0(\mathbf{r}). \tag{30}$$

To get a well-posed problem, (30) requires the specification of suitable boundary conditions. The simplest conditions one can consider are homogeneous Dirichlet

boundary conditions but they are physically incorrect [19, 124]. In fact, in applications (30) is often supplemented with a Robin-type boundary condition, also known as the partial current boundary condition [18, 19, 124, 127]. Briefly, we introduce the total inward- and outward-directed photon fluxes at a point $\mathbf{r} \in \partial\Omega$ as follows

$$J_-(\mathbf{r}) := -\int_{\widehat{\mathbf{s}}\cdot\mathbf{n}(\mathbf{r})<0} \widehat{\mathbf{s}}\cdot\mathbf{n}(\mathbf{r}) L(\mathbf{r},\widehat{\mathbf{s}})\, d\widehat{\mathbf{s}}, \tag{31}$$

$$J_+(\mathbf{r}) := \int_{\widehat{\mathbf{s}}\cdot\mathbf{n}(\mathbf{r})>0} \widehat{\mathbf{s}}\cdot\mathbf{n}(\mathbf{r}) L(\mathbf{r},\widehat{\mathbf{s}})\, d\widehat{\mathbf{s}}. \tag{32}$$

It is straightforward to show that, thanks to the P_1 approximation (23),

$$J_-(\mathbf{r}) = c_d u(\mathbf{r}) - \frac{1}{2}\mathbf{n}\cdot\mathbf{F}(\mathbf{r}), \qquad J_+(\mathbf{r}) = c_d u(\mathbf{r}) + \frac{1}{2}\mathbf{n}\cdot\mathbf{F}(\mathbf{r}), \tag{33}$$

where c_d is a constant depending on the space dimension ($c_d = 1/\pi$ if $d = 2$, and $c_d = 1/4$ if $d = 3$). In the simplest setting, one can assume that the total inward-directed photon flux on $\partial\Omega$ is zero, i.e. $J_-(\mathbf{r}) = 0$, $\mathbf{r} \in \partial\Omega$. The use of the Fick's law (27) in the first equation of (33), with the assumption that $J_-(\mathbf{r}) = 0$ for $\mathbf{r} \in \partial\Omega$, yields the Robin boundary condition

$$u(\mathbf{r}) + \frac{D}{2c_d}\frac{\partial u(\mathbf{r})}{\partial \mathbf{n}} = 0, \qquad \mathbf{r} \in \partial\Omega. \tag{34}$$

This represents the physical situation of a nonscattering medium surrounding Ω, i.e. no diffuse surface reflection on $\partial\Omega$ occurs [127]. However, in practical situations, it is mandatory to consider the case where the object has a different refractive index to the surrounding medium, as depicted in Figure 1a). In fact, in this circumstance, it could happen that a large portion of the radiant energy is reflected back into the object. Therefore, the feasible relation to consider between the fluxes is

$$J_-(\mathbf{r}) = R\, J_+(\mathbf{r}), \qquad \mathbf{r} \in \partial\Omega, \tag{35}$$

where $0 \le R \le 1$ is the reflection coefficient on the boundary $\partial\Omega$. Note that, from the previous relation, we recover the special case of no boundary reflection (34) when $R = 0$. There are different options to calculate the parameter R, see for more details [127]. Usually it is derived from the Fresnel's law [127], that is, denoting by ν_1 and ν_2 the refractive indices of the medium and of the external part respectively, the reflection coefficient R can be also derived, when $\nu_2 = 1$, by an experimental fit [1], that is

$$R \simeq -1.4399\nu_1^{-2} + 0.7099\nu_1^{-1} + 0.6681 + 0.0636\nu_1.$$

From the equality (35) and using again the Fick's law (27) in both the equations of (33), we eventually obtain the following Robin boundary condition

$$u(\mathbf{r}) + \frac{\zeta D(\mathbf{r})}{2c_d}\frac{\partial u(\mathbf{r})}{\partial \mathbf{n}} = 0, \qquad \mathbf{r} \in \partial\Omega, \tag{36}$$

where $\zeta = (1+R)/(1-R)$. Summarizing, the DA model is reduced to the study of the following boundary value problem

$$-\nabla\cdot(D\nabla u) + \mu_a u = f_0, \qquad \text{in } \Omega, \tag{37}$$

$$u + \frac{\zeta D}{2c_d}\frac{\partial u}{\partial \mathbf{n}} = 0, \qquad \text{on } \partial\Omega. \tag{38}$$

In some technological arrangements, part of the boundary $\partial\Omega$ of the object under investigation may be supported on a solid plate containing a certain number of light sources (see, e.g., [48]). Therefore, different boundary conditions have to be chosen on the contact surface between the medium and the plate [31, 30, 19, 124] or the light sources on the boundary of the domain can be approximated as source terms in the equation (37), see for example [19, 86, 115]. In fact, in the DA approximation, light sources can be modeled by using the so-called *collimated model*: the source (assuming that only one boundary source is activated) is modeled as an isotropic internal point source localized to within at a depth $\mathbf{r}_s = 1/\mu_s'$ below the object surface

$$f_0(\mathbf{r}) = Q_0\,\delta(\mathbf{r} - \mathbf{r}_s),$$

Q_0 begin an intensity constant. Therefore, in this specific model, we have a non zero source term f_0 in the equation (30), an homogeneous Dirichlet boundary condition for the fluence u on the portion of the boundary corresponding to the supported part, and homogeneous Robin conditions on the non-supported boundary part. This model, which we will refer to in the following as DOT-DA, provides good results at distances from the source larger than the mean free path, but less accurate results close to the source [127]. For a better description of the light intensity in a neighborhood of source terms, a Monte Carlo method was introduced in [143]. We refer the reader to [127] for other possible configurations to model the light source incident at a point on the boundary of the domain which corresponds to the case of a non-homogeneous Robin condition on a portion of the boundary.

3.1.2. Measurements: The exitance. In DOT the measurable quantity on the boundary of the domain Ω is the so-called *exitance*, that is the energy transfer through a unit area with normal vector $\mathbf{n}$, mathematically defined as

$$J_{ex}(\mathbf{r}) := \int_{S^{d-1}} \widehat{\mathbf{s}} \cdot \mathbf{n}(\mathbf{r}) L(\mathbf{r}, \mathbf{s})\, d\widehat{\mathbf{s}}, \qquad \mathbf{r} \in \partial\Omega. \tag{39}$$

Using definitions (31), the corresponding P_1 approximation (33), the Fick's law and the Robin boundary conditions (36), the exitance (39) is equal to

$$J_{ex}(\mathbf{r}) = J_+(\mathbf{r}) - J_-(\mathbf{r}) = \mathbf{n}(\mathbf{r}) \cdot F(\mathbf{r}) = -D\frac{\partial u(\mathbf{r})}{\partial \mathbf{n}} = \frac{2c_d}{\zeta} u(\mathbf{r}), \qquad \mathbf{r} \in \partial\Omega. \tag{40}$$

3.1.3. The inverse problem and known results. The general inverse problem for the DOT is to determine the coefficients D, μ_a from the knowledge of all possible pairs of Neumann and Dirichlet boundary values, that is, from the use of the Neumann-to-Dirichlet map, thanks to the fact that we are assigning Robin boundary conditions, see (38), and measuring the exitance (40). It has been shown by Arridge and Lionheart [20] that this inverse problem is in general not uniquely solvable, in the sense that it is not possible to determine uniquely the pair D, μ_a from infinitely many pairs of Neumann and Dirichlet boundary values. In fact, in the case where D is sufficiently regular and strictly positive, the DOT-DA equation

$$-\nabla \cdot (D\nabla u) + \mu_a u = 0, \qquad \text{in } \Omega \tag{41}$$

can be transformed into a Schrödinger type equation, by setting $v = \sqrt{D}u$, as in [132], that is

$$-\Delta v + \mu v = 0, \qquad \text{in } \Omega, \qquad\qquad \mu := \frac{\Delta\sqrt{D}}{\sqrt{D}} + \frac{\mu_a}{D}, \tag{42}$$

where $\Delta\sqrt{D}$ is the laplacian of $\sqrt{D}$. If, for example, $D = 1$ in a neighborhood of $\partial\Omega$, then the Cauchy data of v coincides with those of u ($v = u$ and $\partial v/\partial\mathbf{n} = \partial u/\partial\mathbf{n}$ on $\partial\Omega$). This means that under suitable regularity assumptions on the coefficients D and μ_a, hence on μ, we can recover μ only, see [91, 139], that is we cannot extract a unique value for D and μ_a [20, 82]. Non-uniqueness results have been also proven in presence of a source term in (41), see [20].
A uniqueness result can be obtained requiring suitable regularity assumptions on the coefficients, as in the case of a piecewise constant diffusion and a piecewise analytic absorption coefficients [82, 83].

Recent lines of research concern the case of nonlocal models [152] and of anisotropic media [65]. An instance in the latter direction is given by [69] where the case of an anisotropic diffuse tensor, which is assumed a real matrix-valued function, has been studied. It is shown that μ_a and its derivatives at the boundary of Ω can be determined in a stable way by boundary data with a modulus of continuity of Lipschitz and Hölder type respectively, assuming that the scattering coefficient μ_s is known. These results have been extended to the anisotropic time-harmonic case, that is when the diffusion tensor and the absorption coefficients are complex-valued tensors and functions respectively. Precisely in [58] a Lipschitz stability result of the boundary value of μ_a in terms of boundary data, when μ_s is again assumed known, was established. The Hölder stability of the derivatives of the absorption coefficient on the boundary of the domain has been established in [55]. The same results hold in the case where μ_a is known and μ_s has to be determined [55, 58, 69]. The stable determination of μ_a (or equivalently μ_s) and its derivatives at the boundary are useful tools to infer uniqueness and stability of μ_a (or μ_s) in the interior of Ω [69]. We finally also mention [66] for the determination of an anisotropic inclusion in a Schrödinger type equation from local Cauchy data.

We give here some references on related problems which are intimately connected with those presented above and equations (41), (42). We previously mentioned that the coefficient μ in (42) can be uniquely reconstructed from Cauchy data. This question has mainly been studied in the context of the Calderon's problem [42], corresponding to the case where μ_a and f_0 in (41) are zero, that is in the case of the equation $\operatorname{div}(D\nabla u) = 0$. For seminal contributions in this field, when infinitely many pairs of Dirichlet and Neumann data are known on the entire boundary, we refer the reader to the following papers: Kohn-Vogelius [100, 101] for $d \geq 2$ and when D is assumed piecewise analytic, Sylvester and Uhlmann [132] for $d \geq 3$ and $D \in C^2(\overline{\Omega})$, Haberman-Tataru [74] and Caro-Rogers [44] for $D \in W^{1,\infty}(\Omega)$. When $d = 2$, we mention the results of Nachman [113] for $D \in W^{2,p}(\Omega)$ and of Astala-Päivärinta [23] when $D \in L^\infty(\Omega)$. The stability issue was first investigated by Alessandrini in [10] for $d \geq 3$ and $D \in W^{2,\infty}(\Omega)$. Recently Alberti-Santacesaria presented a general framework to study uniqueness, stability and

reconstruction for infinite-dimensional inverse problems, such as Calderon's problem, when only a finite-dimensional approximation of the measurements is available [7, 8, 9]. The determination of a L^∞ potential in the Schrödinger equation in three and higher dimensions when only a finite number of measurements is available and the diffusion depends on finitely many parameters has been studied in [7]. Harrach derived Lipschitz stability from a finite number of measurements for the complete electrode model in electrical impedance tomography in [84].

Equation (41) appears also in the analyis of hybrid inverse problems, which are a particular class of inverse problems using internal data coming from the use of coupled-physics phenomena [6]. For example, the DOT inverse problem is completely analogous to the one of quantitative photoacoustic tomography (qPAT), except for the fact that the photon current data in DOT are known and available only on the boundary of the domain, while in qPAT the available data, given by the absorbed energy, are known in the interior of the domain [122], see also [6, 13]. The literature on qPAT is generously vast. We refer, for example, the reader to [11, 24, 25, 26, 27, 28, 41, 52, 60, 114] for theoretical results related to uniqueness and stability issues in qPAT and to [5, 14, 22, 24, 36, 54, 61, 70, 76, 77, 118, 119, 122, 135] for possible reconstruction methods.

The results above show that the inverse problem of identifying both the diffusion and the absorption coefficients is ill-posed and nonlinear. In order to mitigate the ill-posed nature of the problem and get feasible numerical solutions, regularization methods must be adopted.

3.2. Reconstruction methods

In order to end up with an approximate solution of the problem, one has to introduce a discretization. The standard approach is to decompose the computational domain Ω into triangular/trapezoidal areas (2D) or into voxels of suitable shape (3D). While mesh generation is significantly easier today than in the past, meshing irregular objects with complex internal structure in an acceptable time may still be a challenge, as already observed in the more than 15 years old review by Gibson and Arridge [71]. The construction of the geometry may be benefit from the simultaneous collection of data with other techniques, as for example in the realistic test case presented in [45], where an anatomically coherent mesh of the patient's breast was created by manually segmenting 2D coronal MRI slices to generate 3D surface maps and then a volume mesh. Upon introducing a discretization of the domain, one then operates with functions belonging to finite dimensional subspaces parametrized by the mesh diameter h. For example, using a FEM approach, one approximates the solution u of the DOT-DA problem (41) by a piecewise polynomial function $u_h \in U_h$ given by

$$u_h = \sum_{j=1}^{N_h} u_j \phi_j(\mathbf{r})$$

where U_h is a finite dimensional subspace spanned by the basis functions $\phi_j, j = 1, \ldots, N_h$ and where the nodal values u_j are to be determined. Piecewise polynomial approximations to the continuous optical coefficients are constructed in the same way. Applying a Galerkin approach transforms the continuous problem of into an N_h-dimensional discrete problem of finding the nodal field values at all nodes, given the set of nodal parameters.

In the classical variational framework, one then rephrases the inverse problem (9) as the optimization procedure [125]: seek the optimal parameter set q^* such that

$$q^* = \arg\min_{q \in Q} \mathcal{D}(y_q, y_\delta), \tag{43}$$

where $\mathcal{D} : Y \to [0, \infty]$ is a loss function which typically contains a discrepancy measure (or data fidelity) term. A very common assumption considers the noise perturbing the data of Gaussian nature: this hypothesis naturally leads to choose $\mathcal{D}$ as the Least Square functional:

$$\mathcal{D}(y_q, y_\delta) := \frac{1}{2}\|y_q - y_\delta\|_2^2. \tag{44}$$

Notice that in applications like Time Domain DOT (TD-DOT) one could consider also a mixture of Poisson and Gaussian noise [64, 150]: in such a case, the fidelity functional may for example be of Kullback-Leibler type. Problem (43) with the choice (44) is nonlinear and several techniques have been used to obtain a linearized approximation. Supposing that the forward map $\mathcal{F}$ is Fréchet differentiable, a Taylor expansion around the reference point q_0 reads

$$\mathcal{F}(q) \approx \mathcal{F}(q_0) + \mathcal{F}'(q_0)\Delta q \tag{45}$$

where $\mathcal{F}'$ is the first order Fréchet derivative of the forward operator with respect to q.

Referring to SSD-DOT technologies, many source-detector pairs are typically spread out over the sample surface. Given N_s sources and N_d detectors (for a total of $M = N_s \times N_d$ pairs), the forward operator maps the vector of absorption (and diffusion) coefficients q_h at voxels within the sample to the vector of detector measurements for a given source. Let the index i stand for the detector/source pair, i.e. $i = 1, \ldots, M \to \{l, k\}$, $l = 1, \ldots, N_d$, $k = 1, \ldots, N_s$ and the index j for the voxel number, $j = 1, \ldots, N_v$. Thus, a common notation in the discrete setting for the first order Fréchet derivative is the Jacobian matrix $\mathbf{J}$ of size $M \times N_v$. Inserting (45) into (43) and truncating to the first order yield in the discrete space

$$\Delta\mathbf{q}^* = \arg\min_{\Delta\mathbf{q}} \frac{1}{2} \left\|\mathbf{J}\Delta\mathbf{q} - \mathbf{y}_\delta\right\|_2^2, \tag{46}$$

where $\Delta\mathbf{q}$ and $\mathbf{y}_\delta$ are the discrete counterparts of Δq and y_δ, respectively. The optimization problem (46) can be solved by one of the many available algorithms, that, however, generally imply a significant computational cost. With this respect, a technique which plays a very relevant role in this field is the so–called Rytov perturbation approach, which we discuss in detail here below.

3.3. Linear inversion by perturbation approach: the Rytov approximation

As observed above, DOT inverse problem is in general nonlinear, but in the limit that the optical properties are close to a specified background value, it can be considered approximately linear. This is usually the case for imaging differences in optical properties with respect to a reference condition. To obtain a linearization, one can approximate the change in the field when a change in the state occurs, making use of an expansion of the solution of the problem in series. There are two possible approaches: the Born and the Rytov approximations. In the Born approximation, $u(\mathbf{r})$ is a linear superposition of the incident (background, u_0) and scattered (heterogeneous, u_{sc}) diffusive waves such that $u = u_0 + u_{sc}$. In the Rytov approximation, the contributions from the incident and scattered parts are expressed in exponential fashion as

$$u(\mathbf{r}) = e^{\psi(\mathbf{r})}, \qquad \text{with} \qquad \psi(\mathbf{r}) \approx \sum_{i=0}^{N} \psi_i(\mathbf{r}), \qquad N \geq 1. \tag{47}$$

For details on Born approximation, we refer the reader to [18, 19, 124] and we consider here in detail the Rytov approximation. This choice is motivated by the fact that in experimental and clinical researches on optical tomography, quite often the Born approximation is impractical and tomographic images are obtained with the Rytov approximation [111]. In addition, the Rytov approximation tends to be more accurate as it accounts for some non-linear saturation due to increasing perturbation in the absorption coefficient [39]. We mention however the fact that both the linearized approaches have clear limitations, see for example [38], where - in addition - the Rytov approximation is considered to be inferior to the Born approximation.

We denote then by (D_0, μ_{a0}) the reference state and by u_0 the corresponding solution of (37)-(38) with $D = D_0$ and $\mu_a = \mu_{a0}$, i.e.

$$-\nabla \cdot (D_0 \nabla u_0) + \mu_{a0} u_0 = f_0, \qquad \text{in } \Omega, \tag{48}$$

$$u_0 + \frac{\zeta D_0}{2c_d} \frac{\partial u_0}{\partial \mathbf{n}} = 0, \qquad \text{on } \partial\Omega. \tag{49}$$

Let us consider a perturbation of the reference state inside the domain, that is $\widetilde{D} = D_0 + \delta D$ and $\widetilde{\mu_a} = \mu_{a0} + \delta\mu_a$ in Ω, with the further assumption that

$$\widetilde{D} = D_0 \qquad \text{on } \partial\Omega.$$

We denote by $\widetilde{u}$ the solution of (37)-(38) with $D = \widetilde{D}$ and $\mu = \widetilde{\mu_a}$, i.e.,

$$-\nabla \cdot \left(\widetilde{D} \nabla \widetilde{u}\right) + \widetilde{\mu_a} \widetilde{u} = f_0, \qquad \text{in } \Omega, \tag{50}$$

$$\widetilde{u} + \frac{\zeta D_0}{2c_d} \frac{\partial \widetilde{u}}{\partial \mathbf{n}} = 0, \qquad \text{on } \partial\Omega. \tag{51}$$

Choosing $N = 1$ in (47), we get that

$$\widetilde{u} = e^{\psi_0 + \psi_1}, \quad \text{with} \quad u_0 = e^{\psi_0}, \quad \text{hence} \quad \ln \widetilde{u} = \ln u_0 + \psi_1.$$

Using the fact that $u_0 = e^{\psi_0}$ in (48) and $\widetilde{u} = e^{\psi_0+\psi_1}$ in (50), subtracting the resulting equations and then using some differential identities we obtain

$$\begin{aligned}
&-\nabla\cdot(D_0\nabla(\psi_1 u_0)) + \mu_{a0}u_0\psi_1 \\
&= \nabla\cdot(\delta D\nabla u_0) + u_0\delta D|\nabla\psi_1|^2 + u_0\nabla\cdot(\delta D\nabla\psi_1) + 2u_0\delta D\nabla\psi_0\cdot\nabla\psi_1 \\
&-\delta\mu_a u_0 + f_0(e^{-\psi_1} - 1 + \psi_1).
\end{aligned}$$

The Rytov approximation comes in neglecting the second order terms on the right-hand side and noticing that $e^{-\psi_1} - 1 + \psi_1 = O(\psi_1^2)$ for a small perturbation ψ_1, so we finally find that

$$-\nabla\cdot(D_0\nabla(u_0\psi_1)) + \mu_{a0}u_0\psi_1 \approx \nabla\cdot(\delta D\nabla u_0) - \delta\mu_a u_0. \tag{52}$$

Taking this equation together with the Robin boundary conditions for $u_0\psi_1$, which can be deduced straighforwardly by (49) and (51), the first Rytov approximation gives the following boundary value problem

$$-\nabla\cdot(D_0\nabla(u_0\psi_1)) + \mu_{a0}u_0\psi_1 = \nabla\cdot(\delta D\nabla u_0) - \delta\mu_a u_0, \qquad \text{in } \Omega, \tag{53}$$

$$u_0\psi_1 + \frac{\zeta D_0}{2c_d}\frac{\partial(u_0\psi_1)}{\partial\mathbf{n}} = 0, \qquad \text{on } \partial\Omega. \tag{54}$$

Using Green's functions, the solution of the boundary value problem (53)-(54) can be expressed in an integral form. For, we define the Green's function associated to the Robin boundary value problem (53)-(54), that is the solution to the following problem

$$-\nabla\cdot(D_0(\mathbf{r})\nabla(G(\mathbf{r},\mathbf{r}'))) + \mu_{a0}(\mathbf{r})G(\mathbf{r},\mathbf{r}') = \delta(\mathbf{r}'), \qquad \mathbf{r},\ \mathbf{r}'\in\Omega \tag{55}$$

$$G(\mathbf{r},\mathbf{r}') + \frac{\zeta D_0(\mathbf{r})}{2c_d}\frac{\partial G(\mathbf{r},\mathbf{r}')}{\partial\mathbf{n}} = 0, \qquad \mathbf{r}\in\partial\Omega. \tag{56}$$

where $\delta(\mathbf{r}')$ is the delta distribution centered at $\mathbf{r}'$. Using classical arguments, we find that

$$u_0(\mathbf{r})\psi_1(\mathbf{r}) = \int_\Omega G(\mathbf{r},\mathbf{r}')\left[\operatorname{div}\left(\delta D(\mathbf{r}')\nabla u_0(\mathbf{r}')\right) - \delta\mu_a(\mathbf{r}')u_0(\mathbf{r}')\right]\,d\mathbf{r}',$$

hence, integrating by parts the first term on the righthand side, we get

$$\psi_1(\mathbf{r}) = -\frac{1}{u_0(\mathbf{r})}\int_\Omega\left[\delta D(\mathbf{r}')\nabla u_0(\mathbf{r}')\cdot\nabla G(\mathbf{r},\mathbf{r}') + \delta\mu_a(\mathbf{r}')u_0(\mathbf{r}')G(\mathbf{r},\mathbf{r}')\right]\,d\mathbf{r}'. \tag{57}$$

In the following analysis we assume that D_0 is constant in Ω, as it is customary done in CW-DOT applications (see, *e.g.*, papers [39, 102, 131] for examples of such a simplification which is customarily applied when a fast computation is required). Hence a reduced version of Eq.(57) can be considered. In fact, repeating the same calculations above, with the assumption that D_0 is constant and $\delta D = 0$, that is considering the boundary value problem

$$-\Delta(u_0\psi_1) + \frac{\mu_{a0}}{D_0}u_0\psi_1 = -\frac{\delta\mu_a}{D_0}u_0, \qquad \text{in } \Omega,$$

$$u_0\psi_1 + \frac{\zeta D_0}{2c_d}\frac{\partial(u_0\psi_1)}{\partial\mathbf{n}} = 0, \qquad \text{on } \partial\Omega,$$

an analogous simpler expression holds, which is

$$u_0(\mathbf{r})\hat{\psi}_1(\mathbf{r}) = -\int_\Omega \frac{\delta\mu_a}{D_0}(\mathbf{r}')u_0(\mathbf{r}')\hat{G}(\mathbf{r},\mathbf{r}')\,d\mathbf{r}', \tag{58}$$

where $\hat{G}$ is the Green's function, solution to

$$\begin{aligned} -\Delta\hat{G}(\mathbf{r},\mathbf{r}') + \frac{\mu_{a0}(\mathbf{r})}{D_0}\hat{G}(\mathbf{r},\mathbf{r}') &= \delta(\mathbf{r}'), \qquad \mathbf{r},\ \mathbf{r}' \in \Omega, \\ \hat{G}(\mathbf{r},\mathbf{r}') + \frac{\zeta D_0(\mathbf{r})}{2c_d}\frac{\partial\hat{G}(\mathbf{r},\mathbf{r}')}{\partial\mathbf{n}} &= 0, \qquad \mathbf{r} \in \partial\Omega. \end{aligned}$$

Moreover, let $\boldsymbol{\delta\mu}_a \in \mathbb{R}^{(N_v\times 1)}$ be the vector of unknown variations in absorption coefficient and $\boldsymbol{\psi}_1 \in \mathbb{R}^{(M\times 1)}$ be the vector of logarithmic fluctuations evaluated in the detector positions for each source. One obtains

$$\mathbf{J}\boldsymbol{\delta\mu}_a = \boldsymbol{\psi}_1, \tag{59}$$

where the sensitivity matrix $\mathbf{J} \in \mathbb{R}^{(M\times N_v)}$ stems from the discretization of the integral in Eq. 57 and, namely,

$$(\mathbf{J})_{ij} = \left[\frac{\Delta V}{u_0(\mathbf{r}_s^k,\mathbf{r}_d^l)}\hat{G}(\mathbf{r}_d^l,\mathbf{r}_j)\frac{1}{D}u_0(\mathbf{r}_s^k,\mathbf{r}_j)\right]. \tag{60}$$

Now, let $\mathbf{y}_\delta \in \mathbb{R}^{(M\times 1)}$ be the vector of logarithmic fluctuations obtained from measured data in the detector positions for each source

$$(\mathbf{y}_\delta)_i = \log\frac{U^M(r_s^k,r_d^l)}{U_0^M(r_s^k,r_d^l)}, \qquad i = 1,\ldots,M. \tag{61}$$

One ends up with the following discrete problem: find the best vector $\boldsymbol{\delta\mu_a}$ which minimizes $||\boldsymbol{\psi}_1 - \mathbf{y}_\delta||_2^2$.

3.4. Bayesian approaches

In Bayesian methods, both the measurements and the unknowns are modeled as random variables. The construction of the likelihood (observation) and the prior models is the starting point of the Bayesian approach to inverse problems. The most common likelihood models considers the case of an additive error model, i.e.,

$$y = \mathcal{F}(q) + \varepsilon,$$

where ε is the additive error term with a distribution $\pi_\varepsilon(\varepsilon)$, usually assumed to be mutually independent of q. Denoting by $\pi(q,y)$ the joint distribution and by $\pi(q)$ the *prior distribution*, then by the definition of the conditional probability we have that *posterior distribution* $\pi(q|y)$, describing the uncertainty of the unknown q given the measurements y, is then well-defined and given by

$$\pi(q,y) = \pi(y|q)\pi(q) = \pi(q|y)\pi(y),$$

that is

$$\pi(q|y) = \pi^{-1}(y)\pi(y|q)\pi(q) = \pi^{-1}(y)\pi(q,y),$$

which is the Bayes' theorem. Knowing the joint distribution, the posterior distribution can be computed easily but, unfortunately, the joint distribution is not practically available [19]. On the other hand, many times, the derivation of the *likelihood density* $\pi(y|q)$ can be obtained. In fact, by means of the Bayes' theorem

$$\pi(y|q) = \pi_\varepsilon(y - \mathcal{F}(q)).$$

In the special case of Gaussian additive noise, i.e. $\pi_\varepsilon(\varepsilon) = \mathcal{N}(\overline{\varepsilon}, \Gamma_\varepsilon)$, where $\overline{\varepsilon}$ is the mean and Γ_ε is the covariance function, we get

$$\pi(y|q) \propto \exp\Big(-\frac{1}{2}\|L_\varepsilon(y - \mathcal{F}(q) - \overline{\varepsilon}\|^2\Big),$$

where $\Gamma_\varepsilon^{-1} = L_\varepsilon^T L_\varepsilon$. In this context, the most common choice for the prior distribution is of the form

$$\pi(q) \propto \exp\Big(-\frac{1}{2}\|L_q(q - \overline{q})\|^2\Big),$$

where L_q is a usually a differential operator, and $\overline{q}$ is an a priori information. Using again the Bayes' theorem, we get

$$\pi(q|y) = \pi(y|q)\pi(q) \propto \exp\bigg(-\frac{1}{2}\Big(\|L_\varepsilon(y - \mathcal{F}(q) - \overline{\varepsilon}\|^2 + \|L_q(q - \overline{q})\|^2\Big)\bigg).$$

Then, the main goal is to compute the points of maximal probability. The most common point estimates are the *maximum a posteriori* estimate (MAP) or the *conditional mean* estimate (CM), defined by

$$q_{MAP} = \operatorname{argmax} \pi(q|y)$$

and

$$q_{CM} = \mathbb{E}(q|y) = \int q\pi(q|y)\, dq,$$

where $\mathbb{E}(\cdot)$ denotes the expectation. It can be shown that in the case of Gaussian processes then the MAP estimator coincides with a minimizers of a Tikhonov regularization [19]. The so-called "approximation error method" [98] introduces an uncertainty in the modelling, then the aim is to determine the statistical properties of the modeling errors and compensate for them making use of a Bayesian approach. We refer the reader to [17], [19] for an application of this method to the case of optical tomography.

3.5. Regularization methods

Due to predominant scattering phenomena in light transmission, the DOT inverse problem is severely ill-posed. Mathematically, this means that it does not fulfill all the Hadamard conditions [75], in the sense that there might be lack of uniqueness (several possible model parameter values are consistent with the data) and that there is lack of continuous dependence of the solution on the measurements (small errors in the data may lead to large errors in the model parameters). In these circumstances, it becomes essential the adoption of regularization techniques to solve the problem. One possible

way is to introduce in the optimization procedure (43) a regularization term such that the problem becomes [125]: seek the optimal parameter set q^* such that

$$q^* = \arg\min_{q\in Q}[\mathcal{D}(y_q, y_\delta) + \mathcal{R}(q)], \tag{62}$$

where $\mathcal{R} : Q \to [0, \infty]$ is a regularization functional which encodes soft or hard prior knowledge about the hidden structures to be investigated, favouring appropriate minimizers or penalizing those with undesired structures [16, 35, 62]. A commonly used form of (62) is the formulation

$$q^* = \arg\min_{q\in Q}[\mathcal{D}(y_q, y_\delta) + \lambda\mathcal{R}(q)], \tag{63}$$

which decouples the regularization functional $\mathcal{R}$ from the regularization parameter λ, whose role is to balance the trade–off the weight between the discrepancy term and the penalty functional. Solving (63) is however, again, not a trivial problem and several approaches have been proposed, which we will review in the following in the SSD-DOT reconstruction context.

Referring to the general linearized problem (linearized full problem or Rytov linearization) arising from the discretization, one has then to solve

$$\Delta\mathbf{q}^* = \arg\min_{\Delta\mathbf{q}} \frac{1}{2} \|\mathcal{J}\Delta\mathbf{q} - \mathbf{b}\|_2^2 + \lambda\mathcal{R}(\Delta\mathbf{q}), \tag{64}$$

where here we have used the general notation $\mathcal{J}$ and $\mathbf{b}$ for the tangent matrix and measured data, respectively. Notice that in the case of the Rytov approximation, we have that $\mathcal{J}$ corresponds to the sensitivity matrix defined in (60), $\Delta\mathbf{q}$ coincides with the vector $\boldsymbol{\delta\mu}_a$, and $\mathbf{b}$ is the vector whose components are defined in (61). Problem (64) can be tackled by several approaches, which we review in the following, focusing on the most relevant in the present field.

Truncated Singular Value Decomposition. When $\lambda = 0$ (no regularization), one can use the Singular Value Decomposition (SVD) of $\mathcal{J}$, $\mathcal{J} = \mathbf{V}\Sigma\mathbf{U}^\top$, where $\mathbf{V}, \mathbf{U}$ are orthogonal matrices whose columns represents the left and right eigenvectors and Σ is the diagonal matrix whose diagonal elements are the ordered singular values, $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_{n_r}$, n_r being the rank of $\mathcal{J}$. A solution to the problem can be written as

$$\Delta\mathbf{q} = \sum_{i=1}^{n_r} \frac{\langle \mathbf{b}, \mathbf{u}_i \rangle}{\sigma_i} \mathbf{v}_i. \tag{65}$$

In the case of ill-conditioned problems, singular values typically span several order of magnitude and the lowest singular values exacerbate the influence of the noise on the computed solution, inducing artifacts and unreliable reconstruction. A well known regularization technique is to consider a Truncated Singular Value Decomposition (TSVD), which consists in truncating the above sum to a lower number of addends, neglecting the smallest singular values.

Tikhonov Regularization. A widely explored choice is to set $\mathcal{R}(\mathbf{q}) = \frac{1}{2}\|\Gamma\mathbf{q}\|_2^2$, where Γ is a linear operator. This approach takes the name of Tikhonov regularization. When Γ is chosen as the identity operator, such strategy is named ℓ_2 regularization, or ridge regression in statistical frameworks, and promotes solutions with smaller norms. Other suitable choices for Γ are differential operators, which enhance smoothness of the computed solution. In this framework, problem (64) reads

$$\arg\min_{\Delta\mathbf{q}} \frac{1}{2}\|\mathcal{J}\Delta\mathbf{q} - \mathbf{b}\|_2^2 + \frac{\lambda}{2}\|\Gamma\Delta\mathbf{q}\|_2^2, \tag{66}$$

and admits the one-step optimal solution

$$\Delta\mathbf{q}^* = \left(\mathcal{J}^\top\mathcal{J} + \lambda\Gamma^\top\Gamma\right)^{-1}\mathcal{J}^\top\mathbf{b}. \tag{67}$$

Notice that when $\Gamma = \mathbf{I}$, then the solution can be written as

$$\Delta\mathbf{q}^* = \sum_{i=1}^{n_r} \frac{\sigma_i}{\sigma_i^2 + \lambda}\langle\mathbf{b}, \mathbf{u}_i\rangle\mathbf{v}_i, \tag{68}$$

where the regularization parameter at the denominator prevents the lowest singular values to exacerbate noise.

Remark 3.1 *Under suitable hypothesis, Gradient Descent methods can be applied for solving (66) (see also Remark 3.2 and Algorithm 1 for the more general case). When the regularization functional $\mathcal{R}$ is convex but not differentiable, one can resort to Bregman iterative techniques [33].*

Remark 3.2 *Under the more general framework (44), when Rytov approach or a similar one is not employed, one may use a Levenberg–Marquardt algorithm to solve*

$$\arg\min_{\mathbf{q}} \frac{1}{2}\|\mathcal{F}(\mathbf{q}) - \mathbf{y}_\delta\|^2 + \frac{\lambda}{2}\|\Gamma\mathbf{q}\|^2.$$

Algorithm 1 presents a pseudo-code of a possible implementation.

Algorithm 1 Levenberg–Marquardt Algorithm

Set the initial iterate $\mathbf{q}^{(0)}$, set $\mathbf{J}$ the first order Frechét derivative. Set $\boldsymbol{\beta} = \mathbf{y}_\delta - \mathcal{F}(\mathbf{q}_0)$
for $k = 0, 1, \ldots$ **do**
  Compute the residual: $\boldsymbol{\rho}^{(k)} \leftarrow \boldsymbol{\beta} - \mathcal{J}\mathbf{q}^{(k)}$.
  Solve $(\mathcal{J}^\top\mathcal{J} + \lambda\Gamma^\top\Gamma)\mathbf{d} = \mathcal{J}^\top\boldsymbol{\rho}^{(k)}$
  Update $\mathbf{q}^{(k+1)} \leftarrow \mathbf{q}^{(k)} + \mathbf{d}$
  $\beta \leftarrow \mathbf{y}_\delta - \mathcal{F}(\mathbf{q}^{k+1})$
  Stop if a convergence criterion is satisfied.
end for

The choice of the parameter λ is pretty delicate: it can be made *a priori* but it can also depend on the noisy data y_δ and then it is called an *a posteriori* strategy, which consists in adapting this parameter along the iterations. In this latter case, λ is typically

chosen in dependence of the noise level δ, i.e. if σ^2 is the variance of data y_δ, a well known choice is $\lambda \to 0$, $\sigma^2\lambda \to 0$ for $\delta \to 0$.

In practice, several semi-empirical methods exist: the L-curve method [80, 81] is the first of this list. It is an *a posteriori* strategy which employs only the data y_δ and the computed solution. The L-curve method computes several regularized solution $\{\mathbf{q}_\lambda\}$ for pre-selected set of values λ chosen in a fairly ample range. Then the ℓ_2 norm of the residuals $\{\|\mathbf{J}\mathbf{q}_\lambda - y_\delta\|_2\}$ and the ℓ_2 norm of the regularized solutions $\{\|\mathbf{q}_\lambda\|_2\}$ are plotted against each other in a log-log plot: the optimal value (and the corresponding optimal solution) is the one related to the corner of the L-shape. Such point can be computed by estimating the maximum curvature of the curve [81, 47, 79]. Numerically speaking, the L-curve method is easily tractable, but it has significant inner limitations: the distance of optimal regularized solution from the ground truth increases as $\delta \to 0$ [78], or when the dimension of the problem is too large [141]. The method can be tailored for other regularizers rather then ℓ_2 norm [107]. Examples of the use of the L-curve strategy to solve DOT problems are reported in [151, 103, 138, 108].

The Generalized Cross Validation (GCV) [72] consists in estimating the optimal value for λ as the minimizer of the function

$$\Psi(\lambda) = \frac{\|(\mathbf{I} + \mathcal{J}\left(\mathcal{J}^T\mathcal{J} + \lambda\Gamma^\top\Gamma\right)^{-1}\mathcal{J}^\top)\mathbf{b}\|^2}{\left[\mathrm{tr}(\mathbf{I} + \mathcal{J}\left(\mathcal{J}^T\mathcal{J} + \lambda\Gamma^\top\Gamma\right)^{-1}\mathcal{J}^\top)\right]^2}. \tag{69}$$

Despite having a complicated appearance, the function $\Psi(\lambda)$ can be easily minimized using the Fourier transform or, when $\Gamma = \mathbf{I}$, the Singular Value Decomposition (SVD) of $\mathcal{J}$. The idea beyond this procedure relies on the principle that when some data value is missing, then a suitable choice for λ should provide a good estimation of such missing value [80]. The major strength of GCV approach is that it does not need any estimation of noise level and its minimization is rather easy. On the other hand, the graph of Ψ can present a plateau close to its minimizer and numerical methods may fail in localizing it [140].

The Minimal Residual Method (MRM, [94]) consists in a steepest descent method where the regularization parameter is estimated online, using the last approximated solution. The main drawback of such method is that the theoretical analysis assumes that $\delta = 0$, *i.e.* no noise is affecting the data: in practical applications this leads to a sub-optimal choice for the regularization parameter.

The Discrepancy Principle is an *a posteriori* method for selecting the optimal regularization parameter. Denote with Δq_λ a solution of (66), where the dependency on λ is explicit. The optimal parameter is given by

$$\lambda_{\mathrm{opt}} = \sup\{\lambda > 0 \,|\, \|\mathcal{J}\Delta\mathbf{q}_\lambda - \mathbf{b}\| < \tau\delta\},$$

where δ is the noise level and $\tau > 1$, see [62] for the technical details. This strategy is very useful from a theoretical point of view, but it requires the information about the noise level that is realistically not often available in practical problems. Even when δ is known, this approach provides oversmoothed solution [80]. For overcoming such

issue, several methods use this principle for dynamically estimating λ. In [145] the authors develop a Model Function Method (MFM) based on a damped version of such principle. In [97] a Padé approximation is employed for generalizing also to non smooth functionals. As a matter of fact, Morozov's principle aims to provide the optimal value for λ not only for Tikhonov regularization, but also for other type of functionals (*e.g.* Total Variation).

Remark 3.3 *One may use the SVD for estimating the noise level perturbing the data [80, Section 6.6], [137], or statistical methods [90, 51].*

As shown in Eq. (66), the regularization term can encompass a linear operator Γ: this latter may include *a–priori* information on the tissue structure, for example the operator Γ can be generated using spatial information, typically derived from other imaging modalities. In [56] Γ is a matrix (of suitable dimension) that links all the elements (pixel, voxels or nodes) in a particular tissue type (glandular or fatty) as recognized from MRI, so that a second differential operator is approximated within each region, in a way conceptually similar to Total Variation functional. Each node in the mesh is labeled according to the region, or tissue type, with which it is associated. Namely one can choose:

$$\Gamma_{i,j} = \begin{cases} -1/n & \text{region}_i = \text{region}_j \\ 1 & i = j \\ 0 & \text{otherwise} \end{cases} \tag{70}$$

where i and j are points within a region and n is the total number of unknowns in a given region. Another strategy to construct the linear operator Γ is to employ previous knowledge from ultrasound images [12].

The authors in [59] use L-curve in combination with damped Gauss-Newton iterations to solve the (complex) DE upon finite element discretization. A quasi-optimal Tikhonov parameter is identified, based on a coarse mesh for the control-space. A finer mesh is used afterwards for better reconstructions exploiting the quasi–independence of the quasi-optimal Tikhonov parameter with the control-space discretization.

The above methods for the selection of the parameter λ refer to the case in which the noise perturbing the data is of Gaussian type. In presence of Poisson noise such estimation can be achieved via a discrepancy principle [37, 149].

Sparsity Promoting Regularization. In practice, contrast regions where the absorption coefficient is altered are restricted to limited areas embedded in the healthy tissue and the discrepancy with respect to the physiological value can be significant (up to 5-10 times the physiological value). This implies that using the ℓ_2 norm as regularization often causes an unphysical smoothing of the coefficient value and an enlarged distribution. Hence, one can resort to regularization functionals that promote sparsity and enhance differences between healthy and altered regions. A suitable choice is the ℓ_1 norm

$$\mathcal{R}(\Delta\mathbf{q}) = ||\Delta\mathbf{q}||_1 = \sum_i |\Delta\mathbf{q}_i|. \tag{71}$$

In [43] such choice is employed and some numerical tests on a 3D phantom show reliable results. The authors in [142] solve problem (44) when $\mathcal{R}$ is the ℓ_1 norm employing the FISTA algorithm [29]. In [31] it is shown that in some cases a pure ℓ_1 regularization may fail in estimating the shape, area and value of even simple heterogenieties, whilst employing a Bregman technique with this regularization functional provides more reliable results, particularly when there are contrast regions with different absorption coefficients. Compressed sensing theory [147] suggests that the regularization parameter should be selected as

$$\lambda > \|\mathcal{J}^\top \mathbf{b}\|_\infty$$

in order to avoid algorithms to reach a trivial zero solution.

Total Variation regularization. Suppose that the function q is differentiable. The Total Variation regularization is defined as

$$\mathcal{R}(\Delta q) = \int_\Omega |\nabla(\Delta q)|\, d\Omega, \tag{72}$$

where ∇q denotes the gradient of q, Ω is an open and bounded subset and $d\Omega$ a suitable measure. When q is not differentiable, one has to refer to the more general definition [50]. Its discrete counterpart reads

$$TV(\Delta q) = \sum_i \|\mathbf{A}_i \Delta q\|,$$

where $\mathbf{A}_i$ is the 2D discrete first order difference operator of suitable dimension acting on the ith element of q. The authors in [107] develop a method which couples an homotopy strategy with Bregman iterations [33], using as regularization the Total Variation function together with a Tikhonov term. The regularization parameter λ is chosen by an adaptive strategy, that relies on the knowledge of the noise level δ. In [110] the Total Variation regularization is implemented under a graph framework and implementing an alternating direction method of multipliers (ADMM). In this work the authors use the L-curve method for the estimation of λ. A similar approach is adopted in [106], where the objective function consists in a convex and in a nonconvex term, the latter being a Huber function subtracted from the general Total Variation functional. In [133] it is presented a solid and complete theoretical analysis for the treatment of Eq. (46) when $\mathcal{R}$ is chosen as a nonconvex nonsmooth total variation based regularization functional.

Mixed functionals. The functional $\mathcal{R}$ may consists in more than one term, leading to a mixed functional: for example, in [107] the Total Variation is coupled with the ℓ_2 norm. The authors in [134] used a split Bregman method including a weighted combination of TV and ℓ_1 norms to compute a stable approximation of the discrepancy functional when the forward model is represented by the RTE. They tested the approach on a 2D synthetic case consisting in a circular domain with different targets. The

experimentation is done by manually setting the weights of the two terms, and it led to conclude that the TV+ℓ_1 regularization performs better than the single counterparts in reconstructing and locating the small inclusion. Specifically, the weight of ℓ_1 norm was chosen one order of magnitude larger that the TV one (10^{-3} vs 10^{-4}). When the pathological tissue covers large regions of the investigated body, there were no significant differences in using TV+ℓ_1 or TV regularization, beside a slight improvement in deblurring for the combined regularization. Also in [120] a combination of TV and ℓ_1 is employed.

In [31] the authors employ the Elastic Net functional:

$$\mathcal{R}(\Delta q) = (1-\alpha)\|\Delta q\|_2^2 + \alpha\|\Delta q\|_1, \quad \alpha \in [0,1]. \tag{73}$$

Originally introduced in the statistical framework [153], it consists in a convex combination of ℓ_2 and ℓ_1 norms, and it aims to couple the positive priors imposed by both functionals: it preserves smooth components and at the same time promotes sparsity of the solution. Upon a suitable choice of the parameters, the Elastic Net provides reliable results, both when Δq is localized in one region and when Δq presents several components with different values. In [48] this procedure has been applied also to 3D cases.

Numerical tests. We present here a series of numerical tests aimed at comparing the quality of the reconstruction obtained via some of the methodologies discussed above. This discussion is not aimed at being an exhaustive compendium of the results obtained in this field, rather at showing some significant aspects. All the tests are solved via in-house implementations of the authors of the present paper. We consider both on 2D and 3D geometries. The 2D setting consists of a semi-circular domain of radius 5 cm discretized through a uniform voxel-based mesh composed of 3822 voxels, which corresponds to a spatial resolution of about 0.1 cm. The location of detectors (200) and sources (20) is depicted in Fig. 3(a). The 3D setting consists of a semi-spherical domain of radius 5 cm discretized through a uniform voxel-based mesh composed of 72407 voxels, which corresponds to a spatial resolution of about 0.15 cm. These geometries, albeit in a reduced dimensional setting and idealized version, represents a female breast supported on a solid plate on which embedded light sources are alternatively turned on for the DOT exam. The location of detectors (200) and sources (16) for this case is depicted in Fig. 3(b). We embed in a uniform background, with absorption coefficient $\mu_a = 0.01$ cm^{-1}, one or two contrast regions represented as circular (2D) or spherical (3D) areas, respectively. If not specified differently, we solve the problems using the Rytov perturbative approach and regularization represented by the elastic net (73) with parameters $\alpha = 0.5$ and λ of the order of 10^{-4}. Moreover, we use the Green's function method to address the solution, with a dipole approximation to enforce the null Dirichlet boundary condition on the flat portion of the boundary.

Test: reconstruction with different strength of the contrast regions. We consider the 3D

setting and we embed in the domain two spherical contrast regions, both having radius 0.5 cm and centered respectively at $(-2.5, 0, 2)$ and $(2.5, 0, 2)$. We consider the cases of equal ($2\times$) and different ($5\times$ and $2\times$ on the leftmost and rightmost region) strengths for the contrast regions with respect to the background, respectively. Results are shown in Figure 4b). In the case of equal strength, both regions are identified, well localized and their intensity is accurately estimated. In the case of two regions with different strength, both regions are identified as well, although the reconstruction has a lower quality both in terms of the localization of the interested voxels and in the estimation of the contrast intensity, especially for the weaker contrast. Similar results are obtained in the 2D case (not shown here). Note that one should pay much attention in the choice of the regularization parameter λ. As a matter of fact, exceeding in regularizing causes the merging, or even the overlap, of the regions if they have equal intensities, or the identification of the only higher strength inclusion if the intensities are different.

Test: comparison of different methods in the 2D case. Three different cases are considered:

i) a single circular inclusion is centered in $(0, 2)$ with radius 0.3 cm. The absorption coefficient of the inclusion is set to $2 \times \mu_{a,0}$;

ii) two circular inclusions are centered in $(\pm 2, 2)$ both with radius 0.3 cm. The absorption coefficient of the inclusions is set to $2 \times \mu_{a,0}$;

iii) as in case ii) but the absorption coefficient of the left inclusion is set to $5 \times \mu_{a,0}$, while for the right one is set to $2 \times \mu_{a,0}$.

Tests are performed using the elastic-net regularization method implemented in the `glmnet` software package [68] and alternatively using the Bregman procedure as described in detail in [32]. The results are depicted in Fig. 5, where the two approaches are compared with other techniques. We observe that the elastic-net approach correctly locates the inclusion(s) with their absorption coefficient for tests i) and ii). In test iii) the localization is less precise, but the expected values of the absorption coefficient are well estimated. Using the Bregman procedure, choosing the 2-norm regularization, albeit the quality of the reconstruction is improved by the contrast enhancement, the inclusion is not clearly identified and the recovered halo can be confused with some artefacts. On the other hand, setting the 1-norm regularization really benefits from the Bregman implementation: the results for test i) show that the location of the inclusion is properly recognized and its absorption value is correctly reconstructed. The reconstruction of test ii) suffers from a small displacement, but the value of the coefficient is well recovered. Eventually, in test iii), both inclusions are not completely placed in the correct location, but the absorption value is close to the correct one. This misplacement is due to the different values of absorption coefficient, hence the regularization parameter promotes the higher one.

Test: effect of boundary conditions. We consider here both the 2D and 3D settings. In

the 2D case we embed a single circular perturbed region with radius 0.3 cm, centered at $(2, 2)$ and with contrast $2\times$ with respect to the background value. We consider two different strategies:

a) standard approach, which neglects the Robin boundary condition on the curved boundary and uses to represent the solution u_0 in (60) as Green's function corresponding to a semi–infinite domain, with a dipole approximation on the straight portion of the boundary (see Fig. 6a);

b) do not neglect the Robin boundary condition for the representation of the solution u_0 in relation (60) using more accurate numerical method. Here we consider the Method of Fundamental Solution (MFS, [67]) which allows, albeit in an approximate manner and in a moderately more computationally expensive manner, to enforce boundary conditions on all the boundary in the framework of a Green's function solution (see Fig. 6b).

An evident improvement in terms of localization of the contrast region is observed when correct boundary conditions are taken into account.
We perform the same experiment in the 3D setting using again the Rytov perturbative approach with elastic net regularization and investigating the same strategies as in the 2D case. Figure 7 represents from left to right for the upper and lower rows (first and second strategy, respectively): i) the shape and position of the simulated inclusion (red sphere) and the voxels for which the computed solution has a non-zero component (i.e. the reconstructed inclusion); ii) the computed solution restricted to the vertical plane containing the center of the contrast region; iii) the computed solution restricted to the horizontal plane containing the center of the contrast region. Again, it is apparent the improvement when physically correct boundary conditions are considered. One should consider however that in the 3D case the computational times are significantly larger and one should balance this aspect with the effective improvement.

Test: adaptation of the voxel size. A possibility to reduce the execution time is to first employ a coarse-mesh to locate in a rough manner the contrast area and then use this information as a soft-prior to cover with the voxel-based mesh only the identified ROI. We report the results for the same setting of the previous test case in the 3D geometry, using the second strategy. Figure 8 show the interested voxels superposed to the exact contrast region, along with the voxel size. Notice that different values of the elastic net parameters are required to obtain optimal results when the voxel size is changed.

4. DOT reconstruction aided by Machine Learning

Physically–driven approaches present mainly two drawbacks. First, PDE models give only an approximate description of the physical phenomena, being accurate in some circumstances, less in other. Even if one considers pretty sophisticated models, as the RTE, which translate in increasing complexity, this usually imply i) the introduction

of several new parameters to be estimated; ii) a strong dependence on regularization techniques which require a delicate tuning; iii) in general, an increasing request of computational resources that often cannot be met in practical applications. Nowadays, the possibility of acquiring a large number of data is bringing the attention of the inverse problems community to data-driven methods that are able to capture data features without making any particular a priori assumption in the reconstruction procedure and without the need to postulate complex mathematical models. In this respect, machine learning, with specific emphasis on deep neural networks (NN)-based techniques, has shown great potential in solving problems in image processing with special relation to the biomedical field [136]. The application of deep learning in DOT reconstruction is very recent and studies in this field are still not abundant. In the following, we provide a overview of such attempts, along with the description of their most salient features, highlighting trends and patterns. For a general introduction to machine learning and deep learning we refer to the many available textbooks, e.g. [73, 3].

4.1. Earlier contributions

Patra et al [117] used an embryonic version of a fully connected NN with 2 hidden layers to obtain a prior on the spatial localization of a single contrast region embedded in a 2D circular domain. Accordingly, in the iterative reconstruction procedure they updated only the parameters in a neighborhood around the detected contrast region. Sun et al [130] addressed the multiple scattering problem of microwaves in biological samples using a two-step reconstruction method, where an analytical method based on the linear backprojection operator provided a first image estimate, followed by a U-Net decoder for image reconstruction. This approach avoided the iterative evaluation of the nonlinear discrete Lippmann-Schwinger operator (or its Jacobian). Feng et al. [63] adopted a NN based on a fully connected architecture made of 3 layers for end-to-end DOT image reconstruction, the internal hidden layer comprising 695 neurons, for a total of about 1.56M parameters in the test examples. Training was performed using 20,000 synthetically generated 2D samples (with 2% added Gaussian noise), requiring about 26 hours of computation on a mid-level personal computer. For comparison, DOT reconstruction based on Tikhonov regularization-based was also performed, as well as ℓ_1 and TV-based reconstructions with manually chosen optimal regularization parameters. The results showed that the NN-based reconstruction outperformed the Tikhonov regularization-based on in terms of accuracy and image quality and also the ℓ_1 and TV-based reconstructions, even if to a lesser extent for these latter. They also explored the use of different activation functions and a reduced number of neurons in the hidden layer, albeit obtaining results far from optimal.

4.2. Fully data-driven approaches

Fully data-driven approaches completely rely on data and they do not assume the existence of a physical model of the underlying processes. The network parameters

are tuned according to supervised training procedures which compare the ground truth with the net predictions. In this framework, Yoo et al [148] investigated an end-to-end model based on a DNN-architecture to reconstruct heterogeneous optical maps in small animals. A fully connected layer performed a first inversion, followed by an encoder-decoder structure which implemented a deep 3D convolutional framelet. It is interesting to notice that, despite this approach being fully data-driven, the neural network structure was inspired by the Schwinger–Lippmann integral equation model which provides a nonlinear representation of the scattered field. Deng et al [57] proposed an architecture in which a fully connected layer input the data into an encoder-decoder structure followed by a U-Net for image denoising and quality improvement. Skip connections further enhanced high-resolution features for reconstruction. The proposed loss function penalized the inaccurate contributions from the contrast regions rather than the whole volume. Yedder er al [146] exploited an architecture formed by a first fully–connected layer followed by deep nets with spatial-wise residual attention mechanisms (implemented by convolutional blocks) to focus on important features. A fuzzy Jaccard loss component was used in addition to the reconstruction error loss to address the imbalance between the minority pixels corresponding to one or more lesions versus the majority of background pixels. Transfer learning techniques were used to bridge the gap between in silico simulated data and equivalent data acquired with the real device on phantoms. Upon training of the net on the sole basis of simulated data, the authors performed DOT image reconstruction on phantom and patients diagnosed with breast cancer (2 participants). They contrasted the results of the proposed method to several existing methods, and, namely, with conventional approaches and other deep–learning based methods (including their previous work), showing the increased accuracy of these latter, albeit without being able to remove the notorious difficulty of reconstructing deeper and multiple lesions. Benfenati et al [34] investigated a NN-based regularization technique, inspired by the Learned-SVD method originally proposed in [40] for general inverse problem. This approach is a fully data–driven strategy which adopts two autoencoder networks bridged by an operator which mimics the effect of a (truncated) singular value matrix in a nonlinear fashion. In the numerical experiments, noise-free data as well as noisy measurements were considered, obtained by adding white Gaussian noise to the fluence values at the observations points with variance 1%, 3% or 5%, respectively. The performance was assessed by 1) the average value of the absorption coefficient (ACR) inside the reconstructed contrast regions; 2) the True Positive Ratio (TPR), which assesses the spatial accuracy of the positioning of the reconstructed contrast regions by checking how many pixels in the reconstruction actually belong to the true contrast region. The results are summarized in Table 2, which presents the ACR and TPR metrics averaged on 150 test samples for the Learned-SVD approach and, for comparison, for the elastic net and Bregman variational approaches. For all the considered approaches, increasing levels of noise significantly worsen the TPR, as well as the ACR. However, the quality loss exhibited by the Learned-SVD approach is significantly less pronounced, especially for the ACR

Table 2. ACR (binned according to the intensity of the contrast region) and TPR metrics for different noise levels for the Learned-SVD [34] and for comparison the Elastic Net and Bregman variational approaches. Values are averaged over 150 test samples. Noise level 5% is not considered for the variational approaches due to the extreme difficulty to obtain a sensible reconstruction.

Noise Level	$3\mu_{a,0}$ (GT: 3e-2)	$4\mu_{a,0}$ (GT: 4e-2)	$5\mu_{a,0}$ (GT: 5e-02)	TPR
	Learned-SVD			
0%	3.05e-02 ± 2.00e-03	4.08e-02 ± 3.72e-03	4.74e-02 ± 2.19e-03	0.94
1%	3.07e-02 ± 2.77e-03	3.76e-02 ± 5.54e-03	4.38e-02 ± 4.01e-03	0.76
3%	3.06e-02 ± 5.56e-03	3.61e-02 ± 6.10e-03	4.11e-02 ± 4.65e-03	0.52
5%	3.00e-02 ± 3.82e-03	3.15e-02 ± 4.44e-03	3.39e-02 ± 6.04e-03	0.46
	Elastic Net			
0%	2.73e-02 ± 4.74e-03	3.45e-02 ± 4.96e-03	3.90e-02 ± 8.69e-03	0.45
1%	2.91e-02 ± 9.36e-03	4.30e-02 ± 8.70e-03	5.01e-02 ± 9.80e-03	0.17
3%	9.55e-02 ± 1.92e-02	1.25e-01 ± 2.83e-02	1.23e-01 ± 3.36e-02	0.05
	Bregman			
0%	4.02e-02 ± 8.95e-03	5.93e-02 ± 1.61e-02	8.54e-02 ± 2.98e-02	0.26
1%	4.77e-02 ± 1.81e-02	5.85e-02 ± 1.84e-02	8.34e-02 ± 2.85e-02	0.17
3%	1.28e-01 ± 5.50e-02	1.45e-01 ± 9.06e-02	1.36e-01 ± 6.80e-02	0.03

metric. As a matter of fact, the reconstructed regions with the NN-based method still show a notable contrast with respect to the background value, and 46% of the extension of the perturbed regions is correctly recovered on average even for the highest noise level (vs 5% and 3% for the variational approaches, respectively). One should notice that in the Elastic Net approach the reconstructed intensities constantly approximate the nominal one by defect for low noise level: this is not fully surprising, since regularization comes at the cost of a certain degree of smearing/blurring of the solution. On the other hand, Bregman technique increases the contrast, as already observed in previous works. In presence of high level of noise, the results are completely unreliable. Classical variational methods suffer from this problem to a higher degree especially for a rather coarse voxelization as the one we considered here (results, not reported here, for finer voxelizations, do confirm in any case this general trend).

4.3. Hybrid model/data-driven approaches

In contrast to directly learning a reconstruction operator, one can combine deep learning with model-based (knowledge–driven) approaches. In this framework, Jiang et al [96] investigated DOT reconstruction with limited available boundary measurements. The Direct Sampling Method was used as the idea guiding the design of a neural network

which learned from data the index function representing the location and shape of contrast regions. Mozumder et al [112] followed an approach based on iterative model-based techniques where learned components given by a CNN were intertwined with the model physical equations. Specifically, a finite element solver was used to approximate the solution of the physical model, embedded in a Gauss-Newton method whose update step was chosen accordingly to a CNN. The method was shown to favorably compare with a standard Bayesian inversion method. Zou et al [154] used an autoencoder-like architecture comprising two neural networks. The first net learned the photon transport process using as an input the optical coefficient map ground truth and producing the predicted perturbation while the second net tried to reconstruct the optical properties from measurements. The weights obtained from the initial training process were further refined by adding physical constraints based on the difference between the predicted perturbation and the perturbation produced by the Born model and anatomical information by co-registered US.

4.4. General trends in machine learning-based DOT reconstruction

Despite apparent differences, the above works show common trends which we highlight here to provide a rationale of the present state of the art in this field:

- fully data-driven approaches generally adopt a NN architecture consisting in a fully connected layer which performs a first inversion and then refine/denoise the result via a series of convolutional layers. Existing frameworks, such as U-nets, can be conveniently adapted to the refinement task with minimal changes
- end-to-end approaches based on NNs can benefit from knowledge coming from physical models by incorporating in the net structure information about the physical process, for example in the choice of the convolutional filters. Physical constraints can also be added in the form of the loss function
- skip connections/attention mechanisms may enhance the recognition of the contrast regions and deviate computational resources on these areas, alleviating the computation from the background
- all the above mentioned papers only train the network on synthetic data, for which the ground truth is readily available. However, the trained nets show a good generalization capability to real data (albeit tested on very limited realistic datasets). With this respect, some authors argue that NNs, which directly learn from the measurement data, suffer less from prior conditions that might bias the search space and this may favor generalization. This is however still a poorly studied topic, which will require more investigation on larger datasets.

A synthetic illustration of the above concepts is presented in Figure 9.

5. Conclusions and perspectives

DOT has clear advantages over other imaging techniques in medical imaging, being able to detect lesions without ionization. As such it can be employed safely in both diagnosis and therapy follow up. However, much room is open to improvements in order for DOT to enter into clinical practice. Unlike X-ray CT, DOT image reconstruction is indeed affected by the scattering nature of light propagation in biological media, which significantly reduces the spatial resolution and quantification of images. Aim of this work was to review the mathematical foundations of the following issues: i) derivation of optical coefficients, which are the reconstructed quantities; ii) derivation and well posedness of the mathematical models used to describe the underlying physical processes; iii) approximation methods and numerical techniques both of physics-driven and data-driven type. In this paper we have focused our attention on the so-called Steady State-Domain DOT technology. We refer to [15] for a recent review on other existing DOT technologies. We also refer to [116] and [89] for interesting discussions about clinical aspects in the use of DOT and its integration with other imaging modalities.

ORCID iDs

Andrea Aspri https://orcid.org/0000-0002-2995-8365
Alessandro Benfenati https://orcid.org/0000-0002-2985-374X
Paola Causin https://orcid.org/0000-0002-8285-8101
Cecilia Cavaterra https://orcid.org/0000-0002-2754-7714
Giovanni Naldi https://orcid.org/0000-0001-9191-6003

Acknowledgments

The authors thank the referee for the helpful comments that improved the quality of the manuscript.
This research has been partially performed in the framework of the MIUR-PRIN Grant 2020F3NCPX "Mathematics for industry 4.0 (Math4I4)" and GNCS Project "Metodi di ottimizzazione data-driven: nuove prospettive teoriche e pratiche". Andrea Aspri and Cecilia Cavaterra are members of GNAMPA (Gruppo Nazionale per l'Analisi Matematica, la Probabilità e le loro Applicazioni) of INdAM (Istituto Nazionale di Alta Matematica). Alessandro Benfenati, Paola Causin and Giovanni Naldi are members of GNCS (Gruppo Nazionale Calcolo Scientifico) of INdAM.

References

[1] In WG Egan and TW Hilgeman, editors, *Optical Properties of Inhomogeneous Materials*, pages 221–226. Academic Press, 1979.

[2] E Abuelhia and A Alghamdi. Evaluation of arising exposure of ionizing radiation from computed tomography and the associated health concerns. *Journal of Radiation Research and Applied Sciences*, 13(1):295–300, 2020.

[3] CC Aggarwal et al. Neural networks and deep learning. *Springer*, 10:978–3, 2018.

[4] V Agoshkov. *Boundary Value Problems for Transport Equations*. Modeling and Simulation in Science, Engineering and Technology. Birkhäuser Basel, 1998.

[5] GS Alberti and H Ammari. Disjoint sparsity for signal separation and applications to hybrid inverse problems in medical imaging. *Appl. Comput. Harmon. Anal.*, 42(2):319–349, 2017.

[6] GS Alberti and Y Capdeboscq. *Lectures on elliptic methods for hybrid inverse problems*, volume 25 of *Cours Spécialisés [Specialized Courses]*. Société Mathématique de France, Paris, 2018.

[7] GS Alberti and M Santacesaria. Calderón's inverse problem with a finite number of measurements. *Forum Math. Sigma*, 7:Paper No. e35, 20, 2019.

[8] GS Alberti and M Santacesaria. Calderón's inverse problem with a finite number of measurements II: independent data. *Appl. Anal.*, 101(10):3636–3654, 2022.

[9] GS Alberti and M Santacesaria. Infinite-dimensional inverse problems with finite measurements. *Arch. Ration. Mech. Anal.*, 243(1):1–31, 2022.

[10] G Alessandrini. Stable determination of conductivity by boundary measurements. *Appl. Anal.*, 27(1-3):153–172, 1988.

[11] G Alessandrini, M Di Cristo, E Francini, and S Vessella. Stability for quantitative photoacoustic tomography with well-chosen illuminations. *Ann. Mat. Pura Appl. (4)*, 196(2):395–406, 2017.

[12] M Althobaiti, H Vavadi, and Q Zhu. Diffuse optical tomography reconstruction method using ultrasound images as prior for regularization matrix. *Journal of Biomedical Optics*, 22(2):026002, 2017.

[13] H Ammari. *An introduction to mathematics of emerging biomedical imaging*, volume 62 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer, Berlin, 2008.

[14] H Ammari, E Bossy, V Jugnon, and H Kang. Reconstruction of the optical absorption coefficient of a small absorber from the absorbed energy density. *SIAM J. Appl. Math.*, 71(3):676–693, 2011.

[15] MB Applegate, RE Istfan, S Spink, A Tank, and D Roblyer. Recent advances in high speed diffuse optical imaging in biomedicine. *APL Photonics*, 5(4):040802, 2020.

[16] S Arridge, P Maass, O Öktem, and C-B Schönlieb. Solving inverse problems using data-driven models. *Acta Numer.*, 28:1–174, 2019.

[17] S. R. Arridge, J. P. Kaipio, V. Kolehmainen, M. Schweiger, E. Somersalo, T. Tarvainen, and M. Vauhkonen. Approximation errors and model reduction with an application in optical diffusion tomography. *Inverse Problems*, 22(1):175–195, 2006.

[18] SR Arridge. Optical tomography in medical imaging. *Inverse problems*, 15(2):R41, 1999.

[19] SR Arridge, JP Kaipio, V Kolehmainen, and T Tarvainen. Optical imaging. In *Handbook of mathematical methods in imaging. Vol. 1, 2, 3*, pages 1033–1079. Springer, New York, 2015.

[20] SR Arridge and WRB Lionheart. Nonuniqueness in diffusion-based optical tomography. *Opt. Lett.*, 23(11):882–884, Jun 1998.

[21] SR Arridge and JC Schotland. Optical tomography: forward and inverse problems. *Inverse problems*, 25(12):123010, 2009.

[22] A Aspri, E Beretta, O Scherzer, and M Muszkieta. Asymptotic expansions for higher order elliptic equations with an application to quantitative photoacoustic tomography. *SIAM J. Imaging Sci.*, 13(4):1781–1833, 2020.

[23] K Astala and L Päivärinta. Calderón's inverse conductivity problem in the plane. *Ann. of Math. (2)*, 163(1):265–299, 2006.

[24] G Bal and K Ren. Multi-source quantitative photoacoustic tomography in a diffusive regime. *Inverse Problems*, 27(7):075003, 20, 2011.

[25] G Bal and K Ren. Non-uniqueness result for a hybrid inverse problem. In *Tomography and inverse*

transport theory, volume 559 of *Contemp. Math.*, pages 29–38. Amer. Math. Soc., Providence, RI, 2011.

[26] G Bal and K Ren. On multi-spectral quantitative photoacoustic tomography in diffusive regime. *Inverse Problems*, 28(2):025010, 13, 2012.

[27] G Bal and G Uhlmann. Inverse diffusion theory of photoacoustics. *Inverse Problems*, 26(8):085010, 20, 2010.

[28] G Bal and G Uhlmann. Reconstruction of coefficients in scalar second-order elliptic equations from knowledge of their solutions. *Comm. Pure Appl. Math.*, 66(10):1629–1652, 2013.

[29] A Beck and M Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM Journal on Imaging Sciences*, 2(1):183–202, 2009.

[30] A Benfenati, G Bisazza, and P Causin. A Learned SVD approach for inverse problem regularization in diffuse optical tomography, 2021.

[31] A Benfenati, P Causin, MG Lupieri, and G Naldi. Regularization techniques for inverse problem in DOT applications. *Journal of Physics: Conference Series*, 1476(1):012007, mar 2020.

[32] A Benfenati, P Causin, MG Lupieri, and G Naldi. Regularization techniques for inverse problem in dot applications. In *Journal of Physics: Conference Series*, volume 1476, page 012007. IOP Publishing, 2020.

[33] A Benfenati and V Ruggiero. Inexact Bregman iteration with an application to Poisson data reconstruction. *Inverse Problems*, 29(6):065016, may 2013.

[34] Alessandro Benfenati, Giuseppe Bisazza, and Paola Causin. A learned-svd approach for regularization in diffuse optical tomography. *arXiv preprint arXiv:2111.13401*, 2021.

[35] M Benning and M Burger. Modern regularization methods for inverse problems. *Acta Numerica*, 27:1–111, 2018.

[36] E Beretta, M Muszkieta, W Naetar, and O Scherzer. A variational method for quantitative photoacoustic tomography with piecewise constant coefficients. In *Variational methods*, volume 18 of *Radon Ser. Comput. Appl. Math.*, pages 202–224. De Gruyter, Berlin, 2017.

[37] M Bertero, P Boccacci, G Talenti, R Zanella, and L Zanni. A discrepancy principle for Poisson data. *Inverse Problems*, 26(10):105004, aug 2010.

[38] D. A. Boas. A fundamental limitation of linearized algorithms for diffuse optical tomography. *Opt. Express*, 1(13):404–413, Dec 1997.

[39] David A Boas, Thomas Gaudette, and Simon R Arridge. Simultaneous imaging and optode calibration with diffuse optical tomography. *Optics express*, 8(5):263–270, 2001.

[40] Yoeri E. Boink and Christoph Brune. Learned SVD: solving inverse problems via hybrid autoencoding. *CoRR*, abs/1912.10840, 2019.

[41] E Bonnetier, M Choulli, and F Triki. Stability for quantitative photoacoustic tomography revisited. *Res. Math. Sci.*, 9(2):Paper No. 24, 30, 2022.

[42] A-P Calderón. On an inverse boundary value problem. In *Seminar on Numerical Analysis and its Applications to Continuum Physics (Rio de Janeiro, 1980)*, pages 65–73. Soc. Brasil. Mat., Rio de Janeiro, 1980.

[43] N Cao, A Nehorai, and M Jacob. Image reconstruction for diffuse optical tomography using sparsity regularization and expectation-maximization algorithm. *Opt. Express*, 15(21):13695–13708, Oct 2007.

[44] P Caro and KM Rogers. Global uniqueness for the Calderón problem with Lipschitz conductivities. *Forum Math. Pi*, 4:e2, 28, 2016.

[45] CM Carpenter and H Dehghani. In *European Conference on Biomedical Optics*, page 6629_56. Optica Publishing Group, 2007.

[46] KM Case and PF Zweifel. *Linear transport theory*. Addison-Wesley Educational Publishers Inc., US, 1967.

[47] JL Castellanos, S Gómez, and V Guerra. The triangle method for finding the corner of the l-curve. *Applied Numerical Mathematics*, 43(4):359–373, 2002.

[48] P Causin, MG Lupieri, G Naldi, and R-M Weishaeupl. Mathematical and numerical challenges in

optical screening of female breast. *International Journal for Numerical Methods in Biomedical Engineering*, 36(2):e3286, 2020.

[49] AE Cerussi, AJ Berger, F Bevilacqua, N Shah, D Jakubowski, J Butler, RF Holcombe, and BJ Tromberg. Sources of absorption and scattering contrast for near-infrared optical mammography. *Academic radiology*, 8(3):211–218, 2001.

[50] A Chambolle, V Caselles, D Cremers, M Novaga, and T Pock. An introduction to total variation for image analysis. *Theoretical foundations and numerical methods for sparse recovery*, 9(263-340):227, 2010.

[51] G Chen, F Zhu, and PA Heng. An efficient statistical method for image noise level estimation. In *2015 IEEE International Conference on Computer Vision (ICCV)*, pages 477–485, 2015.

[52] M Choulli. Some stability inequalities for hybrid inverse problems. *C. R. Math. Acad. Sci. Paris*, 359:1251–1265, 2021.

[53] M Choulli and P Stefanov. An inverse boundary value problem for the stationary transport equation. *Osaka Journal of Mathematics*, 36(1):87 – 104, 1999.

[54] BT Cox, SR Arridge, KP Köstli, and PC Beard. Two-dimensional quantitative photoacoustic image reconstruction of absorption distributions in scattering media by use of a simple iterative method. *Appl. Opt.*, 45(8):1866–1875, Mar 2006.

[55] J Curran, R Gaburro, C Nolan, and E Somersalo. Time-harmonic diffuse optical tomography: Hölder stability of the derivatives of the optical properties of a medium at the boundary. *Inverse Problems and Imaging*, pages 0–0, 2022.

[56] H Dehghani, S Srinivasan, BW Pogue, and A Gibson. Numerical modelling and image reconstruction in diffuse optical tomography. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 367(1900):3073–3093, 2009.

[57] B Deng, H Gu, and SA Carp. Deep learning enabled high-speed image reconstruction for breast diffuse optical tomography. In *Optical Tomography and Spectroscopy of Tissue XIV*, volume 11639, page 116390B. International Society for Optics and Photonics, 2021.

[58] O Doeva, R Gaburro, WRB Lionheart, and CJ Nolan. Lipschitz stability at the boundary for time-harmonic diffuse optical tomography. *Appl. Anal.*, 101(10):3697–3715, 2022.

[59] F Dubot, Y Favennec, B Rousseau, and DR Rousse. Regularization opportunities for the diffuse optical tomography problem. *International Journal of Thermal Sciences*, 98:1–23, 2015.

[60] P Elbau, L Mindrinos, and O Scherzer. Inverse problems of combined photoacoustic and optical coherence tomography. *Math. Methods Appl. Sci.*, 40(3):505–522, 2017.

[61] P Elbau, L Mindrinos, and O Scherzer. Quantitative reconstructions in multi-modal photoacoustic and optical coherence tomography imaging. *Inverse Problems*, 34(1):014006, 22, 2018.

[62] HW Engl, M Hanke, and A Neubauer. *Regularization of Inverse Problems*. Mathematics and Its Applications. Springer Netherlands, 1996.

[63] J Feng, Q Sun, Z Li, Z Sun, and K Jia. Back-propagation neural network-based reconstruction algorithm for diffuse optical tomography. *Journal of biomedical optics*, 24(5):051407, 2018.

[64] E Ferocino, A Pifferi, S Arridge, F Martelli, P Taroni, and A Farina. Multi simulation platform for time domain diffuse optical tomography: An application to a compact hand-held reflectance probe. *Applied Sciences*, 9(14), 2019.

[65] S Foschiatti. Lipschitz stability estimate for the simultaneous recovery of two coefficients in the anisotropic Schrödinger type equation via local Cauchy data. *J. Math. Anal. Appl.*, 531(1):Paper No. 127753, 2024.

[66] S Foschiatti and E Sincich. Stable determination of an anisotropic inclusion in the Schrödinger equation from local Cauchy data. *Inverse Probl. Imaging*, 17(3):584–613, 2023.

[67] Raymond J Fretterd and Richard L Longini. Diffusion dipole source. *JOSA*, 63(3):336–337, 1973.

[68] Jerome Friedman, Robert Tibshirani, and Trevor Hastie. Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, 33(1):1–22, 2010.

[69] R Gaburro. Stable determination at the boundary of the optical properties of a medium: the static case. *Rend. Istit. Mat. Univ. Trieste*, 48:407–431, 2016.

[70] H Gao, S Osher, and H Zhao. Quantitative photoacoustic tomography. In *Mathematical modeling in biomedical imaging. II*, volume 2035 of *Lecture Notes in Math.*, pages 131–158. Springer, Heidelberg, 2012.

[71] AP Gibson, JC Hebden, and SR Arridge. Recent advances in diffuse optical imaging. *Physics in medicine & biology*, 50(4):R1, 2005.

[72] GH Golub and U Von Matt. Generalized cross-validation for large-scale problems. *Journal of Computational and Graphical Statistics*, 6(1):1–34, 1997.

[73] I Goodfellow, Y Bengio, and A Courville. *Deep learning*. Springer, 2016.

[74] B Haberman and D Tataru. Uniqueness in Calderón's problem with Lipschitz conductivities. *Duke Math. J.*, 162(3):496–516, 2013.

[75] J Hadamard. Sur les problèmes aux dérivés partielles et leur signification physique. *Princeton University Bulletin*, 13:49–52, 1902.

[76] M Haltmeier, L Neumann, L Nguyen, and S Rabanser. Analysis of the linearized problem of quantitative photoacoustic tomography. *SIAM J. Appl. Math.*, 78(1):457–478, 2018.

[77] M Haltmeier, L Neumann, and S Rabanser. Single-stage reconstruction algorithm for quantitative photoacoustic tomography. *Inverse Problems*, 31(6):065005, 24, 2015.

[78] M Hanke. Limitations of the l-curve method in ill-posed problems. *BIT Numerical Mathematics*, 36(2):287–301, 1996.

[79] PC Hansen, TK Jensen, and G Rodriguez. An adaptive pruning algorithm for the discrete l-curve criterion. *Journal of Computational and Applied Mathematics*, 198(2):483–492, 2007. Special Issue: Applied Computational Inverse Problems.

[80] PC Hansen, JG Nagy, and DP O'Leary. *Deblurring Images: Matrices, Spectra, and Filtering*. Fundamentals of Algorithms. Society for Industrial and Applied Mathematics, 2006.

[81] PC Hansen and DP O'Leary. The use of the l-curve in the regularization of discrete ill-posed problems. *SIAM Journal on Scientific Computing*, 14(6):1487–1503, 1993.

[82] B Harrach. On uniqueness in diffuse optical tomography. *Inverse Problems*, 25(5):055010, 14, 2009.

[83] B Harrach. Simultaneous determination of the diffusion and absorption coefficient from boundary data. *Inverse Probl. Imaging*, 6(4):663–679, 2012.

[84] B Harrach. Uniqueness and Lipschitz stability in electrical impedance tomography with finitely many electrodes. *Inverse Problems*, 35(2):024005, 19, 2019.

[85] J Heino, S Arridge, J Sikora, and E Somersalo. Anisotropic effects in highly scattering media. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, 68(3 1):319081–319088, 2003. Cited By :117.

[86] J Heino and E Somersalo. Estimation of optical absorption in anisotropic background. *Inverse Problems*, 18(3):559–573, 2002.

[87] LG Henyey and JL Greenstein. Diffuse radiation in the galaxy. *Astrophysical Journal*, 93:70–83, 1941.

[88] AH Hielscher, AY Bluestone, GS Abdoulaev, AD Klose, J Lasker, M Stewart, U Netz, and J Beuthan. Near-infrared diffuse optical tomography. *Disease markers*, 18(5-6):313–337, 2002.

[89] Y Hoshi and Y Yamada. Overview of diffuse optical tomography and its clinical applications. *Journal of biomedical optics*, 21(9):091312, 2016.

[90] M Ikeda, R Makino, K Imai, M Matsumoto, and R Hitomi. A method for estimating noise variance of ct image. *Computerized Medical Imaging and Graphics*, 34(8):642–650, 2010.

[91] V Isakov. *Inverse problems for partial differential equations*, volume 127 of *Applied Mathematical Sciences*. Springer, Cham, third edition, 2017.

[92] A Ishimaru. Theory and application of wave propagation and scattering in random media. *Proceedings of the IEEE*, 65(7):1030 – 1061, 1977.

[93] SL Jacques, CA Alter, and SA Prahl. Angular dependence of HeNe laser light scattering by

human dermis. *Lasers Life Sci.*, 1:309–333, 1987.

[94] RPK Jagannath and PK Yalavarthy. Minimal residual method provides optimal regularization parameter for diffuse optical tomography. *Journal of Biomedical Optics*, 17(10):1 – 7, 2012.

[95] AK Jha, Y Zhu, S Arridge, DF Wong, and A Rahmim. Incorporating reflection boundary conditions in the neumann series radiative transport equation: application to photon propagation and reconstruction in diffuse optical imaging. *Biomed. Opt. Express*, 9(4):1389–1407, Apr 2018.

[96] J Jiang, Y Li, and R Guo. Learn an index operator by cnn for solving diffusive optical tomography: a deep direct sampling method. *arXiv preprint arXiv:2104.07703*, 2021.

[97] B Jin, Y Zhao, and J Zou. Iterative parameter choice by discrepancy principle. *IMA Journal of Numerical Analysis*, 32(4):1714–1732, 02 2012.

[98] J Kaipio and E Somersalo. *Statistical and computational inverse problems*, volume 160 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2005.

[99] B Kaltenbacher. Regularization based on all-at-once formulations for inverse problems. *SIAM J. Numer. Anal.*, 54(4):2594–2618, 2016.

[100] R Kohn and M Vogelius. Determining conductivity by boundary measurements. *Comm. Pure Appl. Math.*, 37(3):289–298, 1984.

[101] RV Kohn and M Vogelius. Determining conductivity by boundary measurements. II. Interior results. *Comm. Pure Appl. Math.*, 38(5):643–667, 1985.

[102] Soren D Konecky, George Y Panasyuk, Kijoon Lee, Vadim Markel, Arjun G Yodh, and John C Schotland. Imaging complex structures with diffuse light. *Optics express*, 16(7):5048–5060, 2008.

[103] F Larusson, S Fantini, and EL Miller. Hyperspectral image reconstruction for diffuse optical tomography. *Biomed. Opt. Express*, 2(4):946–965, Apr 2011.

[104] K Lee. Optical mammography: Diffuse optical imaging of breast cancer. *World journal of clinical oncology*, 2(1):64, 2011.

[105] DR Leff, OJ Warren, LC Enfield, A Gibson, T Athanasiou, DK Patten, J Hebden, GZ Yang, and A Darzi. Diffuse optical imaging of the healthy and diseased breast: a systematic review. *Breast cancer research and treatment*, 108(1):9–22, 2008.

[106] J Li, Z Xie, G Liu, L Yang, and J Zou. Diffusion optical tomography reconstruction based on convex–nonconvex graph total variation regularization. *Mathematical Methods in the Applied Sciences*, 46(4):4534–4545, 2023.

[107] L Li, C Chen, and B Bi. A new total variational regularization method for nonlinear inverse problems in fluorescence molecular tomography. *Journal of Computational and Applied Mathematics*, 365:112408, 2020.

[108] PA Lo and HK Chiang. Three-dimensional fluorescence diffuse optical tomography using the adaptive spatial prior approach. *Journal of Medical and Biological Engineering*, 39(6):827–834, 2019.

[109] JR Lorenzo. *Principles of diffuse light propagation: light propagation in tissues with applications in biology and medicine*. World Scientific, 2012.

[110] W Lu, J Duan, D Orive-Miguel, L Herve, and IB Styles. Graph- and finite element-based total variation models for the inverse problem in diffuse optical tomography. *Biomed. Opt. Express*, 10(6):2684–2707, Jun 2019.

[111] Manabu Machida. The inverse rytov series for diffuse optical tomography. *Inverse Problems*, 39(10):105012, 2023.

[112] M Mozumder, A Hauptmann, I Nissilä, SR Arridge, and T Tarvainen. A model-based iterative learning approach for diffuse optical tomography. *arXiv preprint arXiv:2104.09579*, 2021.

[113] AI Nachman. Global uniqueness for a two-dimensional inverse boundary value problem. *Ann. of Math. (2)*, 143(1):71–96, 1996.

[114] W Naetar and O Scherzer. Quantitative photoacoustic tomography with piecewise constant material parameters. *SIAM J. Imaging Sci.*, 7(3):1755–1774, 2014.

[115] I Nissilä, T Noponen, J Heino, T Kajava, and T Katila. *Diffuse Optical Imaging*, pages 77–129. Springer US, Boston, MA, 2005.

[116] Shinpei Okawa and Yoko Hoshi. A review of image reconstruction algorithms for diffuse optical tomography. *Applied Sciences*, 13(8):5016, 2023.

[117] R Patra and PK Dutta. Improved dot reconstruction by estimating the inclusion location using artificial neural network. In *Medical Imaging 2013: Physics of Medical Imaging*, volume 8668, page 86684C. International Society for Optics and Photonics, 2013.

[118] A Pulkkinen, BT Cox, SR Arridge, JP Kaipio, and T Tarvainen. A Bayesian approach to spectral quantitative photoacoustic tomography. *Inverse Problems*, 30(6):065012, 18, 2014.

[119] A Pulkkinen, V Kolehmainen, JP Kaipio, BT Cox, SR Arridge, and T Tarvainen. Approximate marginalization of unknown scattering in quantitative photoacoustic tomography. *Inverse Probl. Imaging*, 8(3):811–829, 2014.

[120] J Radford, A Lyons, F Tonolini, and D Faccio. Role of late photons in diffuse optical imaging. *Opt. Express*, 28(20):29486–29495, Sep 2020.

[121] K Ren, G Bal, and AH Hielscher. Frequency domain optical tomography based on the equation of radiative transfer. *SIAM J. Sci. Comput.*, 28(4):1463–1489, 2006.

[122] K Ren, H Gao, and H Zhao. A hybrid reconstruction method for quantitative PAT. *SIAM J. Imaging Sci.*, 6(1):32–55, 2013.

[123] Julia L Sandell and Timothy C Zhu. A review of in-vivo optical properties of human tissues and its impact on pdt. *Journal of biophotonics*, 4(11-12):773–787, 2011.

[124] O Scherzer, editor. *Handbook of mathematical methods in imaging. Vol. 1, 2, 3.* Springer, New York, second edition, 2015.

[125] O Scherzer, M Grasmair, H Grossauer, M Haltmeier, and F Lenzen. *Variational methods in imaging*, volume 167 of *Applied Mathematical Sciences*. Springer, New York, 2009.

[126] Raymond A Schulz, Jay A Stein, and Norbert J Pelc. How ct happened: the early development of medical computed tomography. *Journal of Medical Imaging*, 8(5):052110–052110, 2021.

[127] M Schweiger, SR Arridge, M Hiraoka, and DT Delpy. The finite element method for the propagation of light in scattering media: Boundary and source conditions. *Medical physics*, 22(11):1779–1792, 1995. Cited By :569.

[128] H Singh, RJ Cooper, CW Lee, L Dempsey, A Edwards, S Brigadoi, D Airantzis, N Everdell, A Michell, D Holder, et al. Mapping cortical haemodynamics during neonatal seizures using diffuse optical tomography: a case study. *NeuroImage: Clinical*, 5:256–265, 2014.

[129] S Srinivasan, BW Pogue, S Jiang, H Dehghani, C Kogel, S Soho, JJ Gibson, TD Tosteson, SP Poplack, and KD Paulsen. Interpreting hemoglobin and water concentration, oxygen saturation, and scattering measured in vivo by near-infrared breast tomography. *Proceedings of the National Academy of Sciences*, 100(21):12349–12354, 2003.

[130] Y Sun, Z Xia, and US Kamilov. Efficient and accurate inversion of multiple scattering with deep learning. *Optics express*, 26(11):14678–14688, 2018.

[131] Z Sun, Y Wang, K Jia, and J Feng. Comprehensive study of methods for automatic choice of regularization parameter for diffuse optical tomography. *Optical Engineering*, 56(4):041310, 2016.

[132] J Sylvester and G Uhlmann. A global uniqueness theorem for an inverse boundary value problem. *Ann. of Math. (2)*, 125(1):153–169, 1987.

[133] J Tang. Nonconvex and nonsmooth total variation regularization method for diffuse optical tomography based on RTE*. *Inverse Problems*, 37(6):065001, may 2021.

[134] J Tang, B Han, W Han, B Bi, and L Li. Mixed total variation and regularization method for optical tomography based on radiative transfer equation. *Computational and mathematical methods in medicine*, 2017, 2017.

[135] T Tarvainen, BT Cox, JP Kaipio, and SR Arridge. Reconstructing absorption and scattering distributions in quantitative photoacoustic tomography. *Inverse Problems*, 28(8):084009, 17, 2012.

[136] L Tian, B Hunt, MAL Bell, J Yi, JT Smith, M Ochoa, X Intes, and NJ Durr. Deep learning in biomedical optics. *Lasers in Surgery and Medicine*, 53(6):748–775, 2021.

[137] E Turajlic. Adaptive svd domain-based white gaussian noise level estimation in images. *IEEE Access*, 6:72735–72747, 2018.

[138] KMS Uddin, A Mostofa, M Anastasio, and Q Zhu. Improved two step reconstruction method in ultrasound guided diffuse optical tomography. In *Biophotonics Congress: Biomedical Optics Congress 2018 (Microscopy/Translational/Brain/OTS)*, page OF1D.5. Optica Publishing Group, 2018.

[139] G Uhlmann. Electrical impedance tomography and Calderón's problem. *Inverse Problems*, 25(12):123011, 39, 2009.

[140] JM Varah. Pitfalls in the numerical solution of linear ill-posed problems. *SIAM Journal on Scientific and Statistical Computing*, 4(2):164–176, 1983.

[141] CR Vogel. Non-convergence of the l-curve regularization parameter selection method. *Inverse Problems*, 12(4):535–547, aug 1996.

[142] B Wang, Y Zhang, D Liu, X Ding, M Dan, T Pan, H Zhao, and F Gao. Sparsity-regularized approaches to directly reconstructing hemodynamic response in brain functional diffuse optical tomography. *Appl. Opt.*, 58(4):863–870, Feb 2019.

[143] L Wang and SL Jacques. Hybrid model of monte carlo simulation and diffusion theory for light reflectance by turbid media. *J. Opt. Soc. Am. A*, 10(8):1746–1752, Aug 1993.

[144] LV Wang and H-I Wu. *Biomedical optics: principles and imaging*. John Wiley & Sons, 2012.

[145] J Xie and J Zou. An improved model function method for choosing regularization parameters in linear inverse problems. *Inverse Problems*, 18(3):631–643, apr 2002.

[146] HB Yedder, B Cardoen, M Shokoufi, F Golnaraghi, and G Hamarneh. Multitask deep learning reconstruction and localization of lesions in limited angle diffuse optical tomography. *IEEE Transactions on Medical Imaging*, 2021.

[147] W Yin, S Osher, D Goldfarb, and J Darbon. Bregman iterative algorithms for ℓ_1-minimization with applications to compressed sensing. *SIAM Journal on Imaging Sciences*, 1(1):143–168, 2008.

[148] J Yoo, S Sabir, D Heo, KH Kim, A Wahab, Y Choi, S-I Lee, EY Chae, HH Kim, YM Bae, et al. Deep learning diffuse optical tomography. *IEEE transactions on medical imaging*, 39(4):877–887, 2019.

[149] L Zanni, A Benfenati, M Bertero, and V Ruggiero. Numerical methods for parameter estimation in Poisson data inversion. *Journal of Mathematical Imaging and Vision*, 52(3):397 – 413, 2015.

[150] Y Zha, A Raghuram, HK Kim, AH Hielscher, JT Robinson, and A Veeraraghavan. High resolution, deep imaging using confocal time-of-flight diffuse optical tomography. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(7):2206–2219, 2021.

[151] Y Zhao, MA Mastanduno, S Jiang, F El-Ghussein, J Gui, BW Pogue, and KD Paulsen. Optimization of image reconstruction for magnetic resonance imaging–guided near-infrared diffuse optical spectroscopy in breast. *Journal of Biomedical Optics*, 20(5):1 – 10, 2015.

[152] P Zimmermann. Inverse problem for a nonlocal diffuse optical tomography equation. *Inverse Problems*, 39(9):094001, jul 2023.

[153] H Zou and T Hastie. Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 67(2):301–320, 2005.

[154] Y Zou, Y Zeng, S Li, and Q Zhu. Machine learning model with physical constraints for diffuse optical tomography. *Biomedical Optics Express*, 12(9):5720–5735, 2021.

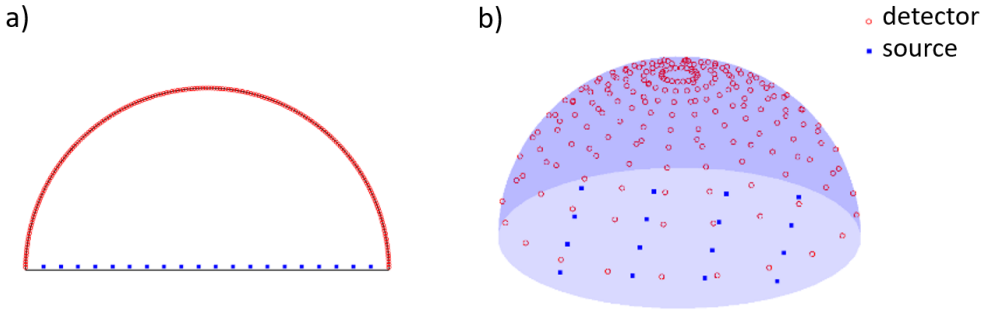

Figure 3. Geometry and position of sources and detectors used in the numerical tests in the: (a) 2D setting and (b) 3D setting.

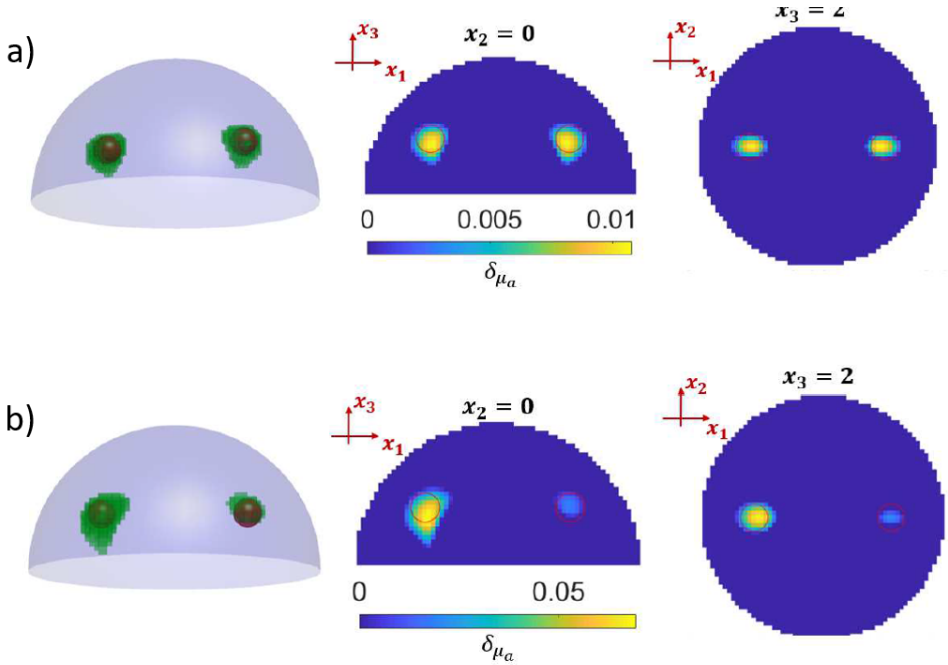

Figure 4. Reconstruction with contrast regions with the same (a) or different (b) strength. In the second case the algorithm struggles to obtain a good quality both in terms of localization and intensity.

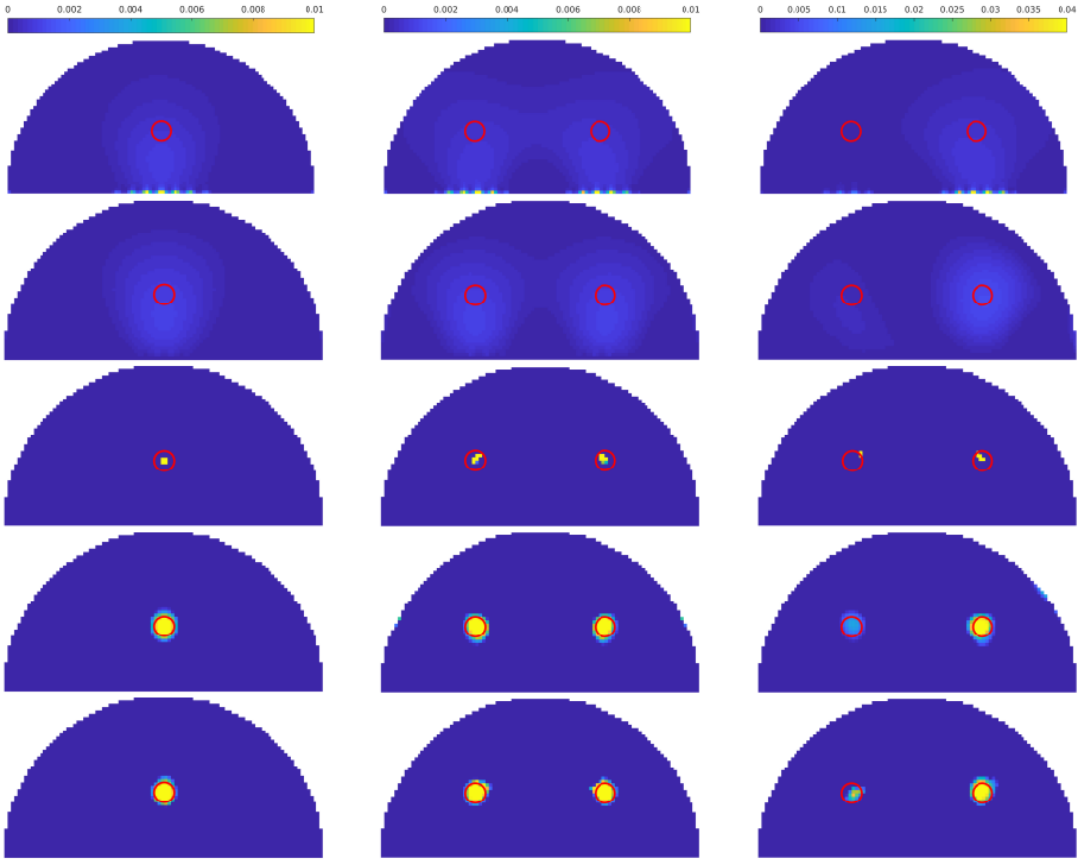

Figure 5. Results for test cases i), ii), iii) in column–wise order. Top to bottom: Tikhonov regularization, Bregman procedure coupled with 2–norm, pure LASSO approach, Bregman coupled with 1–norm and elastic–net procedure with $\alpha = 0.5$.

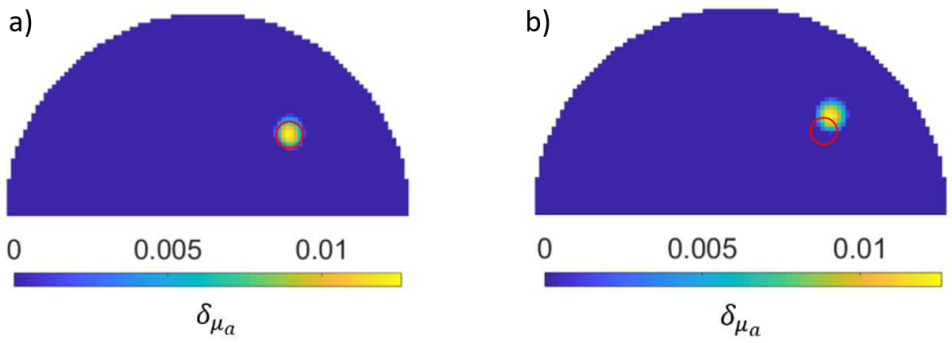

Figure 6. Results in the 2D test setting with: (a) no enforcement of Robin boundary conditions on the background solution; (b): enforcement of Robin boundary conditions on the background solution via the MFS strategy. The red circle represents the exact location of the contrast region.

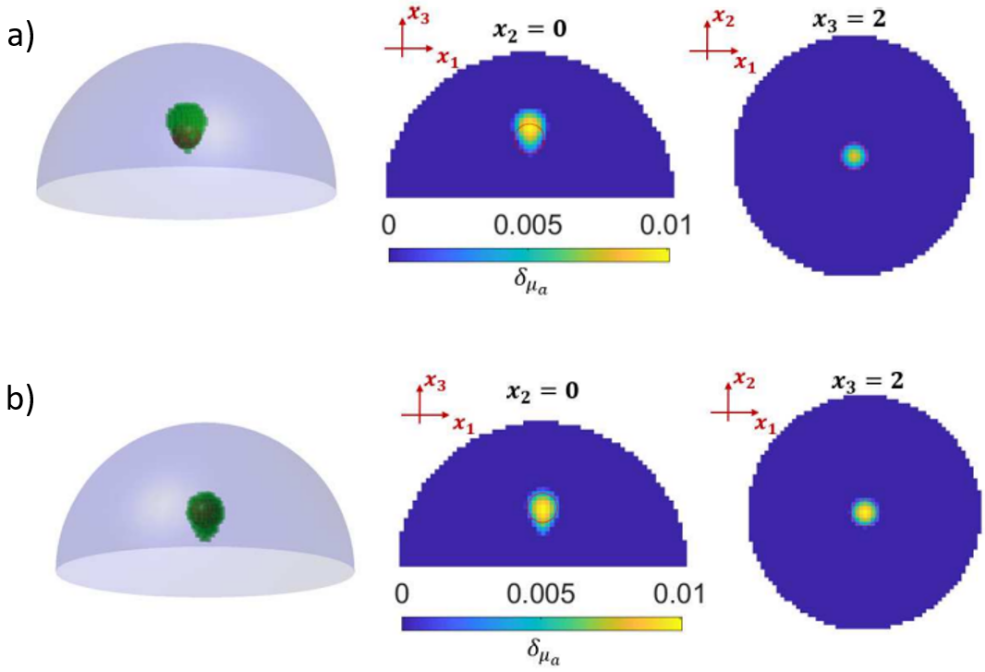

Figure 7. Results in the 3D test setting with: (a) no enforcement of Robin boundary conditions on the background solution; (b): enforcement of Robin boundary conditions on the background solution via the MFS strategy. The red circle represents the exact location of the contrast region.

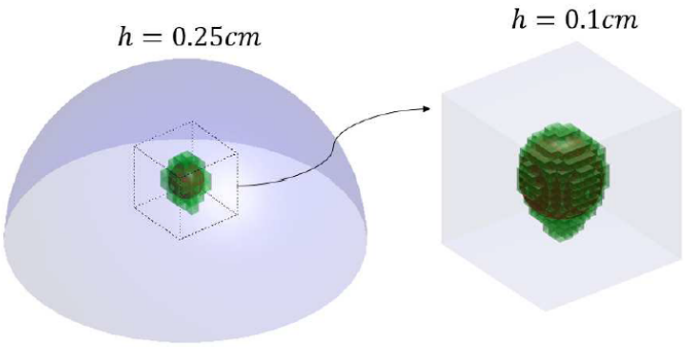

Figure 8. Improvement (right) of the spatial resolution in the 3D geometry using as soft-prior a first rough estimation of the position of the contrast region from a coarse voxelization (left).

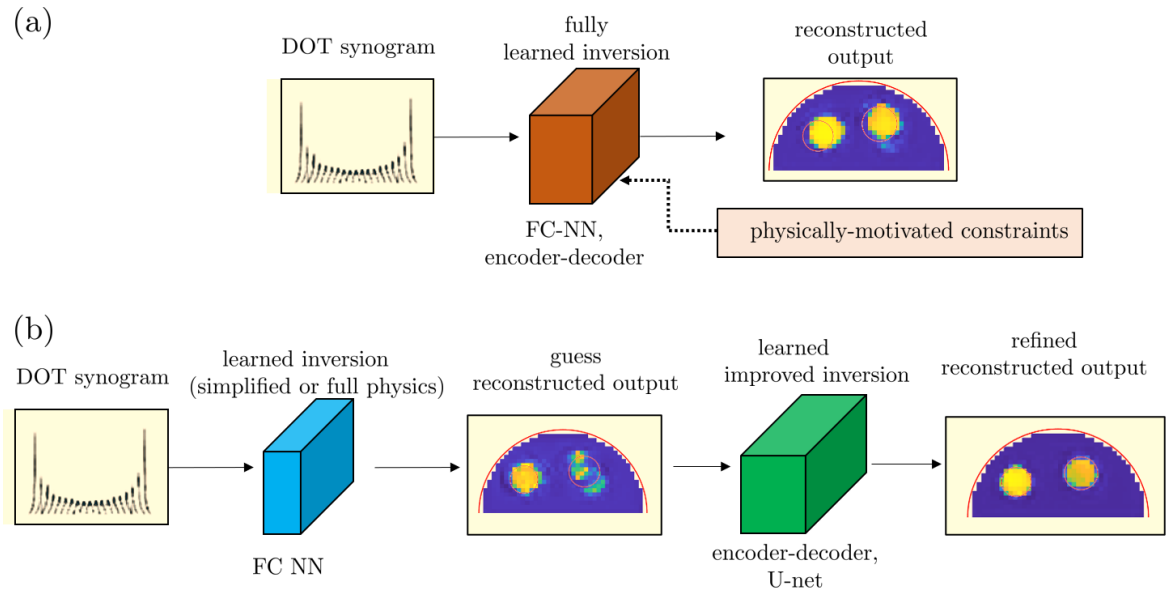

Figure 9. Classes of existing DL-based approach for DOT reconstruction: a) methods of this family are fully data-driven. Sometimes the loss function used in the training of the net can be enriched with physically motivated constraints; b) methods of this family partially rely on physical models that can produce a first guess for the solution and then are supported by nets which improve it.