.

Inverse Problems and Dynamical Systems in Tomography and Optics

Lauri Ylinen

Doctoral dissertation

To be presented for public discussion with the permission of the Faculty of Science of the University of Helsinki, in Auditorium PIII, Porthania Building, on the 23rd of November, 2022 at 12 o'clock.

Department of Mathematics and Statistics
Faculty of Science
University of Helsinki

Helsinki 2022

Advisor:
Academy Professor Matti Lassas, University of Helsinki

Pre-Examiners:
Professor Otmar Scherzer, University of Vienna
Associate Professor Lassi Roininen, LUT University

Opponent:
Professor Jaan Janno, Tallinn University of Technology
Custos:
Academy Professor Matti Lassas, University of Helsinki

ISBN 978-951-51-8718-5 (softcover)
ISBN 978-951-51-8719-2 (PDF)
Unigrafia
Helsinki 2022

Abstract

The problem of determining an unknown physical quantity from indirect measurements is called an inverse problem. A mathematical model that describes how a quantity evolves in time is called a dynamical system. This thesis is about inverse problems and dynamical systems.

Inverse problems research, as a subfield of mathematics, studies the mathematical theory of indirect measurements. It is an active area of research with extensive mathematical theory and numerous applications. Many inverse problems are concerned with physical systems that depend on time. These problems are called dynamical inverse problems. Dynamical systems research, by itself, is a wide field of mathematic that has its origins in the study of celestial mechanics.

In the first publication of this thesis, we consider an inverse problem in X-ray computed tomography. This is a technique where a cross-section of an object is imaged by X-ray measurements from different angles. The classical inverse problem is to reconstruct the unknown object from these measurements. In the classical problem it is assumed that the angles are known, in the problem we consider both the object being imaged and the angles from which it was imaged are unknown. This problem is called tomography with unknown view angles. This kind of problem may arise due to unknown orientation of the object being imaged, as is the case in cryogenic electron microscopy of viral particles, for example.

In the first publication we show that under certain assumptions it is possible to reconstruct the unknown object from tomography with unknown view angles. We also derive explicit conditions under which this is true. The proofs of these results use the Helgason–Ludwig consistency conditions for the Radon transform and results from algebraic geometry.

The second publication of the thesis considers dynamical inverse problems for diffusion and anomalous diffusion. Diffusion is the flow of some substance from areas of high concentration to areas of low concentration, and the diffusion equation is a mathematical model of this process. The space–time fractional diffusion equation is a generalization of the diffusion equation that covers anomalous diffusion processes, such as those sometimes observed in porous or fractured geological formations.

In the second publication we consider an inverse problem for the space–time fractional diffusion equation in the geometric setting. This means that the unknown is a Riemannian manifold on which the space–time fractional diffusion equation is defined. The measurement corresponds to applying a source supported on a fixed open set of the manifold, and observing the restriction of the corresponding solution for the equation on the same set. We show that it is possible to construct a source such that a single measurement of this type determines the manifold uniquely.

In the third and fourth publications of the thesis we consider a dynamical system that models injection locking in a laser. Injection locking is a technique where light from one laser is injected into another laser's cavity (the part of a laser where the

emitted light is created) with the intention of altering the laser's properties. The idea in the publications is to consider injection locking as a process that can provide the basis for optical computing devices.

In the third publication we derive an approximation for the nonlinear relationship between the injected light and the injection-locked emitted light, and we show that it is possible to construct an optical logic gate based on this relationship. In the fourth publication we do a detailed analysis of the dynamical system that models injection locking in lasers, and based on this analysis, we propose a design for an optical neural network.

Acknowledgments

I express my deepest gratitude to my advisor Matti Lassas whose guidance and support during these years has been invaluable. I am deeply grateful for the opportunity to be part of the inverse problems research group of the University of Helsinki.

I also wish to thank my other collaborators for the pleasant cooperation. In particular, I wish to thank Tapio Helin, Lars Lamberg, and Tuomo von Lerber for all the inspiring discussions. I also thank Lassi Roininen and Otmar Scherzer for their careful pre-examination of my thesis, and Jaan Janno for kindly agreeing to be my opponent.

During the many years I have been involved in inverse problems research I have had the pleasure to work with a lot of people. In particular, I wish to thank Lassi Päivärinta and Gunther Uhlmann for their advice and support during the early years of my studies. I am also indebted to all my friends and coworkers in Kumpula, and my colleagues in the Finnish Inverse Problems Society.

I am grateful to my daughter Tiia and her mother Malla, my parents Satu and Aarne, and my siblings Artturi, Tyyne and Väinö for their unconditional support.

Finally, I thank the Academy of Finland and the Centre of Excellence of Inverse Modelling and Imaging for financial support.

List of Included Publications

I L. Lamberg and L. Ylinen. "Two-dimensional tomography with unknown view angles". In: *Inverse Probl. Imaging* 1.4 (2007), pp. 623–642. DOI: 10.3934/ipi.2007.1.623

II T. Helin, M. Lassas, L. Ylinen, and Z. Zhang. "Inverse problems for heat equation and space–time fractional diffusion equation with one measurement". In: *J. Differential Equations* 269.9 (2020), pp. 7498–7528. DOI: 10.1016/j.jde.2020.05.022

III T. von Lerber, M. Lassas, V. S. Lyubopytov, L. Ylinen, A. Chipouline, K. Hofmann, and F. Küppers. "All-optical majority gate based on an injection-locked laser". In: *Scientific reports* 9.1 (2019), pp. 1–7. DOI: 10.1038/s41598-019-51025-y

IV L. Ylinen, T. von Lerber, F. Küppers, and M. Lassas. "Analysis of a dynamical system modeling lasers and applications for optical neural networks". In: *SIAM J. Appl. Dyn. Syst.* 21.2 (2022), pp. 840–878. DOI: 10.1137/21M1405976

Author's Contribution

I The authors had an equal role in proving the theoretical results. LY made major contribution in writing the paper.

II LY had a leading role in analyzing the direct problem and the single measurement problem. The authors had an equal role in the analysis of the inverse problem with multiple measurements. LY made major contribution in writing the proof sections of the paper.

III LY helped to derive the equation for the steady state solution. Simulations of the full adder and ripple-carry adder were programmed by LY.

IV LY had a leading role in mathematical analysis, simulation, and writing of the article.

Contents

1 Introduction

1.1 Inverse problems

A ubiquitous problem in the natural sciences is to infer a physical quantity that cannot be directly measured from quantities that can be directly measured. For example, in geophysics one wants to know the deep structure of Earth, but the deep structure is inaccessible and cannot be directly measured. Consequently, one has to content with indirect measurements on the surface of Earth, like measuring the propagation of seismic waves caused by earthquakes. The seismic waves are affected by the medium where they propagate, and since some of the waves have traveled deep into the mantle of Earth before surfacing, they bring with them information from the deep.

The problem of determining a quantity from indirect measurements is called an *inverse problem*. Inverse problems research, as a subfield of mathematics, studies the mathematical theory of indirect measurements.

In the context of inverse problems the mathematical model of a measurement is called a *direct problem* (also called a *forward problem*). In the above example the mathematical model is the equation that governs the propagation of seismic waves inside Earth. Solving the direct problem amounts to finding what the observations (seismic waves on the surface) would be given the process that produces them (an earthquake and the equation for wave propagation inside Earth). The inverse problem goes in the opposite direction: From the observation (seismic waves on the surface) the problem is to recover the equation that governed their propagation. Usually the general structure of the direct problem is known from a physical model of the measurement, but some specific information relevant to the particular situation is unknown, such as the variable speed of wave propagation inside Earth.

A direct problem may be written in terms of an integral transform, or a distance function on a boundary of a manifold, or a partial differential equation, for example. An example of the last case is the Calderón problem—a problem that has motivated many developments in the inverse problems. In 1980 Calderón published a paper [19] where he asked whether it is possible to determine the electrical conductivity inside a domain by making voltage and current measurements on the boundary. Given a potential f on the boundary of a bounded domain $\Omega \subset \mathbb{R}^N$, the potential u inside the domain satisfies the Dirichlet problem for the conductivity equation:

$$\nabla \cdot (\gamma \nabla u) = 0 \text{ in } \Omega,\ u|_{\partial\Omega} = f. \tag{1}$$

Here $\gamma(x)$ is the conductivity (a positive function). Equation (1) is the direct problem, its general form can be derived from the physical principles, but the conductivity γ depends on the particular domain. The inverse problem Calderón posed is whether the map

$$f \mapsto \int_\Omega \gamma |\nabla u|^2 \, dx \tag{2}$$

uniquely determines the conductivity γ, and if so, how to calculate it. Often the problem is posed in an equivalent form where (2) is replaced by the Dirichlet-to-Neumann map $f \mapsto (\gamma \partial_\nu u)|_{\partial\Omega}$, where ∂_ν is the boundary normal derivative.

In dimensions three and higher Calderón's question about the uniqueness was answered in the positive in 1987 in a breakthrough by Sylvester and Uhlmann [122]. They showed that a smooth conductivity is indeed uniquely determined by the map (2). On the plane the question of uniqueness was solved in the positive in 2006 by Astala and Päivärinta [6]. Their result only assumes L^∞-regularity of the conductivity. Calderón's second question, namely, how to actually calculate the conductivity from the measurements, is still an active field of research. The measurement method of recovering the conductivity from the boundary measurements is known as electrical impedance tomography.

During the decades since the breakthrough by Sylvester and Uhlmann the field of inverse problems has grown into an active area of research with extensive mathematical theory and many subfields of its own. The success of inverse problems as a research field can in part be contributed to the endless supply of problems from the many fields of sciences that rely on indirect measurements, like medical imaging, astronomy, and remote sensing. On the other hand, the theory of inverse problems contains deep and beautiful mathematics that gratifies those who find the mathematical beauty as important as its potential application.

A rich source of deep and beautiful mathematics for inverse problems—at least in the author's opinion—have been the *geometric inverse problems*, which have become an important subfield of inverse problems research. It was observed already in 1989 by Lee and Uhlmann [81] that a generalization of the problem posed by Calderón has a geometric nature. Calderón's original question assumed an isotropic (equal in every direction) conductivity, but the same question can be asked for anisotropic conductivities by replacing the scalar-valued function γ in (1) by a symmetric, positive definite matrix-valued function. Lee and Uhlmann realized that in three and higher dimensions the equation can then be written equivalently as

$$\Delta_g u = 0 \text{ in } M,\ u|_{\partial M} = f. \tag{3}$$

Here (M, g) is a compact Riemannian manifold with boundary ∂M, the Riemannian metric g is related to the anisotropic conductivity γ, and Δ_g is the Laplace–Beltrami operator on the manifold. The geometric formulation allows one to take the manifold M (with its topology, differentiable structure, and Riemannian metric) as the unknown, and to ask whether the manifold is determined uniquely by the boundary measurements. Calderón's problem for anisotropic conductivity was solved in the plane by Astala, Päivärinta, and Lassas [7], but, as of the time of writing, in higher dimensions the problem is open. It is considered to be one of the major open problems in the field.

In this thesis we consider two types of inverse problems. The first inverse problem is a problem with direct problem given in terms of an integral transform called

the Radon transform. This is the content of the next section. The second inverse problem is a geometric inverse problem with direct problem given in terms of a (fractional) partial differential equation—the so-called space–time fractional diffusion equation. This problem will be introduced in section 1.3.

1.2 The Radon transform and the X-ray transform

In 1917 Radon introduced an integral transform that takes a function f on the plane to a function $\mathcal{R}f$ on the space of lines in the plane [100] (see [99] for an English translation). For a line ℓ he defined the value of $\mathcal{R}f(\ell)$ as the integral of f over ℓ. This transform is now known as the *Radon transform*. An analogous transform for a function on the sphere had been introduced earlier [37], but the Radon transform has become the most influential transform of its type.

Main reason for the present-day importance of the Radon transform is its relevance to imaging applications. Computed tomography (CT) is a medical imaging technique where multiple X-ray images of a cross-section of a body are measured in various directions, and these one-dimensional images are processed on a computer to reconstruct a two-dimensional image of the (X-ray attenuation coefficient of the) cross-section of the body [62]. In a mathematical idealization of CT, the X-ray images determine the Radon transform of the cross-section of the body, and the reconstruction corresponds to inverting the Radon transform. The first clinical CT scan was performed in 1971, however, the relevance of Radon's work for the application was realized only later [27, 62].

In N dimensions the Radon transform is defined by

$$\mathcal{R}f(\theta, s) := \int_{\theta^\perp} f(s\theta + y)\, dy,$$

where $\theta \in S^{N-1}$, the $(N-1)$-dimensional unit sphere of $\mathbb{R}^N$, and $s \in \mathbb{R}$. The direction θ is sometimes called the *view angle*. The integral is over $\theta^\perp$, the hyperplane of $\mathbb{R}^N$ perpendicular to θ, and it is with respect to the Euclidean measure. The function $f : \mathbb{R}^N \to \mathbb{R}$ is taken from some suitable function space in which the integral makes sense. The other obvious generalization of Radon's original transform to higher dimensions, namely the one where one integrates over lines instead of hyperplanes, is called the *X-ray transform*. For $(\theta, x) \in S^{N-1} \times \mathbb{R}^N$ and for a suitable function $f : \mathbb{R}^N \to \mathbb{R}$, the X-ray transform of f is

$$\mathcal{P}f(\theta, x) := \int_{\mathbb{R}} f(x + t\theta)\, dt. \tag{4}$$

Because the integral in (4) does not change if x moves along the direction θ, the x-variable is often restricted to $\theta^\perp$. In two dimensions $\mathcal{R}$ and $\mathcal{P}$ coincide, except for the notation of their arguments.

For a mathematical exposition on the Radon transform, the X-ray transform, and their applications, we refer the reader to [49, 92, 30]. For an illustration of the

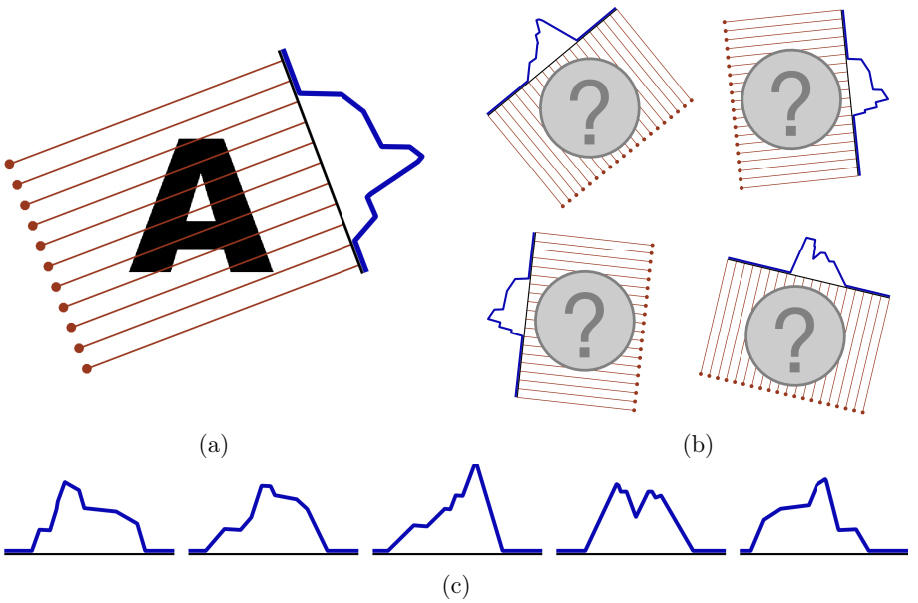


Figure 1: Illustration of the Radon transform and two related inverse problems.[1] (a) Radon transform of the characteristic function of the letter 'A' in one direction. Values of the transform (in blue) are obtained by integrating along the parallel lines. (b) The inverse problem of determining an unknown function from a set of its Radon transforms in known directions. (c) The inverse problem of determining an unknown function from a set of its Radon transforms in unknown directions.

Radon transform and two related inverse problems, see figure 1.

The Radon and X-ray transforms enjoy nice mapping properties. For example, the Radon transform maps the Schwartz space $\mathcal{S}(\mathbb{R}^N)$ into the Schwartz space $\mathcal{S}(S^{N-1} \times \mathbb{R})$. Between these spaces the range of the Radon transform admits an elegant characterization: If $f \in \mathcal{S}(\mathbb{R}^N)$, then $\mathcal{R}f$ is an even function, i.e., $\mathcal{R}f(-\theta, -s) = \mathcal{R}f(\theta, s)$, and for $m = 0, 1, 2, \ldots$

$$\begin{aligned} \int_{\mathbb{R}} s^m \mathcal{R}f(\theta, s)\, ds &= \int_{\mathbb{R}} s^m \int_{\theta^\perp} f(s\theta + y)\, dy\, ds \\ &= \int_{\mathbb{R}^N} (x \cdot \theta)^m f(x)\, dx \\ &= \sum_{|\alpha|=m} \binom{|\alpha|}{\alpha} \theta^\alpha \int_{\mathbb{R}^N} x^\alpha f(x)\, dx =: p_m(\theta), \end{aligned} \tag{5}$$

where p_m is a homogeneous polynomial of degree m. The second equality above follows from the Fubini's theorem and the change of variable $x = s\theta + y$, and the last equality follows from the multinomial theorem. On the last line the sum is taken

[1] Figures were created with code by Samuli Siltanen.

over all multiindices $\alpha = (\alpha_1, \alpha_{,2}, \ldots, \alpha_N)$ with $|\alpha| = m$, and

$$\theta^\alpha := \theta_1^{\alpha_1}\, \theta_2^{\alpha_2}\, \cdots \theta_N^{\alpha_N} \text{ and } \binom{|\alpha|}{\alpha} := \frac{|\alpha|!}{\alpha_1!\, \alpha_2! \cdots \alpha_N!}.$$

Conversely, if $g \in \mathcal{S}(S^{N-1} \times \mathbb{R})$ is even and for every $m = 0, 1, 2, \ldots$ the integral

$$\int_{\mathbb{R}} s^m g(\theta, s)\, ds$$

is a homogeneous polynomial of degree m in θ, then $g = \mathcal{R}f$ for some $f \in \mathcal{S}(\mathbb{R}^N)$ [92, Section II.4].

Condition (5), which characterizes the range of the Radon transform, and the corresponding condition for the X-ray transform together are called the Helgason–Ludwig consistency conditions. The integral

$$m_\alpha(f) := \int_{\mathbb{R}^N} x^\alpha f(x)\, dx \tag{6}$$

on the last line of (5) is called the *geometric moment of order* α of the function f.

Interestingly, calculation (5) together with some elementary algebraic properties of polynomials already imply the injectivity of the Radon transform on, say, $L^1(B)$, where $B \subset \mathbb{R}^N$ is the unit ball. Namely, if we define

$$c_\alpha := \binom{|\alpha|}{\alpha} m_\alpha(f), \tag{7}$$

then (5) and the fact that a homogeneous polynomial is determined by its values on the unit sphere imply that the Radon transform $\mathcal{R}f$ determines the homogeneous polynomials

$$p_m(z) = \sum_{|\alpha|=m} c_\alpha z^\alpha,\ z \in \mathbb{R}^N,\ m = 0, 1, 2, \ldots$$

The polynomials p_m uniquely determine their coefficients c_α, and consequently the geometric moments $m_\alpha(f)$. The geometric moments in turn uniquely determine the integral of f against any polynomial, and therefore also f itself.

It is possible to derive explicit inversion formulas for the Radon transform, which is significant for the development of inversion algorithms. A way to calculate a function from its Radon transform is as follows [92]. Let $\mathcal{H}$ denote the Hilbert transform of a function $h \in \mathcal{S}(\mathbb{R})$ defined by

$$\mathcal{F}(\mathcal{H}h)(\sigma) := -i \operatorname{sgn}(\sigma) \mathcal{F}h(\sigma),$$

where $\mathcal{F}$ is the Fourier transform. For a function $g \in \mathcal{S}(S^{N-1} \times \mathbb{R})$ define

$$(\Lambda g)(\theta, s) := \begin{cases} \frac{d^{(N-1)}}{ds^{(N-1)}} g(\theta, s), & N \text{ odd}, \\ \mathcal{H} \frac{d^{(N-1)}}{ds^{(N-1)}} g(\theta, s), & N \text{ even}, \end{cases} \tag{8}$$

where the Hilbert transform $\mathcal{H}$ acts on the variable s. Then

$$f(x) = c_N \int_{S^{N-1}} h(\theta, \theta \cdot x)\, d\theta, \tag{9}$$

where the measure is the (unnormalized) spherical measure, $h := \Lambda \mathcal{R} f$, and

$$c_N := \begin{cases} \frac{1}{2}(2\pi)^{1-N}(-1)^{(N-1)/2}, & N \text{ odd}, \\ \frac{1}{2}(2\pi)^{1-N}(-1)^{(N-2)/2}, & N \text{ even}. \end{cases}$$

Note that, for a fixed θ, the integrand in (9) is constant in x on each affine hyperplane parallel to $\theta^\perp$, i.e., it is a plane wave. Therefore, the inverse Radon transform (9) can be interpreted to represent f as a superposition of plane waves.

Numerous other inversion formulas for the Radon transform have been derived, for example, to allow for computationally more efficient implementations, or to enlarge the domain of the inverse transform [49]. Some boundedness condition in the infinity for the function being transformed is nevertheless essential; for example, there exists smooth functions on the plane that are integrable on every line of the plane, but for which the integral over each line vanishes [135].

The X-ray transform also enjoys various inversion formulas [49, 92].

In the strict mathematical sense the Radon (or X-ray) transform of a function in finitely many directions never determines the function uniquely. Yet, in applications excellent reconstruction resolution is obtained if a large number of measurements distributed uniformly in all directions is available. Nevertheless, in some applications such data cannot be obtained. In *sparse angle tomography* only a small number of measurements are performed. In this case the data is a function of the form $\{\theta_1, \theta_2, \ldots, \theta_n\} \times \mathbb{R} \ni (\theta, s) \mapsto \mathcal{R} f(\theta, s)$, where $n \in \mathbb{N}$ is small. In *limited angle tomography* measurements are obtained only from some restricted angular range, the data is then a function of the form $U \times \mathbb{R} \ni (\theta, s) \mapsto \mathcal{R} f(\theta, s) \in \mathbb{R}$, where $U \subset S^{N-1}$ is a proper subset such as an arc in the two-dimensional case. Examples of reasons leading to sparse or limited angle tomography in applications are concerns over radiation dosage and occlusion by an external object, respectively. Both of these problems are an active field of research, see [116, 102, 63] and references therein.

In *tomography with unknown view angles* the directions in which the object is imaged are unknown. In this case the data is a family of functions

$$(\mathcal{R} f(\theta_j, \cdot))_{j \in J},$$

where the index set J is known, but the map $J \ni j \mapsto \theta_j \in S^{N-1}$ as well as the function f are unknown. In an application the view angles θ_j may be unknown or known only approximately, for example, due to uncontrollable motion of the object being imaged, such as involuntary movement of the patient in a CT scan [48], or the random movement of particles trapped by acoustic or optical forces [123, 34].

In three dimensions tomography with unknown view angles is particularly relevant for cryogenic electron microscopy (cryoEM) of viral particles [127, 111]. In

cryoEM of viral particles multiple images of identical virus specimen are imaged, but the relative orientations of the specimen are unknown. As the specimen is destroyed in the measurement by the beam of electrons that penetrates it, it is not possible to image the same specimen in many directions.

A method to recover the angles in tomography with unknown view angles was proposed in [41, 106]. The method is based on the geometric moments and the Helgason–Ludwig consistency conditions, however, no theoretical justification for the method was given (see also [11, Section VII]). In two dimensions the idea can be explained as follows. Let $f, \widetilde{f} \in L^1(B)$ be two functions, and let $c_\alpha \in \mathbb{R}$ and $\widetilde{c}_\alpha \in \mathbb{R}$ be defined by (7) for f and $\widetilde{f}$, respectively. If $\mathcal{R}f(\theta_j, s) = \mathcal{R}\widetilde{f}(\widetilde{\theta}_j, s)$, $j \in J$, then by the Helgason–Ludwig consistency condition (5) the system of homogeneous polynomials

$$y_1^2 + y_2^2 - z_1^2 - z_2^2 = 0, \tag{10a}$$

$$\sum_{|\alpha|=m} \left(c_\alpha y^\alpha - \widetilde{c}_\alpha z^\alpha \right) = 0,\ m = 0, 1, 2, \ldots, \tag{10b}$$

has at least the zeros $(y, z) = (\theta_j, \widetilde{\theta}_j) \in S^1 \times S^1$, $j \in J$. Bézout's theorem [29] in algebraic geometry provides an upper bound for the number of zeros of a system of polynomials with no common factor. Consequently, if the cardinality of J is large enough, then the polynomials (10) must have a common factor, and the idea is to show that this implies that the view angles $(\theta_j)_{j\in J}$ and $(\widetilde{\theta}_j)_{j\in J}$ must agree up to a common orthogonal transformation.

A theoretical justification for the aforementioned method was provided in [11]. There the authors prove that there exists a "generic" set of functions for which nine measurements from unknown view angles suffice for unique recovery of the view angles. In Publication I, among others, we improve this number to five measurements, and we provide a simple explicit condition for the function being imaged under which the results hold. Justification for the method in the three-dimensional case was considered in [77], and an algorithm implementing the method was proposed in [66].

In three dimensions an entirely different approach, the *common line method*, for the problem of (X-ray) tomography with unknown view angles is also available [128, 117]. This approach is based on the Fourier slice theorem [92, Theorem 1.1].

The Radon transform and the X-ray transform have many generalizations. These include integrating over different geometrical objects than lines or hyperplanes, such as the k-dimensional subspaces of $\mathbb{R}^N$, where $1 < k < N - 1$, or spheres, see [49, 36, 1] and references therein. Another class of generalizations is obtained by replacing the underlying space $\mathbb{R}^N$ and its hyperplanes by a manifold and a subset of its submanifolds. The field of mathematics studying these transformations is called integral geometry, see e.g. [49, 113].

The usefulness of the Radon transform is not limited to imaging applications. In fact, the Radon transform was found useful in the study of partial differential

equations well prior to any imaging applications [61]. This stems from the fact that the inverse Radon transform (9) decomposes a function into a sum (or integral) of plane waves (see [49, Ch. I, §7 A] for an overview of the main idea). In [114] and [101] the Radon transform and the X-ray transform, respectively, were used to solve an inverse problem for the wave equation. The Radon transform has been used also in the theory of neural networks, which is relevant to Publication IV. This will be explained further in Section 1.5.

1.3 Dynamical systems and diffusion models

The study of a (continuous time) dynamical system often starts with a differential equation of the form

$$\frac{d}{dt}u(t) = \Phi(t, u(t)), \tag{11}$$

where t is a real variable on some interval $I \subset \mathbb{R}$, and $\Phi : I \times X \to X$ is a given function on a vector space X. The task is to find all functions $u : I \to X$ that satisfy (11) and to understand their properties. Usually the variable t is interpreted as time, in which case the function Φ is seen to govern the time evolution of u. For a general exposition of dynamical systems we refer the reader to [120, 94, 132].

The field of dynamical systems theory breaks into numerous subfields, two of which are particularly relevant to this thesis. In section 1.4 we consider a system that models the dynamics of semiconductor lasers. That system is an example of a *nonlinear dynamical system*, that is, a system in which the function Φ is nonlinear in u. Nonlinear systems can manifest rich dynamical behavior, such as chaos, even in low dimensional spaces like $X = \mathbb{R}^3$.

In a *linear dynamical system* the function Φ is affine in u, so that equation (11) can be written as

$$\frac{d}{dt}u(t) = A(t)u(t) + f(t), \tag{12}$$

where $t \in I$, $A(t)$ is a linear operator on X for every t, and $f : I \to X$ is a function independent of u. Finite-dimensional linear dynamical systems, that is, those in which $\dim X < \infty$, are quite restricted in their behavior and they are covered by the classical theory, but the *infinite-dimensional* linear systems are more complicated. In infinite-dimensional linear systems the space X is often a function space, and the operator $A(t)$ on X may be unbounded and only densely defined. Frequently the operator $A(t)$ is a partial differential operator. Diffusion equation, the subject of this section, is an example of such a system.

Diffusion—the flow of some substance from areas of high concentration to areas of low concentration—is a commonly observed phenomenon in nature. Heat spreading in a metal rod and salt dissolving in water are examples of diffusion, as is the spreading of an animal population from an area of high population growth to the surroundings. Surprisingly, for various types of diffusion, a single mathematical equation is able to capture the essence of the phenomenon.

A general mathematical model for diffusion is the *diffusion equation* (also called the *heat equation*):

$$\partial_t u(x,t) - \Delta_g u(x,t) = f(x,t). \tag{13}$$

Here $u(x,t)$ is the density of the diffusing material at location x and time t, and $f(x,t)$ is a source term that represents the rate of injection or extraction of diffusing material at (x,t). Operator ∂_t is the partial derivative in t, and Δ_g is the Laplace–Beltrami operator on the underlying Riemannian manifold (M,g), where g is the Riemannian metric. Recall that for a smooth function v on M the Laplace–Beltrami operator is defined in local coordinates by

$$\Delta_g v := \sum_{j,k} |g|^{-1/2} \partial_j (|g|^{1/2} g^{jk} \partial_k v),$$

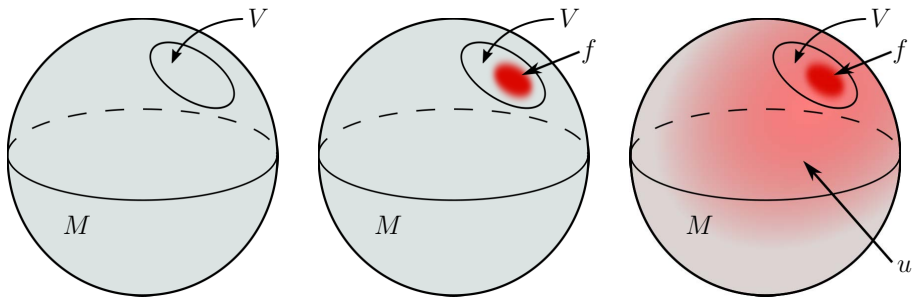
where $|g|$ is the determinant of g. A comprehensive account of the diffusion equation in the manifold setting is [44].

Equation (13) can be written in the form (12) as an infinite-dimensional linear dynamical system by considering u and f as functions of time t only, with values in, say, $X = L^2(M)$. Operator A is then the time-independent densely defined unbounded linear operator $H^2(M) =: D(A) \ni v \mapsto \Delta_g v \in L^2(M)$.

As a physical example, the diffusion equation (13) could model changes in the temperature distribution (represented by function u) caused by an external heat source (function f) being applied on some object (manifold M). The heat conductivity of the object would be represented by the metric g, and the tensor nature of g makes it possible to consider nonconstant and anisotropic (heat flows differently in different directions) conductivities, also. See figure 2.

In 1905 Einstein published his theory that explained Brownian motion, the random motion of microscopic particles suspended in a liquid, as a result of thermal molecular motions [33]. This theory, together with experimental verification of its predictions by Perrin in 1908 [95], confirmed the atomic nature of matter. A key result of Einstein's theory is that the mean-square displacement of the particle is proportional to time. At the continuum limit, it follows that the concentration of a large number of independent particles is governed by the diffusion equation.

Despite the great applicability of the diffusion equation, there are a number of experimental observations of diffusion processes that are not governed by it. *Anomalous diffusion* is a diffusion process where the mean square displacement does not scale linearly in time. In nature such processes have been observed in various settings such as in porous or fractured geological formations, turbulent fluids and plasma, and biological tissues, see [15, 136, 31, 22, 88, 109] and references therein. Continuous time random walk (a random walk in which waiting time and step length are sampled from given probability distribution), random walk in a space with fractal structure, and percolation are examples of mathematical processes giving rise to anomalous diffusion, see [14] and references therein.



(a) $t < 0$ (b) $t = 0$ (c) $t > 0$

Figure 2: Schematic illustration of diffusion of heat on an object, and a related inverse problem. In the beginning ($t < 0$) there are no heat sources on object M, and the object is uniform in temperature. At $t = 0$ a heat source f is applied in a region V of M. Immediately ($t > 0$) the heat starts to spread out from V, the temperature distribution u evolves according to the diffusion equation (13). The inverse problem is to deduce the global structure (shape and heat conductivity) of the whole object by observing the evolution of the temperature distribution only within the region V.

A so-called *(space–time) fractional diffusion equation*,

$$\partial_t^\alpha u(x,t) + (-\Delta_g)^\beta u(x,t) = f(x,t), \tag{14}$$

has been introduced to model anomalous diffusion. Above $0 < \alpha \leq 1$ and $0 < \beta \leq 1$. Operator ∂_t^α is a partial fractional derivative of order α in variable t, and it is defined in the sense of Caputo [21]. The Caputo fractional[2] derivative d^α/dt^α of a smooth scalar function y is defined by

$$\frac{d^\alpha}{dt^\alpha} y(t) := \begin{cases} \frac{1}{\Gamma(1-\alpha)} \int_0^t (t-\tau)^{-\alpha} y'(\tau)\, d\tau, & 0 < \alpha < 1, \\ y'(t), & \alpha = 1, \end{cases} \tag{15}$$

where Γ is Euler's gamma function and y' is the derivative of y. The space fractional part $(-\Delta_g)^\beta$ is defined in the sense of spectral theory [104]. In particular, if M is compact and $(\lambda_k)_{k=1}^\infty$ are the eigenvalues of $(-\Delta_g)$, listed according to their multiplicities, and if $(\varphi_k)_{k=1}^\infty \subset L^2(M)$ is a complete orthonormal sequence consisting of associated eigenfunctions, then

$$(-\Delta_g)^\beta v = \sum_{k=1}^\infty \lambda_k^\beta \langle v, \varphi_k \rangle_{L^2(M)} \varphi_k, \tag{16}$$

[2]The term *fractional* derivative is in use because of historical reasons, the order α can also be irrational.

with domain

$$\mathcal{D}((-\Delta_g)^\beta) := \left\{ v \in L^2(M) : \sum_{k=1}^{\infty} \lambda_k^{2\beta} |\langle v, \varphi_k \rangle_{L^2(M)}|^2 < \infty \right\} = H^{2\beta}(M).$$

It follows immediately from the definitions that equations (13) and (14) coincide if $\alpha = 1$ and $\beta = 1$.

There are in use various nonequivalent definitions of the time fractional derivative. Beside the Caputo fractional derivative, another common fractional derivative is D^α_{RL}, the *Riemann–Liouville fractional derivative* [107]:

$$D^\alpha_{\mathrm{RL}} y(t) := \begin{cases} \frac{1}{\Gamma(1-\alpha)} \frac{d}{dt} \int_0^t (t-\tau)^{-\alpha} y(\tau)\, d\tau, & 0 < \alpha < 1, \\ y'(t), & \alpha = 1. \end{cases}$$

In general the Caputo and Riemann–Liouville fractional derivatives are different operators, however, for a smooth function y with the zero initial value ($y(0) = 0$) they agree. Mathematical work on fractional calculus is extensive, for a general overview see textbooks [32, 75, 107, 74], reviews [17, 52] and references therein. Various generalizations of fractional diffusion equation also exist, such as space dependent time fractional orders where the order α depends on x [121].

Some aspects of linear systems theory generalize to the fractional differential case. For example, for $\lambda \in \mathbb{R}$ and a smooth function g the solution for the fractional initial value problem

$$\frac{d^\alpha}{dt^\alpha} y(t) + \lambda y(t) = g(t), \tag{17}$$

$$y(0) = 0, \tag{18}$$

for $t \geq 0$ is

$$y(t) = \int_0^t (t-\tau)^{\alpha-1} E_{\alpha,\alpha}(-\lambda(t-\tau)^\alpha) g(\tau)\, d\tau, \tag{19}$$

where $E_{\alpha,\alpha}$ is the two parameter Mittag–Leffler function [43], which is defined for general $a > 0$, $b > 0$, and $z \in \mathbb{C}$ by

$$E_{a,b}(z) := \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(ka + b)}. \tag{20}$$

In the fractional order case the Mittag–Leffler function often assumes the role of the exponential function in the standard differential equations. This is because the eigenfunctions of the fractional differential operator can be expressed in terms of $E_{a,b}(z)$. Namely, if $\lambda \in \mathbb{R}$ and $y(t) := E_{\alpha,1}(\lambda t^\alpha)$, $t \geq 0$, then

$$\frac{d^\alpha}{dt^\alpha} y(t) = \lambda y(t).$$

Note that if $\alpha = 1$, then $E_{\alpha,\alpha}$ *is* the exponential function, and equation (19) is the standard variation of parameters formula for a linear inhomogeneous differential equation.

On the other hand, many aspects from the case $\alpha = 1$ do not carry over to the fractional case. In particular, if $0 < \alpha < 1$, then the value of the fractional derivative $\frac{d^\alpha}{dt^\alpha} y(t)$ depends on the values $y(t')$ for all $t' \in (0, t)$, i.e., on the history of y. Extending the theory of fractional differential equations to Sobolev spaces also requires a substantial amount of work [75].

There are also different ways to generalize the Laplace–Beltrami operator to fractional orders. In $\mathbb{R}^N$, where the Fourier transform is available, the fractional Laplacian of order β, $(-\Delta)^\beta$, is often defined as the Fourier multiplier associated to symbol $|\xi|^{2\beta}$:

$$(-\Delta)^\beta v := \mathcal{F}^{-1}(|\xi|^{2\beta} \widehat{v}(\xi)) \tag{21}$$

(here $\widehat{v} = \mathcal{F}v$ is the Fourier transform of v, and $\mathcal{F}^{-1}$ is the inverse Fourier transform). Another common definition is the principal value integral [18]

$$(-\Delta)^\beta u(x) := c_{N,\beta}\, \mathrm{P.V.} \int_{\mathbb{R}^N} \frac{u(x) - u(y)}{|x - y|^{N+2\beta}}\, dy,$$

where $c_{N,\beta}$ is a certain constant. Other definitions involving semigroups of operators, Bochner integrals, or harmonic extensions are also in use, but unlike the different time fractional derivatives, all the above definitions agree, see the comprehensive account [76]. For fractional Laplacians in bounded domains see [103], and for fractional Laplacians arising as limits of random walks see [126]. A variable coefficient version of the fractional Laplacian in $\mathbb{R}^N$ has been considered in [38, 28].

In Publication II we consider an inverse problem for the fractional diffusion equation (14). There we consider sources f that vanish for x outside of some fixed subset $V \subset M$, and we observe $u^f(x, t)$ for $(x, t) \in V \times (0, \infty)$, where u^f is the solution for the fractional diffusion equation (14) with the zero initial value. The inverse problem is to recover the Riemannian manifold M from this data, or more generally, from only a single measurement. Considering sources supported on V and observing the corresponding solutions only on V is motivated by the fact that in an application this data could be measured with access to only a part (represented by V) of an object, as depicted in figure 2.

Inverse problems for the diffusion equation, especially in the setting of boundary measurements, have been studied by several authors. In the Euclidean setting Avdonin and Seidman [9] proved that uniqueness of the potential $q(x)$ in $\partial_t u = \Delta u - qu$ from boundary observations follows from the corresponding uniqueness result for the wave equation. They also constructed a source so that a single measurement suffices to determine q. This idea was extended by Cheng and Yamamoto in the planar case to recover a convection term from a single boundary measurement [25]. Our construction of the source for the determination of the manifold from a single measurement uses similar ideas as [9], but the situation is more complicated because classical results about time-analycity of solutions of parabolic differential equations are not available in the fractional case.

Uniqueness and reconstruction of a scalar conductivity $a(x)$ in $\partial_t u = a(x)\Delta u$ was shown in [8] using the boundary control method [12, 13], and joint uniqueness

of several coefficients was considered in [20]. In [65] Katchalov, Kurylev, and Lassas consider uniqueness and reconstruction problems in the manifold setting. The authors relate the inverse problem for the diffusion equation to an inverse boundary spectral problem and an inverse problem for the wave equation. In [64] it was shown that these three inverse problems are, in fact, equivalent.

Inverse problems for fractional PDEs have attracted major attention in recent years. The review [60] summarizes work on some common fractional inverse problems. Particularly relevant to our work is article [68] by Kian, Oksanen, Soccorsi, and Yamamoto, where they prove the uniqueness of the Riemannian metric for a time fractional PDE given Dirichlet-to-Neumann map at a fixed time at the boundary of the manifold. In [70] the same was done using a single measurement observed over a time interval. Inverse source problems were considered in [59, 71, 73, 84, 58], and an inverse problem with space-dependent order ($\alpha = \alpha(x)$) was considered in [72]. Simultaneous unique determination of several coefficients, sources, and obstacles was considered in [69]. In the static case, the fractional Calderón problem was studied by Ghosh, Salo, and Uhlmann in [40]. They show global uniqueness in the partial data problem where measurements are taken in an arbitrary open subset of the exterior. The fractional Calderón problem with a single measurement was considered in [39].

1.4 The laser as a dynamical system

The laser is a complex physical system whose description from the first principles involves at least electromagnetic theory, atomic physics, condensed matter physics, quantum theory, many-body theory, and plasma physics [26]. Constructing a mathematical model for a laser inevitably involves trade-offs between simplicity and accuracy of the model, with different objectives of the model giving rise to different trade-offs.

On the one hand, an analytical study of the laser as a dynamical system necessitates a relatively simple model, but on the other hand, lasers are known to exhibit complex behavior such as chaos [35, 67]. These things are not exclusive, however, as a simple model may exhibit complex behavior. The chaotic solutions of the well-known three-dimensional system of ordinary differential equations by Lorenz [86] is a standard example of such a system. Fortunately, there exist relatively simple models for lasers that capture their physical behavior with adequate accuracy for many purposes. In fact, the laser has been said to be a nearly ideal system for studying nonlinear phenomena, because they are easy to construct, and they exhibit complex behavior accurately modeled by models simple enough to be mathematically tractable [105]. Interestingly, a certain model of a laser is in fact mathematically equivalent to the Lorenz model [46].

Injecting external light from one laser into another laser's cavity (the part of a laser where the emitted light is created) alters the behavior of the laser that receives

the light. It is well-known that oscillatory systems tend to synchronize with external periodic perturbations, a fact whose first documented observation is by Christiaan Huygens from 1665 [56, 96]. As the laser is an optical oscillator, and as the injected external light creates a source of periodic perturbation, it is expected that under appropriate circumstances the two lasers may synchronize with each other. The first experimental demonstration of this phenomenon is from 1966 [119], and nowadays this technique of altering laser's phase or amplitude (or other properties) is called *optical injection locking* [131, 85]. As an example, this technique can be used to stabilize a strong but unstable laser by a weak but stable laser [35, 79]. Lasers are known to exhibit a rich dynamical behavior under external optical injection, see the comprehensive report [131] and the references therein.

Following [108, 89], we present a mathematical model of optical injection locking in the so-called Vertical-Cavity Surface-Emitting Lasers (VCSEL). In the model the electric field emitted by a laser is written as

$$E(x,t) = E(t)e^{-i(kx-\omega t)}, \tag{22}$$

where $E(t)$ is a $\mathbb{C}^2$-valued *slowly varying amplitude* that multiplies the scalar *carrier wave* $e^{-i(kx-\omega t)}$. The model assumes that $E(t)$ varies slowly in time compared to the carrier wave, which allows one to model the system in terms of $E(t)$ only. This is the so-called *slowly varying envelope approximation*. In (22) parameter k is the wave number, x is the spatial coordinate, ω is the angular frequency, and t is the time. The real-valued physical electric field emitted by a laser is

$$\mathcal{E}(x,t) = \mathrm{Re}\left(E(t)e^{-i(kx-\omega t)}\right).$$

One should note that slowness in the slowly varying amplitude $E(t)$ is relative to the carrier wave. Typically, the timescale for the slowly varying amplitude is in the nanosecond range, whereas for the carrier wave it is in the femtoseconds.

The $\mathbb{C}^2$-valued slowly varying amplitude $E(t)$ can be expressed in various physically meaningful bases. We choose to express $E(t)$ in the *circular polarization basis* (E_-, E_+), where the basis vectors correspond to electric fields that are constant in magnitude and rotating at a constant rate in a plane perpendicular to the propagation direction of the wave. The basis vectors $E_\pm$ are the *right* $(+)$ and *left* $(-)$ *circularly polarized* components, where the name denotes whether the rotation of the electric field is related to the direction of propagation by the right-hand or left-hand rule.

Another commonly used basis is the linear polarization basis (E_x, E_y) that consists of two linear polarizations that are perpendicular to each other. They are related to the circularly polarized components E_+ and E_- by

$$E_x = \frac{E_+ + E_-}{\sqrt{2}} \text{ and } E_y = -i\frac{E_+ - E_-}{\sqrt{2}}.$$

The temporal behavior of a VCSEL under external optical injection is expressed with so-called spin-flip rate equations [108, 89] as

$$\frac{d}{dt}E_{\pm}(t) = \kappa\big(1 + i\alpha\big)\,\big(N(t) \pm n(t) - 1\big)E_{\pm}(t) + \kappa\eta u_{\pm}(t), \tag{23a}$$

$$\begin{aligned}\frac{d}{dt}N(t) = &-\gamma\big(N(t) - \mu\big) - \gamma\big(N(t) + n(t)\big)|E_{+}(t)|^2 \\ &- \gamma\big(N(t) - n(t)\big)|E_{-}(t)|^2,\end{aligned} \tag{23b}$$

$$\begin{aligned}\frac{d}{dt}n(t) = &-\gamma_s n(t) - \gamma\big(N(t) + n(t)\big)|E_{+}(t)|^2 \\ &+ \gamma\big(N(t) - n(t)\big)|E_{-}(t)|^2.\end{aligned} \tag{23c}$$

Here $E_{\pm}(t)$ are the component functions of the slowly varying amplitude $E(t)$ in the circular polarization basis, $N(t)$ and $n(t)$ are real-valued functions related to internal dynamics of the laser, and $u_{\pm}$ are the circularly polarized components of the electric field of an external injection $u \in \mathbb{C}^2$, that is, the amplitude of the external light that is injected into the laser. The fields $E_{\pm}$ and $u_{\pm}$, as well as the functions N and n are written in a non-dimensionalized form so they are dimensionless. The other parameters are $\eta > 0$ (the coupling efficiency factor), $\alpha \geq 0$ (Henry factor), $\mu > 1$ (the normalized injection current), $\kappa > 0$ (decay rate of the cavity electric field), $\gamma > 0$ (decay rate of the total carrier number), and $\gamma_s > 0$ (the excess in the decay rate that accounts for the mixing in the carriers with opposite spins).

Interaction between the injected field u and the emitted field E in model (23) is unidirectional—the injected field u affects the behavior of the laser that emits the field E, but the field E does not affect u. An electric field E corresponding to a stable equilibrium point (a time-independent solution, also called a steady state) of system (23) is said to be *injection-locked*.

Figure 3 depicts a simulation of system (23) with an injected field u that is piecewise constant in time.[3] Parameter values were chosen to be realistic for a typical VCSEL. Simulation shows that after a change in the injected field u, injection locking (stabilization to a new stable equilibrium point of (23)) happens within a nanosecond.

System (23) is a six-dimensional (counting real dimensions) fully nonlinear system of ordinary differential equations. When considering linearly polarized injected fields u, one often restricts to linearly polarized solutions of system (23), in which case the system reduces to a three-dimensional system [35]. The equations can also be augmented with extra terms to model material birefringence or frequency detuning (injecting a field u with different frequency), or they can be modified to account for the Gaussian transverse profile of the laser beam.

In Publication III we propose a design for an optical logic gate that can be used as a building block for optical circuits. The operational principle of the logic gate is based on the physical process of injection locking. In the publication we derive

[3]The simulation was programmed in Julia [16] using the suite DifferentialEquations.jl [98].

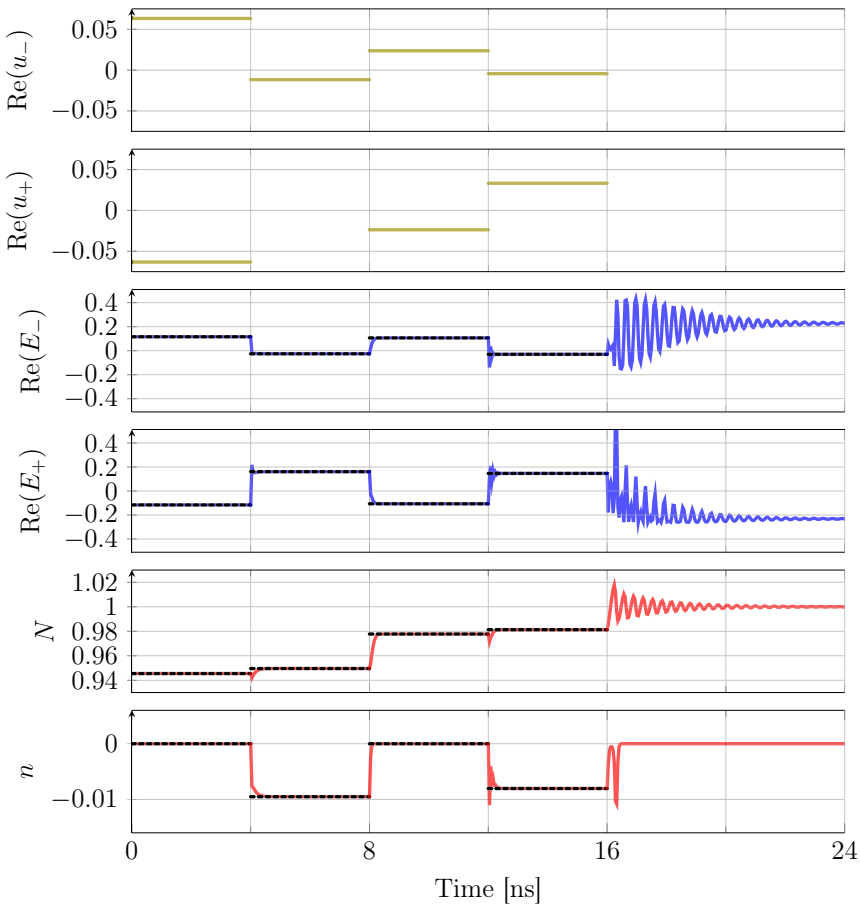


Figure 3: Time evolution of the components $E_\pm$ (blue lines) of the slowly varying amplitude of the electric field emitted by a laser with injected electric field u (olive lines) that is piecewise constant in time, and corresponding time evolution of the parameters N and n (red lines) of the laser. The electric fields are expressed in the circular polarization basis. The injected field u is chosen so that it is real-valued in this basis, N and n are real-valued, $E_\pm$ are complex-valued and only their real parts are shown. The values of the y-axes are unitless.
For the first 16 ns the injected field u changes every 4 ns. The black dotted lines denote the value of the stable equilibrium point of system (23) corresponding to the currently injected field u. After every change in u the emitted electric field E and the functions N and n quickly stabilize to the new equilibrium point. At $t = 16$ ns the injected field u is switched off. After that the emitted field E as well as N and n are no more locked into a specific value, and they start to fluctuate.
In this figure, $\eta = 1$, $\kappa = 1000\ \text{ns}^{-1}$, $\mu = 1.2$, $\alpha = 3$, $\gamma = 1\ \text{ns}^{-1}$, and $\delta = \gamma_s/\gamma = 50$. These values are realistic for a VCSEL.

an approximation for the relationship between the linearly polarized injected field $u \in \mathbb{C}^2$ and the corresponding injection-locked field $E \in \mathbb{C}^2$. The approximation is

$$E = c\,u/|u|^{-1}, \tag{24}$$

where $c = c(\mu, \alpha) \in \mathbb{C}$ is a constant. The idea of the design of the optical logic gate is to consider (24) as an input–output map that is computed by the laser.

A fundamental challenge in optical computing is that photons do not interact with each other in the vacuum but require a nonlinear medium. In this respect, the idea in Publication III is that the laser and the process of injection locking can provide the nonlinearity. In the publication it is experimentally demonstrated that an optical logic gate using the nonlinearity (24) is achievable, and the cascadibility of these gates is studied by simulation. For an overview of optical computing, see [124, 3, 24, 118], and references therein.

In Publication IV we analytically study system (23) with injected field u of arbitrary polarization. There we derive an exact form for the input–output map (24), and we also study the stability properties of the equilibrium points of (23). Based on this analysis we propose a design for an optical neural network.

1.5 Real- and complex-valued neural networks

Artificial neural networks (henceforth just neural networks) are the basis of deep learning, a machine learning technique that during the last decade has made major advances in solving problems in domains traditionally hard for computers, such as speech recognition, image recognition, and natural language processing [80]. Currently, there is a vast ongoing interest in deep learning due to how it has revolutionized machine learning in the aforementioned and other domains. This explosion of interest is relatively recent, it has only been made possible by advances in computer hardware and availability of vast datasets to train the networks, as well as advances in the theory [112]. However, the concept of a neural network has a long history. Modeled after a biological neuron, the (artificial) neuron—the elementary unit of neural networks—was introduced already in 1943 [90]. See [42, 93] for introduction to the theory of neural networks, and [112] for their history.

Most research and applications of neural networks have considered real-valued neural networks, meaning that the numbers that propagate through the network are real numbers. The definition of a neural network generalizes immediately to scalars from any field (or even more generally, a ring), however. In those applications where the inputs are naturally complex-valued, as for example any data represented in the frequency domain is, a complex-valued neural network seems natural. Indeed, in this kind of applications the complex-valued neural networks have been observed to sometimes perform better than the real-valued networks [129, 125, 133, 45]. See [53, 2] for more on complex-valued neural networks.

Despite the fact that real- and complex-valued neural networks formally look

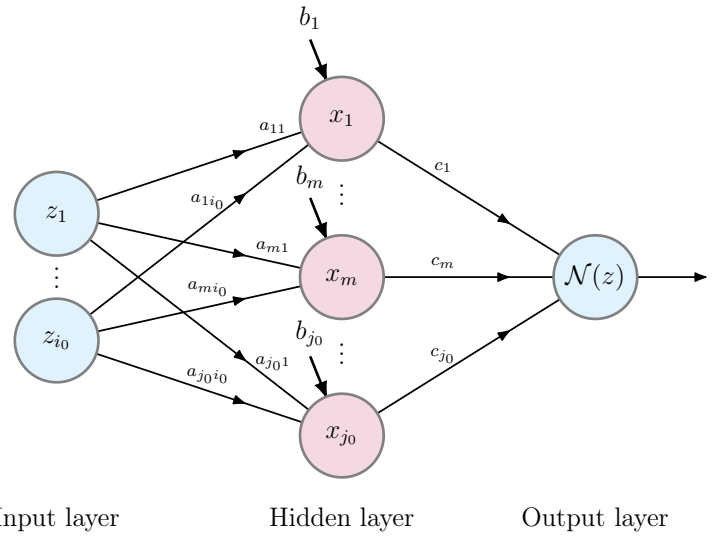


Figure 4: Illustration of a neural network. The network consists of three layers: an input layer, a hidden layer, and an output layer. The input layer contains inputs $z_i \in \mathbb{F}$. The m:th node in the hidden layer receives the inputs z_i multiplied by weights $a_{mi} \in \mathbb{F}$, and calculates the sum $x_m = \sum_i a_{mi} z_i + b_m$, where $b_m \in \mathbb{F}$ is a bias term. It then applies the activation function ρ to x_m and outputs $\rho(x_m)$. The single node in the output layer receives the values $\rho(x_j)$ multiplied by weights c_j, and outputs the sum of these as the final value $\mathcal{N}(z) = \sum_j c_j \rho(x_j)$ of the network.

similar, they have, in fact, quite different properties. Below we introduce both the real- and the complex-valued neural networks and indicate some their differences.

Let $\mathbb{F}$ be either $\mathbb{R}$ or $\mathbb{C}$. An $\mathbb{F}$-*valued neural network* is determined by following objects (also see figure 4):

(i) numbers $i_0 \in \mathbb{Z}_+$ and $j_0 \in \mathbb{Z}_+$, the number of inputs of the network and the *width* of the network, respectively,

(ii) a set $E \subset \mathbb{F}^{i_0}$, the domain of inputs of the network,

(iii) an *activation function* $\rho : U \to \mathbb{F}$, where $U \subset \mathbb{F}$, and

(iv) parameters $a_j = (a_{j1}, a_{j2}, \ldots, a_{ji_0}) \in \mathbb{F}^{i_0}$, $j = 1, 2, \ldots, j_0$, and $b, c \in \mathbb{F}^{j_0}$ that are required to satisfy

$$a_j \cdot z + b_j \in U \text{ for every } z \in E \text{ and } j = 1, 2, \ldots, j_0, \tag{25}$$

where $a_j \cdot z := \sum_i a_{ji} z_i$ (if $U = \mathbb{F}$, then (25) is redundant).

A neural network determined by (i)–(iv) computes the function $\mathcal{N} : E \to \mathbb{F}$

defined by

$$\mathcal{N}(z) := \sum_{j=1}^{j_0} c_j \rho(a_j \cdot z + b_j). \tag{26}$$

Often the function $\mathcal{N}$ itself is also called a neural network.

The neural network described above is more precisely a *shallow feedforward neural network with a single output*. A shallow feedforward neural network with *multiple outputs* computes a function $\mathcal{N} : E \to \mathbb{F}^{\ell_0}$ whose component functions are of the form (26). Several shallow neural networks (with multiple outputs) connected in sequence form a *deep neural network* (we omit the exact definitions, for more details we refer the reader to [42]).

For an activation function ρ, number of inputs i_0, and domain $E \subset \mathbb{F}^{i_0}$, the associated class $\mathcal{NN}(\rho, i_0, E)$ of neural networks is the set of functions

$$\mathcal{NN}(\rho, i_0, E) := \Big\{ E \ni z \mapsto \sum_{j=1}^{j_0} c_j \rho(a_j \cdot z + b_j) \in \mathbb{F} : \\ j_0 \in \mathbb{Z}_+, \text{ and } a_j \in \mathbb{F}^{i_0},\, b, c \in \mathbb{F}^{j_0} \Big\}.$$

A class $\mathcal{NN}(\rho, i_0, E)$ of neural networks is said to have the *universal approximation property*, if for every continuous function $f : E \to \mathbb{F}$, every compact set $K \subset E$, and every $\epsilon > 0$, there exists a neural network $\mathcal{N} \in \mathcal{NN}(\rho, i_0, E)$ such that

$$\sup_{z \in K} |f(z) - \mathcal{N}(z)| < \epsilon.$$

Note that the activation function ρ is not required to be continuous, so in general $\mathcal{NN}(\rho, i_0, E)$ is not a subset of the continuous functions on E.

Approximation capabilities of neural networks have been studied also with respect to other topologies than the above, such as $L^p(E)$ (with respect to the Lebesgue measure, for example) [54, 23], or the Sobolov space $W^{k,p}(E)$ [55, 91]. Also, there exist qualitative results about how the approximation error is related to the width of the network [10, 87, 47]. See [97, 110] for surveys on various approximation properties of neural networks

Regarding the universal approximation property, for *real*-valued neural networks it was proved in the 1990s that under very general assumptions a necessary and sufficient condition for the universal approximation property to hold is that the activation function is not a polynomial [83, 54]. This is the so-called *universal approximation theorem*. Perhaps surprisingly, the same is *not* true for complex-valued neural networks [4]. Several sufficient conditions for very specific types of complex-valued activation functions that lead to classes of neural networks with the universal approximation property were proved[4] also in the 1990s [4, 5], but the exact

[4]Unfortunately the literature contains several often cited results that are not entirely correct, see [130, Section 1.2].

characterization (for a globally defined activation function $\rho : \mathbb{C} \to \mathbb{C}$) was proved only recently [130]. In Publication IV we extend the result of [130] to a locally defined activation function $\rho : U \to \mathbb{C}$, where $U \subset \mathbb{C}$ is an open set. It turns out that for the complex-valued case instead of nonpolynomiality, the correct condition is that the activation function should not be polyharmonic. Recall that by definition the function ρ is polyharmonic if $\Delta^k \rho \equiv 0$ for some $k \in \mathbb{Z}_+$, where Δ is the Laplacian on $\mathbb{C} = \mathbb{R}^2$.

Many of the aforementioned results about the universal approximation property are nonconstructive in nature, and consequently they are of limited use in practice. In practice one would be given a function $f : E \to \mathbb{F}$, or more often just a set of samples $(x_i, f(x_i))_{i=1}^n$ of a function, and one is faced with the challenge of finding parameters for a network that approximates f satisfactorily (in a real world application one is usually free to choose also the activation function and the topology of the network). Interestingly, one of the first constructive methods for choosing the parameters of a neural network is based on an inversion formula for the Radon transform [23]. We next outline the idea behind this method.

Consider a smooth compactly supported function f on $\mathbb{R}^N$. Recall from section 1.2 that then

$$f(x) = c_N \int_{S^{N-1}} h(\theta, \theta \cdot x)\, d\theta, \tag{27}$$

where $h := \Lambda \mathcal{R} f$ (see equation (9)). Integral (27) can be approximated by a Riemann sum as

$$\int_{S^{N-1}} h(\theta, \theta \cdot x)\, d\theta \approx \sum_k \mu_k h(\theta_k, \theta_k \cdot x), \tag{28}$$

where $\mu_k \in \mathbb{R}$ is the measure of the element of the partition of S^{N-1} containing θ_k. Assuming that ρ is a so-called *sigmoidal* activation function, it is furthermore possible to approximate each of the scalar functions $h_k(s) := h(\theta_k, s)$ by a linear combination of scaled translates of ρ:

$$h_k(s) \approx \sum_j c_{k,j} \rho(a_{k,j} s - b_{k,j}), \text{ where } a_{k,j}, b_{k,j} \in \mathbb{R}. \tag{29}$$

In practice this approximation can be constructed by first approximating h_k by a linear combination of translates of the Heaviside step function, and then approximating each of the step functions by the sigmoidal activation function.

Combining equations (27) to (29) shows that

$$f(x) \approx c_N \sum_k \sum_j \mu_k c_{k,j} \rho(a_{k,j} \theta_k \cdot x - b_{k,j}),$$

which is of the form (26), i.e., a neural network. In [23] it was shown that for a square integrable compactly supported function f and a sigmoidal activation function ρ the above approximations can be justified in the L^2-norm, which yields an algorithm to construct a neural network that approximates a given function in the L^2-norm (for nonsmooth f the argument also involves a smoothing step).

The inverse Radon transform can also be applied to construct neural networks with other approximation properties, such as networks that approximate a function together with its derivatives [57].

In Publication IV we consider the dynamics of injection-locked semiconductor lasers, and show that the physical process of injection locking can be used to build an optical neural network. The electric fields in the laser are inherently complex-valued, which means that the resulting neural network is also complex-valued. We show that this type of optical neural network satisfies the assumptions of the universal approximation theorem with locally defined activation functions, meaning that, in principle, it is possible to approximate an arbitrary continuous function with such an optical neural network.

2 A review of the results of Publications I–IV

2.1 Publication I

In this paper we study uniqueness of tomography with unknown view angles in two dimensions. Our main instruments in this study are the Helgason–Ludwig consistency conditions (5) and the (geometric) moments (6). A moment of order (k, l) of a compactly supported integrable function ξ will be denoted by $m_{k,l}(\xi)$, i.e.,

$$m_{k,l}(\xi) := \int_{\mathbb{R}^2} x_1^k x_2^l \xi(\mathbf{x})\, d\mathbf{x}.$$

We denote by B the unit disk of $\mathbb{R}^2$ and call a function in $L^1(B)$ an *object*. If $U \in O(2)$, the group of real orthogonal 2×2-matrices, the *transformed object* ξ_U is defined by $\xi_U(\mathbf{x}) := \xi(U^T\mathbf{x})$, $\mathbf{x} \in \mathbb{R}^2$, where U^T is the transpose of U. For $\theta \in \mathbb{R}$ we define

$$R(\theta) := \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \in SO(2),$$

and

$$\xi_\theta := \xi_{R(\theta)} \text{ and } m_{k,l}(\xi;\theta) := m_{k,l}(\xi_\theta).$$

The *(parallel-beam line integral) projection* of object ξ at view angle $\theta \in \mathbb{R}$ is the function $P(\xi;\theta) \in L^1(\mathbb{R})$ defined almost everywhere by

$$P(\xi;\theta)(x_1) = \int_{\mathbb{R}} \xi_\theta(\mathbf{x})\, dx_2,\ x_1 \in \mathbb{R}.$$

The connection between moments and tomography comes from the fact that the moments $m_{k,0}(\xi;\theta)$ are determined by a projection from the view angle θ. Indeed,

$$m_{k,0}(\xi;\theta) = \int_{\mathbb{R}} x_1^k \left(\int_{\mathbb{R}} \xi_\theta(\mathbf{x})\, dx_2 \right) dx_1 = \int_{\mathbb{R}} x_1^k P(\xi;\theta)(x_1)\, dx_1. \tag{30}$$

A matrix $U \in O(2)$ is either a rotation or the composition of a rotation with a reflection. The reason why we above do not restrict to rotations only, i.e., to the

special orthogonal group $SO(2)$, is that an object ξ and the transformed object ξ_U are indistinguishable as far as tomography with unknown view angles is concerned.

Given projections $P(\xi;\theta_i)$, $i \in I$, with an unknown object ξ and unknown view angles θ_i, it is our aim to study if the projections determine the object and the view angles uniquely. We are most interested in the case where the index set I is finite, in this case we usually represent the indexed set $(\theta_i)_{i\in I}$ in the form of a vector $\boldsymbol{\theta} \in \mathbb{R}^n$.

There are two obstructions for the uniqueness of the object and the view angles. The first obstruction is that they can be uniquely determined only at most up to a common orthogonal transformation. This motivates the following definition:

Definition 1. Two vectors of view angles $\boldsymbol{\theta}, \widehat{\boldsymbol{\theta}} \in \mathbb{R}^n$ are equivalent, denoted $\boldsymbol{\theta} \sim \widehat{\boldsymbol{\theta}}$, if there exists $\sigma \in \{-1, 1\}$ and $\alpha \in \mathbb{R}$ such that for all $i = 1, 2, \ldots, n$ it holds that $\widehat{\theta}_i \equiv \sigma\theta_i + \alpha \mod 2\pi$.

The second obstruction is that uniqueness for the view angles cannot hold unless there is some asymmetry in the object ξ. Our uniqueness considerations are based on the object's moments, so what we assume is asymmetry in the moments:

Definition 2. Let $\xi \in L^1(B)$ and $d \geq 1$ be odd. The object ξ admits *spherical inertia*, if for every $U \in O(2)$ and $k + l = 2$ it holds that $m_{k,l}(\xi; U) = m_{k,l}(\xi)$. The object ξ admits $(2, d)$*-asymmetric moments*, if it does not admit spherical inertia and if the only matrix $U \in O(2)$ that satisfies $m_{k,l}(\xi; U) = m_{k,l}(\xi)$ for all $k+l = 2, d$ is the identity matrix.

Our main result for uniqueness for an object in tomography with unknown view angles is that a sufficiently asymmetric object is uniquely determined (up to an orthogonal transformation) by infinitely many projections with unknown view angles (Theorem 2.2 in Publication I):

Theorem 1. *Suppose $\xi \in L^1(B)$ admits $(2, d)$-asymmetric moments for some odd $d \geq 1$. Let I be an infinite index set and suppose $\theta_i \in \mathbb{R}$ satisfy $\theta_i \not\equiv \theta_j \mod 2\pi$ for $i \neq j$, $i, j \in I$. If $\eta \in L^1(B)$ and $\widehat{\theta}_i \in \mathbb{R}$ are such that $P(\xi;\theta_i) = P(\eta;\widehat{\theta}_i)$ for all $i \in I$, then there exists $S \in O(2)$ such that $\eta = \xi_S$.*

An object is never uniquely determined by finitely many projections even if the view angles are known. Nevertheless, the view angles themselves can be uniquely determined by finitely many projections, at least in the generic case. Here "generic" is defined as follows:

Definition 3. Let $D \subset \mathbb{N}$ be a nonempty finite set and denote by $\mathbf{m}_D(\xi)$ a vector that contains (in some fixed order) all d:th order moments of object ξ, for all $d \in D$. Let k be the dimension of $\mathbf{m}_D(\xi)$. A set S of objects in $L^1(B)$ is said to be D*-generic* if the set $\{\mathbf{m}_D(\xi) : \xi \in L^1(B) \setminus S\} \subset \mathbb{R}^k$ is a nowhere dense set of Lebesgue measure zero.

Above definition quantifies the size of a set of objects in $L^1(B)$ by measuring the size of the set of corresponding moments in $\mathbb{R}^k$. This is quite natural in this context, since the information about an object we use to determine the view angles are its d:th order moments for finitely many d.

A vector $\boldsymbol{\theta} \in \mathbb{R}^n$ of view angles is called *π-distinct* if it holds that

$$\theta_i \not\equiv \theta_j \mod \pi \text{ for all } i, j = 1, 2, \ldots, n,\ i \neq j.$$

Two projections from opposite view angles contain jointly the same information as each of them individually. In the following theorem we are interested in determining the unknown view angles from the least number of projections possible, therefore we assume π-distinct view angles (Theorem 2.6 in Publication I):

Theorem 2. *Let $d \geq 1$ be odd, $\boldsymbol{\theta} \in \mathbb{R}^n$ be π-distinct and suppose $n \geq 4d+1$. Then there exists a $\{2, d\}$-generic set $E = E(\boldsymbol{\theta}) \subset L^1(B)$ such that if $\xi \in E$ and there exist another object η and a vector $\widehat{\boldsymbol{\theta}} \in \mathbb{R}^n$ such that $m_{k,0}(\xi; \theta_i) = m_{k,0}(\eta; \widehat{\theta}_i)$ for all $i = 1, 2, \ldots, n$ and $k = 2, d$, then $\widehat{\boldsymbol{\theta}} \sim \boldsymbol{\theta}$.*

Note that by (30) the equality $m_{k,0}(\xi; \theta_i) = m_{k,0}(\eta; \widehat{\theta}_i)$ above is implied by $P(\xi; \theta_i) = P(\eta; \widehat{\theta}_i)$.

There are classes of objects for which the view angles cannot be uniquely determined by any number of projections; consider the class of objects with n-fold rotational symmetry, for example. For an object that admits $(2, d)$-asymmetric moments our final theorem provides a necessary and sufficient condition under which the view angles in tomography with unknown view angles are determined by moments of certain orders. The precise definition for what this means is as follows:

Definition 4. Let $\xi \in L^1(B)$, $D \subset \mathbb{N}$ and $n \in \mathbb{N}$, $n \geq 2$. We say that *view angles of object ξ are determined uniquely from n projections by D-order moments* if the following holds for all π-distinct $\boldsymbol{\theta} \in \mathbb{R}^n$: If η is an object and $\widehat{\boldsymbol{\theta}} \in \mathbb{R}^n$ a vector of view angles such that $m_{k,0}(\xi; \theta_i) = m_{k,0}(\eta; \widehat{\theta}_i)$ for all $i = 1, 2, \ldots, n$ and all $k \in D$, then $\widehat{\boldsymbol{\theta}} \sim \boldsymbol{\theta}$.

The result is as follows (Theorem 2.7 in Publication I):

Theorem 3. *Let c, d, $c \neq d$, be odd positive integers.*

1. *Suppose $\xi \in L^1(B)$ admits $(2, d)$-asymmetric moments. Let D be any finite set such that $\{2, c, d\} \subset D \subset \mathbb{N}$. Let $n \in \mathbb{N}$, $n \geq \max D + 3d + 1$. View angles of object ξ are determined uniquely from n projections by D-order moments if and only if the following implication is true: $\alpha, \beta \in \mathbb{R}$ and*

   $$m_{k,0}(\xi; \alpha) = m_{k,0}(\xi; \beta) \text{ for all } k \in D$$

   implies

   $$\alpha \equiv \beta \mod 2\pi.$$

2. *Suppose $d < c$ and let $n \geq c + 3d + 1$, $n \in \mathbb{N}$. The set of objects for which view angles are determined uniquely from n projections by $\{2, c, d\}$-order moments is $\{2, c, d\}$-generic.*

2.2 Publication II

In this paper we consider a direct problem and an inverse problem for the inhomogeneous space–time fractional diffusion equation (14) with the zero initial value:

$$\partial_t^\alpha u(x,t) + (-\Delta_g)^\beta u(x,t) = f(x,t), \qquad (x,t) \in M \times (0,\infty), \tag{31a}$$

$$u(x,0) = 0, \qquad x \in M. \tag{31b}$$

Here (M,g) is a compact Riemannian manifold (without boundary), and $0 < \alpha \leq 1$ and $0 < \beta \leq 1$. The Caputo fractional derivative ∂_t^α and the fractional power of the Laplace–Beltrami operator, $(-\Delta_g)^\beta$, are defined by (15) and (16), respectively.

The direct problem consists of proving the existence and uniqueness of a solution for (31). The inverse problem is motivated by the following question: Suppose the manifold M is unknown, except for an open set $V \subset M$ where one can do measurements. A measurement means applying a source f with spatial support in V, and observing the corresponding solution u^f on $V \times (0,\infty)$. Is it possible with such measurements to uniquely determine the global structure of the manifold M?

The main result of Publication II is that it is possible, and in fact it is possible even with a single measurement (Theorem 1 in Publication II):

Theorem 4. *Let (M,g) be a connected compact smooth Riemannian manifold without boundary, with metric g and $\dim M \geq 2$, let $V \subset M$ be a nonempty open subset with smooth boundary and let $T > 0$. Denote by u^f the solution of (31) corresponding to a source f.*

It is possible to construct a source $h \in C_c^\infty((0,T); L^2(V))$ such that the smooth manifold with boundary $\mathrm{cl}(V)$ together with the function $u^h|_{V\times[0,T)}$ determine the manifold (M,g) up to a Riemannian isometry.

The proof of Theorem 4 consists of three steps. First, in order to be able to state the inverse problem we need to prove that (31) admits a unique solution. This means deciding on a function space from which the source f is chosen, as well as finding an appropriate function space from which the solution is sought. This is the first step.

Next we formalize the measurements in the form of a measurement operator L_V, the so-called *local source-to-solution operator*, defined by

$$C_c^2((0,\infty); L^2(V)) \ni f \mapsto u^f|_{V\times[0,\infty)} =: \mathrm{L}_V f \in C^1([0,\infty); L^2(V)).$$

Here u^f is the unique solution for (31) corresponding to the source f, and the domain and range of L_V correspond to the choices of the function spaces made in the first step. The second step in the proof of Theorem 4 is to show that it is possible to construct a source $h \in C_c^\infty((0,T); L^2(V))$ so that the operator L_V is uniquely determined by the single measurement $(\mathrm{L}_V h)|_{[0,T)}$.

The third step in the proof of Theorem 4 is to show that the local source-to-solution operator L_V uniquely determines the manifold (M, g) up to a Riemannian isometry.

We briefly outline the execution of each of the steps.

The function u in (31) is a function of time t and space x, but it is convenient to consider it as a function of t only, with values in a function space in x, usually $L^2(M)$. The Caputo derivative of order α of an $L^2(M)$-valued function $y \in C^1([0,\infty); L^2(M))$, denoted by $\partial_t^\alpha y(t)$, is defined analogously to the scalar case:

$$\partial_t^\alpha y(t) := \begin{cases} \frac{1}{\Gamma(1-\alpha)} \int_0^t (t-\tau)^{-\alpha} y'(\tau)\, d\tau, & 0 < \alpha < 1, \\ y'(t), & \alpha = 1. \end{cases}$$

Here the derivative y' is in the sense of the derivative of an $L^2(M)$-valued function of a real variable, and the integral is in the sense of Bochner.

Consider a source $f \in C_c^2((0,\infty); L^2(M))$. The function $u \in C^1([0,\infty); L^2(M))$ is said to be a *strong solution* for the initial value problem (31), if

(i) $u(0) = 0$,

(ii) $u(t) \in \mathcal{D}((-\Delta_g)^\beta)$ for every $t \geq 0$, and

(iii) $\partial_t^\alpha u(t) + (-\Delta_g)^\beta u(t) = f(t)$ for every $t \geq 0$.

The manifold M is compact by assumption, so the spectrum of $(-\Delta_g)$ is discrete and we can order its eigenvalues into an increasing sequence $(\lambda_k)_{k=1}^\infty \subset [0,\infty)$ where the eigenvalues are listed according to their multiplicities. Let $(\varphi_k)_{k=1}^\infty \subset L^2(M)$ be an orthonormal basis of $L^2(M)$ consisting of corresponding eigenfunctions.

Because the functions φ_k are also eigenfunctions of $(-\Delta_g)^\beta$, we can prove the uniqueness and existence of a strong solution for (31) using an eigenfunction expansion. This also yields a representation formula for the solution (Proposition 6 in Publication II):

Proposition 5. *For every source $f \in C_c^2((0,\infty); L^2(M))$ there exists a unique strong solution $u^f \in C^1([0,\infty); L^2(M))$ for the fractional diffusion equation* (31)*. This solution can be represented as*

$$u^f(t) = \sum_{k=1}^\infty u_k^f(t)\, \varphi_k,\ t \geq 0, \tag{32}$$

where the series converges in $L^2(M)$ for every $t \geq 0$, and

$$u_k^f(t) := \int_0^t (t-\tau)^{\alpha-1} E_{\alpha,\alpha}(-\lambda_k^\beta (t-\tau)^\alpha)\, \langle f(\tau), \varphi_k \rangle_{L^2(M)}\, d\tau,\ t \geq 0, \tag{33}$$

where $E_{\alpha,\alpha}$ is the Mittag–Leffler function (20)*.*

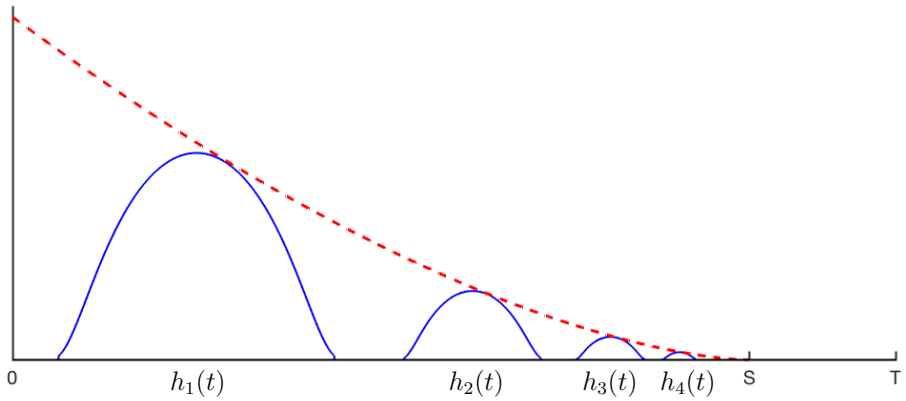


Figure 5: Examples of functions h_k, $k \geq 1$, are plotted in blue with the choice of n being a Gaussian bump. The exponential decay of the amplitude of h_k as k increases is illustrated by the red dashed line. The values of h_k are not presented in scale.

Next we explain the construction of a source $h \in C_c^\infty((0,T);L^2(V))$ such that the operator L_V is uniquely determined by the measurement $(\mathrm{L}_V h)|_{[0,T)}$. This construction uses similar ideas as a construction for a parabolic equation in [9], but now the situation is complicated by the fact that standard results about time-analycity of solutions for the parabolic equation are not available for the fractional case.

Fix numbers $0 < S < T$, a nonzero non-negative function $n \in C_c^\infty(-1,1)$, and a bounded sequence $(\psi_k)_{k=1}^\infty \subset L^2(V)$ of functions that spans a dense subspace of $L^2(V)$. Let $(r(k))_{k=1}^\infty$ be the sequence of positive integers that begins

$$1, \quad 1,2, \quad 1,2,3, \quad 1,2,3,4, \quad \dots$$

(grouping of the terms is chosen to emphasize the pattern). Define the source h by

$$h(t) := \sum_{k=1}^{\infty} h_k(t)\psi_{r(k)},\ t \in \mathbb{R}, \tag{34}$$

where

$$h_k(t) := \frac{S^k}{2^{k(k+2)}m_k} \cdot n\left(2^{k+1}\left(\frac{t}{S}-1\right)+3\right)$$

with $m_k := \max_{0\leq l\leq k} \|n^{(l)}\|_\infty$. See figure 5. Then $h \in C_c^\infty((0,T);L^2(V))$ (Proposition 15 in Publication II).

The rationale behind the construction of h is following. Properties related to the time-analycity of solutions for (31) imply that $(\mathrm{L}_V h)|_{[0,T)}$ uniquely determines the measurements corresponding to the individual terms of h, i.e., the functions

$$\mathrm{L}_V(h_k\psi_{r(k)}),\ k = 1,2,\dots \tag{35}$$

(Lemma 18 in Publication II). From this and the representation formulas (32) and (33) it follows that $(\mathrm{L}_V h)|_{[0,T)}$ uniquely determines also the measurements

$$\mathrm{L}_V(h_k(\cdot - t_0)\psi_{r(k)}), \tag{36}$$

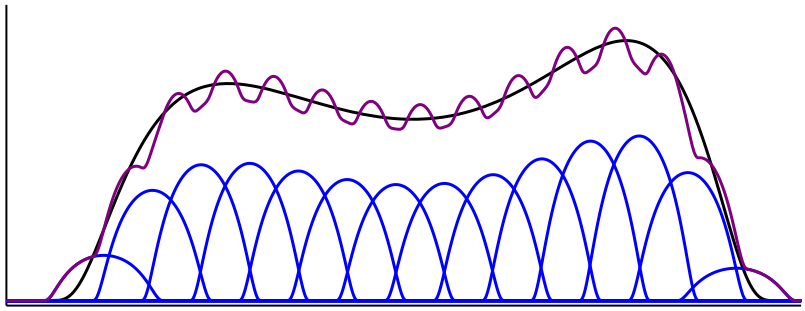


Figure 6: A function (in black) and its approximation (in purple) by a linear combination of translates of a bumb function. The approximation is the sum of the blue bumb functions.

where $t_0 \in \mathbb{R}$ is such that $h_k(\cdot - t_0)$ is supported on $(0, \infty)$ (Lemma 16 in Publication II).

Consider a function $a \in C_c^2(0, \infty)$. The sequence $(h_{r(k)})_{k=1}^{\infty}$ consists of smooth positive bumb functions with shrinking supports, which implies that it is possible to uniformly approximate the function a by a linear combination of functions from the set

$$\{h_k(\cdot - t_0) : t_0 \in \mathbb{R}, \operatorname{supp} h_k(\cdot - t_0) \subset (0, \infty),\ k = 1, 2, \ldots\}$$

(Lemma 19 in Publication II). See figure 6. This approximation property holds also for any subsequence $(h_{r(k_j)})_{j=1}^{\infty}$ of $(h_{r(k)})_{k=1}^{\infty}$.

Let us now group the terms of h as

$$\begin{aligned} h(x,t) \ = \ & h_1(t)\psi_1(x) + \\ & h_2(t)\psi_1(x) + h_3(t)\psi_2(x) + \\ & h_4(t)\psi_1(x) + h_5(t)\psi_2(x) + h_6(t)\psi_3(x) + \\ & h_7(t)\psi_1(x) + h_8(t)\psi_2(x) + h_9(t)\psi_3(x) + h_{10}(t)\psi_4 + \ldots \end{aligned} \tag{37}$$

Above reasoning implies that the source $a(t)\psi_1(x)$ can be approximated by a linear combination of sources of the form

$$\{h_k(\cdot - t_0)\psi_1 : t_0 \in \mathbb{R}, \operatorname{supp} h_k(\cdot - t_0) \subset (0, \infty),\ k = 1, 2, 4, 7, \ldots\},$$

i.e., by time translates of the sources in the first column in (37). Using this approximation, continuity properties of L_V (Proposition 13 in Publication II), and the fact that $(\mathrm{L}_V h)|_{[0,t)}$ uniquely determines the functions (36), it can be shown that $(\mathrm{L}_V h)|_{[0,t)}$ also uniquely determines $\mathrm{L}_V(a(t)\psi_1(x))$. The same is true if ψ_1 is replaced by ψ_k for any k, which together with the fact that the sequence $(\psi_k)_{k=1}^{\infty}$ spans a dense subspace of $L^2(V)$ can be used to complete the proof of the following (Proposition 2 in Publication II):

Proposition 6. *Let $h \in C_c^{\infty}((0,T); L^2(V))$ be the source constructed above. The function $(\mathrm{L}_V h)|_{[0,T)}$ uniquely determines the local source-to-solution operator L_V.*

The third and final step is to prove that the operator L_V uniquely determines the manifold (M, g). This proof is based on a reduction to the wave equation on the same manifold (M, g). Similar reduction was considered in [9] for a parabolic equation in the Euclidean space, and in the manifold setting in [65, 64].

Let $(\lambda_{q_k})_{k=1}^{\infty}$ be the increasing sequence of distinct eigenvalues of $(-\Delta_g)$, and let $P_k : L^2(M) \to L^2(M)$ be the orthogonal projection onto the eigenspace corresponding to the eigenvalue λ_{q_k}. Define the operator $P_{V,k} : L^2(V) \to L^2(V)$ by $P_{V,k}v := (P_k v)|_V$, $v \in L^2(V) \subset L^2(M)$. Then using the representation formulas (32) and (33) it is possible to show that

$$\mathcal{L}\,\mathrm{L}_V\, f(s) = \sum_{k=1}^{\infty} \frac{P_{V,k}\mathcal{L}f(s)}{s^\alpha + \lambda_{q_k}^\beta},\ \mathrm{Re}(s) > 0, \tag{38}$$

where $\mathcal{L}$ is the Laplace transform (Proposition 21 in Publication II). From the poles and residues of the analytic continuation of (38) it can be deduced that L_V uniquely determines the data

$$(\lambda_{q_k}, P_{V,k})_{k=1}^{\infty} \tag{39}$$

(Proposition 22 in Publication II).

The data (39) uniquely determines the local source-to-solution operator $\mathrm{L}_V^{\mathrm{hyp}}$ for the wave equation on (M, g) (see proof of Theorem 3 in Publication II). This operator is defined for a source $p \in C_c^\infty(V \times (0, \infty))$ by

$$\mathrm{L}_V^{\mathrm{hyp}}\, p := w^p|_{V\times(0,\infty)},$$

where w^p is the unique solution for the wave equation

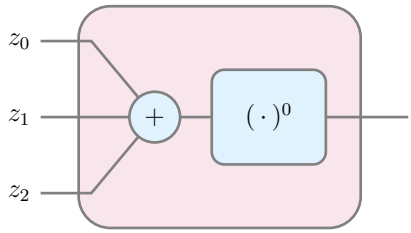
$$\begin{aligned} (\partial_t^2 - \Delta_g)w(x,t) &= p(x,t), && (x,t) \in M \times (0,\infty), \\ w(x,0) &= 0, && x \in M, \\ \partial_t w(x,0) &= 0, && x \in M. \end{aligned} \tag{40}$$

This is enough to prove Theorem 4 due to a result by Helin, Lassas, Oksanen, and Saksala in [50], where they show that $\mathrm{L}_V^{\mathrm{hyp}}$ uniquely determines the manifold (M, g) up to a Riemannian isometry.

2.3 Publication III

This paper consists of an experimental part and a theoretical part. The author of this thesis had no part in the experimental work.

In the theoretical part of the paper we derive an approximation, starting from system (23), for the injection-locked field E of a laser that is injected with a weak electric field u of linear polarization. The principal idea in the paper is to consider the laser as a computing device that computes, by the physical process of injection locking, the operation $u \mapsto E$. We show that by using this operation, it is possible to implement an optical logic gate that can be used as a building block in optical



(a)

z_0	z_1	z_2	$(z_0+z_1+z_2)^0$
+	+	+	+
+	+	−	+
+	−	+	+
+	−	−	−
−	+	+	+
−	+	−	−
−	−	+	−
−	−	−	−

(b)

Figure 7: (a) Schematic illustration of an optical logic gate that implements the majority gate. Purple rectangle denotes the majority gate. The inputs are $z_j \in \{-1,+1\}$, $j = 0,1,2$. In the majority gate the inputs are first combined into a single field with value $z_0+z_1+z_2$, which is then injected into a laser. The output of the gate is the injection-locked field $(z_0+z_1+z_2)^0 \in \{-1,+1\}$ of the laser. (b) Table of values of the map $(z_0,z_1,z_2) \mapsto (z_0+z_1+z_2)^0$. Only the sign of the value is shown. With the interpretation $+$ = TRUE and $-$ = FALSE this is the truth table of the majority gate.

computing devices. Furthermore, we study the cascadibility of this optical logic gate by simulating two different optical circuits that are composed of several of these gates.

The experimental part of the paper is the demonstration of the optical logic gate based on vertical-cavity surface-emitting lasers (VCSEL).

To understand the computation $u \mapsto E$ that the laser performs, let us suppose that the injected field u is of the form $u = z\,u_0$, where $z \in \mathbb{C}$, and $u_0 \in \mathbb{C}^2$, $|u_0| = 1$, is a fixed linear polarization. In Supplement S1 of Publication III we show that the injection-locked field E of a laser that is injected with this field $u = z\,u_0$ can be approximated as

$$E = c\,\frac{z}{|z|}\,u_0, \text{ where } c := e^{i\theta}\sqrt{\mu-1} \text{ and } \theta := -\arg(1+i\alpha). \tag{41}$$

Because c and u_0 are constants that are independent of z, we can think that the laser computes the operation

$$\mathbb{C}\setminus\{0\} \ni z \mapsto \frac{z}{|z|} =: (z)^0 \in \mathbb{C}. \tag{42}$$

The operator $(\cdot)^0$ in (42) is called the *normalization operator* (also called the *signum function*).

The logic gate that we implement is the *majority gate*, which is a logic gate that takes in an odd number of Boolean inputs (with values denoted by TRUE/FALSE), and outputs TRUE if the majority of its inputs are TRUE, and FALSE otherwise. We

explain the principle on which the implementation of this optical logic gate is based using three inputs, the generalization for any odd number of inputs is immediate. See figure 7.

The three inputs of the optical logic gate are electric fields. They could be, for example, the electric fields emitted by three different semiconductor lasers in an integrated optical circuit. The fields are assumed to share a common linear polarization, have an equal intensity, and each of the fields is assumed to have a phase difference of either 0 or π with respect to some common reference signal. From these assumptions it follows that the inputs can be written as $z_j\, u_0$, where $z_j = \pm 1$, $j = 0, 1, 2$, and $u_0 \in \mathbb{C}^2$.

In the optical logic gate the inputs $z_j\, u_0$ are first combined into a single field u,

$$u = (z_0 + z_1 + z_2)\, u_0,$$

and then the combined field u is injected into a laser. According to approximation (41), the injection-locked field E of the laser is then

$$E = c'\, (z_0 + z_1 + z_2)^0\, u_0, \tag{43}$$

where $c' \in \mathbb{C}$ is a constant. The output of the logic gate is this injection-locked field E.

Again, by ignoring the constants c' and u_0 in (43), we can think that the logic gate computes the map

$$\{-1, +1\}^3 \ni (z_0, z_1, z_2) \mapsto (z_0 + z_1 + z_2)^0 \in \{-1, +1\}. \tag{44}$$

An input–output table for this map is shown in figure 7. If we there interpret $+1$ as TRUE and -1 as FALSE, then the table is exactly the truth table of the majority gate.

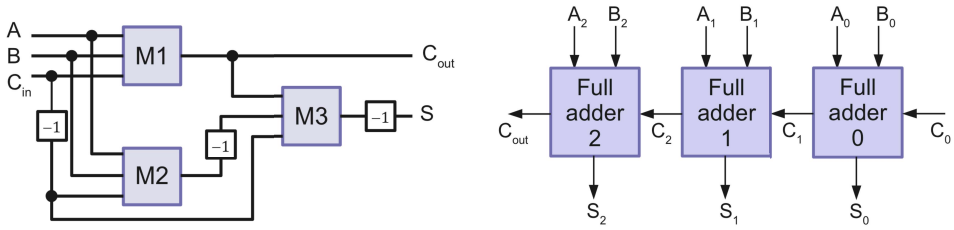
Another way to interpret the 3-input majority gate is to think it as a programmable logic gate where, say, z_0 is the control, and z_1 and z_2 are the inputs. If $z_0 = +1$, then the output is $z_1 \vee z_2$ (logical disjunction), and if $z_0 = -1$, then the output is $z_1 \wedge z_2$ (logical conjunction). That is, depending on the value of z_0 the gate computes either the OR-gate or the AND-gate.

In Supplement S2 of Publication III we simulate two optical circuits that are build with the majority gate as the basic building block. These circuits are:

1. A full adder (see figure 8a). This circuit has three binary inputs (A, B, and C_{in}) and two binary outputs (S and C_{out}). The inputs and outputs are interpreted to be 0/1-valued, and they are related by

   $$A + B + C_{\text{in}} = C_{\text{out}} S, \tag{45}$$

   where on the right-hand side the binary representation $C_{\text{out}} S$ is identified with the number $2C_{\text{out}} + S \in \{0, 1, 2, 3\}$. This circuit is built with three majority gates.



(a) (b)

Figure 8: Schematic diagrams of a full adder and a ripple-carry adder implemented with the majority gate. (a) A full adder composed of three majority gates (M1, M2, and M3). A box labeled with '-1' denotes a phase-shift of π. (b) A 3-bit ripple-carry adder composed of three full adders. In simulations C_0 is set to 0.

2. A 3-bit ripple-carry adder (see figure 8b). This circuit has six inputs, A_i and B_i, $i = 0, 1, 2$, and four outputs, S_i, $i = 0, 1, 2$ and C_{out}. The input and output are related by

$$A_2A_1A_0 + B_2B_1B_0 = C_{\text{out}}S_2S_1S_0,$$

where the expression is interpreted analogously to (45). This circuit is built with nine majority gates.

The simulations show that the optical circuits correctly implement the adders with response times measured in nanoseconds, see figure S2 and figure S3 in Supplement of Publication III.

2.4 Publication IV

In the first part of this paper we do a detailed analysis of the dynamics of injection locking in a semiconductor laser by analyzing system (23). In Publication III we considered injection locking with a weak injected field u that is linearly polarized, and we derived approximation (41) for the injection-locked field E that is valid in the limit $|u| \to 0$. Here we consider an injected field with arbitrary polarization, and we derive asymptotic expressions (up to $o(u)$-terms) for the injection-locked field in the limits $|u| \to 0$ and $|u| \to \infty$.

In the second part of the paper we consider optical neural networks. We propose a design for an optical neural network in which the nonlinear activation function is computed by the physical process of injection locking. Based on the dynamical properties of system (23) that we prove in the first part of the paper, and a local version of the universal approximation theorem for complex-valued neural networks that we also prove, we show that such an optical neural network can, in principle, uniformly approximate an arbitrary continuous function.

Recall that determining the value of the injection-locked field E means finding the equilibrium point of (23), i.e., the zero of its right-hand side. In fact, in the

general situation the right-hand side can have several zeros, and we want to find them all.

In Proposition 3 of Publication IV we reduce the problem of finding the zeros of the right-hand side of (23) to a problem of finding the zeros of a system of polynomials. This bears similarity to the problem considered in Publication I. There we had a system of unknown polynomials (obtained from the Helgason–Ludvig consistency conditions), and the knowledge that they share a certain number of common zeros. From that information we wanted to deduce information about the coefficients of the polynomials (the geometric moments of the objects being imaged). Here we have a system of known polynomials, and we want to deduce information about their zeros in terms of some coefficients (related to the injected field u) of the polynomials.

For finding the equilibrium points of (23), it is convenient to introduce the matrix-valued functions $X, Y : \mathbb{C}^2 \to \mathbb{C}^{2\times 2}$ defined by

$$X(z) := \begin{bmatrix} 1-(z_1-z_2) & 0 \\ 0 & 1-(z_1+z_2) \end{bmatrix}, \tag{46a}$$

$$Y(z) := \begin{bmatrix} 1+|z|^2 & |z_2|^2-|z_1|^2 \\ |z_2|^2-|z_1|^2 & \delta+|z|^2 \end{bmatrix}, \tag{46b}$$

where $\delta := \gamma_s/\gamma > 0$ and $|z|$ is the Euclidean norm of $z \in \mathbb{C}^2$. Then the four equations (23) can be written as a pair

$$\frac{d}{dt}E(t) = -\kappa\big((1+i\alpha)\,X(N(t),n(t))E(t) - u\big), \tag{47a}$$

$$\frac{d}{dt}\begin{bmatrix} N(t) \\ n(t) \end{bmatrix} = -\gamma\left(Y(E(t))\begin{bmatrix} N(t) \\ n(t) \end{bmatrix} - \begin{bmatrix} \mu \\ 0 \end{bmatrix}\right), \tag{47b}$$

where $E = (E_-, E_+)$. In (47) we have assumed without loss of generality $\eta = 1$ for the coupling efficiency factor, because this constant can be incorporated in the field u. The pair (47) is more amenable to algebraic manipulations than the four equations (23), particularly because the matrix Y is always invertible. When solving for the zeros of the right-hand sides of (47), we can first solve (N, n) in (47b) in terms of E,

$$\begin{bmatrix} N \\ n \end{bmatrix} = Y(E)^{-1}\begin{bmatrix} \mu \\ 0 \end{bmatrix} =: y(E),$$

and then substitute $y(E)$ for (N, n) in (47a). This way we get an equation for the equilibrium points in terms of only $E \in \mathbb{C}^2$.

The reason for originally writing system (23) in terms of the circularly polarized basis states $E_\pm$ is that in this basis the matrix $X(N, n)$ is diagonal.

Our main result about the asymptotics of the equilibrium points corresponding to weak injected fields is the following (Theorem 2 in Publication IV):

Theorem 7. *Consider injected external field with amplitude $\lambda\widehat{u}$, where $\lambda \in \mathbb{C}$ and $\widehat{u} = (\widehat{u}_-, \widehat{u}_+) \in \mathbb{C}^2$ satisfies $\widehat{u}_- \neq 0$ and $\widehat{u}_+ \neq 0$. There exists a constant $\ell = \ell(\widehat{u}) > 0$ and a family $\{E^{(j)}_{\widehat{u}}\}_{j\in\mathcal{J}}$ of nine continuous functions*

$$E^{(j)}_{\widehat{u}} : \{\lambda \in \mathbb{C} : 0 < |\lambda| < \ell\} \to \mathbb{C}^2,\ j \in \mathcal{J} := \{0, \pm\mathrm{L}, \pm\mathrm{R}, \pm\mathrm{X}, \pm\mathrm{Y}\}, \tag{48}$$

with pairwise distinct values that have the following properties:

(i) If in system (47) the injected field is of the form $u = \lambda\widehat{u}$ with $0 < |\lambda| < \ell$, then a triple $(E, N, n) \in \mathbb{C}^2 \times \mathbb{R} \times \mathbb{R}$ is an equilibrium point (a time-independent solution) of the system, if and only if

$$E = E^{(j)}_{\widehat{u}}(\lambda) \text{ for some } j \in \mathcal{J}, \text{ and } (N, n) = y(|E_-|, |E_+|).$$

(ii) The functions $E^{(j)}_{\widehat{u}}$ have following asymptotics as $\lambda \to 0$:

$$E^{(0)}_{\widehat{u}}(\lambda) = e^{i\theta}\frac{\lambda}{|\lambda|}\left(|\lambda|\widehat{w}^{(0)} + o(\lambda)\right), \tag{49a}$$

$$E^{(\pm\mathrm{L})}_{\widehat{u}}(\lambda) = e^{i\theta}\frac{\lambda}{|\lambda|}\left(\pm\sqrt{\frac{\delta(\mu-1)}{1+\delta}}\begin{bmatrix}\widehat{u}_-/|\widehat{u}_-|\\ 0\end{bmatrix} + |\lambda|\ \widehat{w}^{(\mathrm{L})} + o(\lambda)\right), \tag{49b}$$

$$E^{(\pm\mathrm{R})}_{\widehat{u}}(\lambda) = e^{i\theta}\frac{\lambda}{|\lambda|}\left(\pm\sqrt{\frac{\delta(\mu-1)}{1+\delta}}\begin{bmatrix}0\\ \widehat{u}_+/|\widehat{u}_+|\end{bmatrix} + |\lambda|\,\widehat{w}^{(\mathrm{R})} + o(\lambda)\right), \tag{49c}$$

$$E^{(\pm\mathrm{X})}_{\widehat{u}}(\lambda) = e^{i\theta}\frac{\lambda}{|\lambda|}\left(\pm\sqrt{\frac{\mu-1}{2}}\begin{bmatrix}\widehat{u}_-/|\widehat{u}_-|\\ \widehat{u}_+/|\widehat{u}_+|\end{bmatrix} + |\lambda|\,\widehat{w}^{(\mathrm{X})} + o(\lambda)\right), \tag{49d}$$

$$E^{(\pm\mathrm{Y})}_{\widehat{u}}(\lambda) = e^{i\theta}\frac{\lambda}{|\lambda|}\left(\pm\sqrt{\frac{\mu-1}{2}}\begin{bmatrix}\widehat{u}_-/|\widehat{u}_-|\\ -\widehat{u}_+/|\widehat{u}_+|\end{bmatrix} + |\lambda|\,\widehat{w}^{(\mathrm{Y})} + o(\lambda)\right), \tag{49e}$$

where $\theta := -\arg(1 + i\alpha)$ and

$$\widehat{w}^{(0)} := \frac{-1}{|1+i\alpha|(\mu-1)}\widehat{u}$$

$$\widehat{w}^{(\mathrm{L})} := \frac{1}{2|1+i\alpha|(\mu-1)}\begin{bmatrix}\mu\,\widehat{u}_-\\ -(1+\delta)\,\widehat{u}_+\end{bmatrix},$$

$$\widehat{w}^{(\mathrm{R})} := \frac{1}{2|1+i\alpha|(\mu-1)}\begin{bmatrix}-(1+\delta)\,\widehat{u}_-\\ \mu\,\widehat{u}_+\end{bmatrix},$$

$$\widehat{w}^{(\mathrm{X})} := \frac{1}{4|1+i\alpha|(\mu-1)}\begin{bmatrix}\big(2\mu+\delta-1+(1-\delta)|\widehat{u}_+|/|\widehat{u}_-|\big)\,\widehat{u}_-\\ \big((1-\delta)|\widehat{u}_-|/|\widehat{u}_+| + 2\mu+\delta-1\big)\,\widehat{u}_+\end{bmatrix},$$

$$\widehat{w}^{(\mathrm{Y})} := \frac{1}{4|1+i\alpha|(\mu-1)}\begin{bmatrix}\big(2\mu+\delta-1+(\delta-1)|\widehat{u}_+|/|\widehat{u}_-|\big)\,\widehat{u}_-\\ \big((\delta-1)|\widehat{u}_-|/|\widehat{u}_+| + 2\mu+\delta-1\big)\,\widehat{u}_+\end{bmatrix}.$$

(iii) Furthermore, if $|\widehat{u}_-| = |\widehat{u}_+|$ and $j \in \{0, \pm\mathrm{X}\}$, then for every λ with $0 < |\lambda| < \ell$ it holds that

$$E^{(j)}_{\widehat{u}}(\lambda) = \rho^{(j)}(\lambda)\widehat{u}$$

for some $\rho^{(j)}(\lambda) \in \mathbb{C}$.

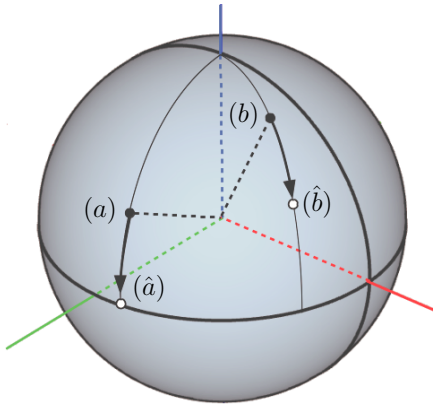


Figure 9: Illustration of amplitudes of injected field u and corresponding injection-locked field E on the normalized Poincaré sphere [115]. Points (a) and (b) denote two different values for the injected field $u = \lambda\widehat{u}$, and $(\hat{a})$ and $(\hat{b})$ are the corresponding values for the injection-locked field $E = E_{\widehat{u}}^{(+\mathrm{X})}(\lambda)$. In (a) the field u is weak, then the injection-locked field E is obtained by projecting (a) onto the equator. In (b) the injected field u is stronger, then the amplitude E of the injection-locked field moves from u toward the equator, yet, does not reach it.

Above the condition $|\widehat{u}_-| = |\widehat{u}_+|$ in the last item means that the injected field is linearly polarized. From the asymptotics (49) it follows that in the limit $|\lambda| \to 0$ the amplitude $E_{\widehat{u}}^{(0)}(\lambda)$ vanishes, the amplitudes $E_{\widehat{u}}^{(\pm\mathrm{L})}(\lambda)$ and $E_{\widehat{u}}^{(\pm\mathrm{R})}(\lambda)$ become left and right circularly polarized, respectively, and the amplitudes $E_{\widehat{u}}^{(\pm\mathrm{X})}(\lambda)$ and $E_{\widehat{u}}^{(\pm\mathrm{Y})}(\lambda)$ become linearly polarized and orthogonal to each other. The index set $\mathcal{J}$ is chosen to reflect this fact.

Following theorem says that, assuming $\alpha = 0$, only one of the above equilibrium points is asymptotically stable, while the rest are unstable. The stable equilibrium point corresponds to the injection-locked field; unstable equilibria are not physical, but they are important because they help to understand the phase space of the system. The result is as follows (Theorem 12 in Publication IV):

Theorem 8. *Consider system* (47) *under the assumption that* $\alpha = 0$ *and that the injected field* u *is of the form* $u = \lambda\widehat{u}$*, where* $\lambda \in \mathbb{C} \setminus \{0\}$ *and* $\widehat{u} = (\widehat{u}_-, \widehat{u}_+) \in \mathbb{C}^2$ *satisfies* $\widehat{u}_- \neq 0$ *and* $\widehat{u}_+ \neq 0$*. With reference to Theorem 7, let* $\ell > 0$ *be a constant and* $E_{\widehat{u}}^{(j)}(\lambda)$*,* $j \in \mathcal{J}$*, the functions with asymptotics* (49) *such that for* $0 < |\lambda| < \ell$ *they determine the nine equilibrium points of system* (47) *with injected field* $u = \lambda\widehat{u}$*.*

There exists a constant $0 < \ell_0 \leq \ell$ *such that for every* $0 < |\lambda| < \ell_0$ *the equilibrium point corresponding to* $E_{\widehat{u}}^{(+\mathrm{X})}(\lambda)$ *is asymptotically stable, and the other eight equilibrium points corresponding to* $E_{\widehat{u}}^{(j)}(\lambda)$ *with* $j \in \{0, \pm\mathrm{L}, \pm\mathrm{R}, -\mathrm{X}, \pm\mathrm{Y}\}$ *are unstable.*

See figure 9 for a geometric interpretation about the relationship between the

injected field $u = \lambda\widehat{u}$ and the injection-locked field E, which is the unique stable equilibrium point $E^{(+\mathrm{x})}_{\widehat{u}}(\lambda)$.

There exists only a single equilibrium point if the injected field is strong enough (Theorem 17 in Publication IV):

Theorem 9. *Consider $\widehat{u} = (\widehat{u}_-, \widehat{u}_+) \in \mathbb{C}^2$ with $\widehat{u}_- \neq 0$ and $\widehat{u}_+ \neq 0$. There exists a constant $L = L(\widehat{u}) > 0$ and a continuous function*

$$E_{\widehat{u}} : \{\lambda \in \mathbb{C} : |\lambda| \geq L\} \to \mathbb{C}^2$$

with the following property: If in system (47) the injected field u is of the form $u = \lambda\widehat{u}$ with $|\lambda| \geq L$, then a triple $(E, N, n) \in \mathbb{C}^2 \times \mathbb{R} \times \mathbb{R}$ is an equilibrium point of the system, if and only if

$$E = E_{\widehat{u}}(\lambda) \text{ and } (N, n) = y(|E_-|, |E_+|).$$

Furthermore, there exists a constant $C = C(\widehat{u}) > 0$ such that the function $E_{\widehat{u}}$ satisfies

$$E_{\widehat{u}}(\lambda) = \frac{\lambda e^{i\theta}}{|1 + i\alpha|}(\widehat{u} + e(\lambda)), \text{ where } |e(\lambda)| \leq \frac{C}{|\lambda|^{2/3}} \text{ and } \theta := -\arg(1 + i\alpha). \quad (50)$$

In the second part of the paper we propose a design for an optical neural network. The signals in this network are complex-valued electric fields (generated by lasers), and the activation function is computed by the physical process of injection locking. See figure 10 for an illustration of the design of the optical network.

The electric fields in the optical network are assumed to share a common linear polarization $\widehat{u}$. If $\alpha = 0$ and the injected field $\lambda\widehat{u}$ is weak enough, then by Theorem 7 and Theorem 8 there exists a unique stable equilibrium point, and that equilibrium point is of the form $\rho^{(+\mathrm{x})}(\lambda)\,\widehat{u}$. This is the injection-locked field. We interpret this so that the process of injection locking computes the function $\lambda \mapsto \rho^{(+\mathrm{x})}(\lambda) =: \rho(\lambda)$ (in Publication III we approximated ρ by the normalization operator $(\cdot)^0$). Note that ρ is a function

$$\rho : \{\lambda \in \mathbb{C} : 0 < |\lambda| < \ell\} \to \mathbb{C},$$

so that it is defined only in some punctured neighborhood of the origin of the complex plane.

The optical network illustrated in figure 10 computes the map $\mathcal{M}$,

$$\mathbb{C}^{i_0} \ni (\lambda_i^{(I)})_{i=1}^{i_0} \mapsto (\lambda_k^{(K)})_{k=1}^{k_0} =: \mathcal{M}((\lambda_i^{(I)})_{i=1}^{i_0}) \in \mathbb{C}^{k_0}, \quad (51)$$

where the k:th component function $\mathcal{M}_k$ of $\mathcal{M}$ is

$$\mathcal{M}_k((\lambda_i^{(I)})_{i=1}^{i_0}) := \sum_{j=1}^{j_0} c_{kj}\,\rho\left(\sum_{i=1}^{i_0} a_{ji}\lambda_i^{(I)} + b_j\right). \quad (52)$$

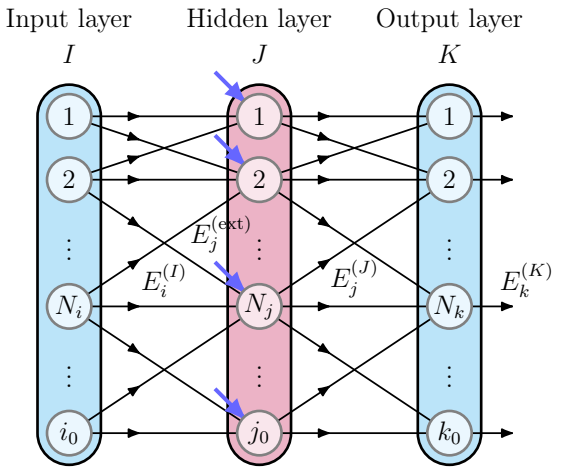


Figure 10: Schematic illustration of an optical neural network. The fields $E_i^{(I)}$ are inputs, they are of the form $E_i^{(I)} = \lambda_i^{(I)}\widehat{u}$, where $\lambda_i^{(I)} \in \mathbb{C}$ and $\widehat{u}$ is a fixed linear polarization. The inputs are passed through passive optical elements (which correspond to multiplication by $a_{ji} \in \mathbb{C}$) and joined with fixed external fields $E_j^{(\mathrm{ext})} = b_j\widehat{u}$ to form a field $(\sum_i a_{ji}\lambda_i^{(I)} + b_j)\widehat{u} = \lambda_j^{(J)}\widehat{u}$ injected into the j:th laser in the hidden layer J. Due to injection locking, the corresponding injection-locked field $E_j^{(J)}$ is $\rho(\lambda_j^{(J)})\widehat{u}$, where $\rho := \rho^{(+\mathrm{x})}$. The injection-locked fields are passed through passive optical elements (which correspond to multiplication by $c_{kj} \in \mathbb{C}$) and joined to form outputs $E_k^{(K)} = \lambda_k^{(K)}\widehat{u}$, $\lambda_k^{(K)} = \sum_j c_{kj}\rho(\lambda_j^{(J)})$, of the network.

Here $a_{ji}, b_j, c_{kj} \in \mathbb{C}$. Because ρ is not globally defined, we assume that the inputs $(\lambda_i^{(I)})_{i=1}^{i_0}$ satisfy

$$(\lambda_i^{(I)})_{i=1}^{i_0} \in \{z \in \mathbb{C}^{i_0} : |z|_{\mathbb{C}^{i_0}} \leq R\} =: \overline{B}_R$$

for some fixed $R > 0$, and that the parameters a_{ji}, b_j, c_{kj} have been chosen so that

$$\left|\sum_{i=1}^{i_0} a_{ji}\lambda_i^{(I)} + b_j\right| < \ell \text{ for every } j = 1, 2, \ldots, j_0 \text{ and } (\lambda_i^{(I)})_{i=1}^{i_0} \in \overline{B}_R. \tag{53}$$

The activation function ρ is not defined at zero, and for some input values $(\lambda_i^{(I)})_{i=1}^{i_0}$ it may happen that $\sum_{i=1}^{i_0} a_{ji}\lambda_i^{(I)} + b_j = 0$. However, the set of those inputs has measure zero in $\mathbb{C}^{i_0}$ (the measure is the $2i_0$-dimensional Lebesgue measure)

We are interested in what kind of functions can be approximated with an optical neural network of the form (51). A universal approximation theorem for complex-valued neural networks was recently proved in [130], however, that theorem assumes that the activation function ρ is globally defined on the complex plane. In our case ρ is defined only locally, so we first extend the universal approximation theorem for locally defined activation functions (Theorem 19 in Publication IV):

Theorem 10. *Consider a complex-valued neural network $\mathcal{N} : \overline{B}_R \to \mathbb{C}^{k_0}$ whose k:th component function is of the form*

$$\mathcal{N}_k(z) := \sum_{j=1}^{j_0} c_{kj}\sigma(a_j \cdot z + b_j), \tag{54}$$

where $a_j \cdot z := \sum_i a_{ji} z_i$, and $\sigma : U \to \mathbb{C}$, where $U \subset \mathbb{C}$ is an open set, is the activation function, and the parameters $a_j = (a_{j1}, a_{j2}, \ldots, a_{ji_0}) \in \mathbb{C}^{i_0}$, $j = 1, 2, \ldots, j_0$, $b \in \mathbb{C}^{j_0}$, and $(c_{kj}) \in \mathbb{C}^{k_0 \times j_0}$ are required to satisfy

$$a_j \cdot z + b_j \in U \text{ for every } z \in \overline{B}_R \text{ and } j = 1, 2, \ldots, j_0. \tag{55}$$

Suppose that

1. *σ is locally bounded and continuous almost everywhere in the nonempty open set $U \subset \mathbb{C} = \mathbb{R}^2$ (the measure is the two-dimensional Lebesgue measure), and*

2. *$\Delta^m \sigma$ does not vanish identically in U for any $m = 0, 1, 2, \ldots$ (here $\Delta = \partial^2/\partial x^2 + \partial^2/\partial y^2$, $z = x + iy$, is the Laplace operator defined in the sense of distributions).*

If $f : \overline{B}_R \to \mathbb{C}^{k_0}$ is continuous and $\epsilon > 0$, then there exists an integer $j_0 > 0$ and parameters $a_j \in \mathbb{C}^{i_0}$, $j = 1, 2, \ldots, j_0$, $b \in \mathbb{C}^{j_0}$, and $(c_{kj}) \in \mathbb{C}^{k_0 \times j_0}$ such that (55) holds, and that the complex-valued neural network $\mathcal{N}$ defined componentwise by (54) satisfies

$$\sup_{z \in \overline{B}_R} \left| \mathcal{N}(z) - f(z) \right|_{\mathbb{C}^{k_0}} \leq \epsilon. \tag{56}$$

Based on the asymptotics of the equilibrium points we proved in Theorem 7, we can conclude that ρ satisfies the conditions in Theorem 10. This implies that the optical neural network (51) is capable of approximating an arbitrary continuous function (Theorem 18 in Publication IV):

Theorem 11. *Consider an arbitrary continuous function $f : \overline{B}_R \to \mathbb{C}^{k_0}$. Let $\epsilon > 0$. There exists an integer $j_0 > 0$ and numbers $a_{ji}, b_j, c_{kj} \in \mathbb{C}$, $j = 1, 2, \ldots, j_0$, $i = 1, 2, \ldots, i_0$, $k = 1, 2, \ldots, k_0$, such that following holds:*

1. *The two-sided inequality*

$$0 < \left| \sum_{i=1}^{i_0} a_{ji} \lambda_i^{(I)} + b_j \right| < \ell$$

   holds for a.e. $(\lambda_i^{(I)})_{i=1}^{i_0} \in \overline{B}_R$ (the measure on $\overline{B}_R \subset \mathbb{C}^{i_0} = \mathbb{R}^{2i_0}$ is the $2i_0$-dimensional Lebesgue measure), and

2. *the function $\mathcal{M}$ defined componentwise a.e. in $\overline{B}_R$ by (52) is measurable and satisfies*

$$\left\| \mathcal{M} - f \right\|_{L^\infty(\overline{B}_R; \mathbb{C}^{k_0})} \leq \epsilon. \tag{57}$$

References

[1] M. L. Agranovsky and E. T. Quinto. "Injectivity sets for the Radon transform over circles and complete systems of radial functions". In: *J. Funct. Anal.* 139.2 (1996), pp. 383–414. DOI: 10.1006/jfan.1996.0090.

[2] I. Aizenberg. *Complex-valued neural networks with multi-valued neurons*. Vol. 353. Studies in Computational Intelligence. Springer-Verlag, Berlin, 2011.

[3] P. Ambs. "Optical Computing: A 60-Year Adventure." In: *Advances in Optical Technologies* (2010).

[4] P. Arena, L. Fortuna, R. Re, and M. Xibilia. "On the capability of neural networks with complex neurons in complex valued functions approximation". In: *1993 IEEE International Symposium on Circuits and Systems*. 1993, 2168–2171 vol.4. DOI: 10.1109/ISCAS.1993.394188.

[5] P. Arena, L. Fortuna, R. Re, and M. G. Xibilia. "Multilayer perceptrons to approximate complex valued functions". In: *International Journal of Neural Systems* 6.04 (1995), pp. 435–446.

[6] K. Astala and L. Päivärinta. "Calderón's inverse conductivity problem in the plane". In: *Ann. of Math. (2)* 163.1 (2006), pp. 265–299. DOI: 10.4007/annals.2006.163.265.

[7] K. Astala, L. Päivärinta, and M. Lassas. "Calderón's inverse problem for anisotropic conductivity in the plane". In: *Comm. Partial Differential Equations* 30.1-3 (2005), pp. 207–224. DOI: 10.1081/PDE-200044485.

[8] S. A. Avdonin, M. I. Belishev, and Y. S. Rozhkov. "The BC-method in the inverse problem for the heat equation". In: *J. Inverse Ill-Posed Probl.* 5.4 (1997), pp. 309–322. DOI: 10.1515/jiip.1997.5.4.309.

[9] S. Avdonin and T. I. Seidman. "Identification of $q(x)$ in $u_t = \Delta u - qu$ from boundary observations". In: *SIAM J. Control Optim.* 33.4 (1995), pp. 1247–1255. DOI: 10.1137/S0363012993249729.

[10] A. Barron. "Universal approximation bounds for superpositions of a sigmoidal function". In: *IEEE Transactions on Information Theory* 39.3 (1993), pp. 930–945. DOI: 10.1109/18.256500.

[11] S. Basu and Y. Bresler. "Uniqueness of tomography with unknown view angles". In: *IEEE Trans. Image Process.* 9.6 (2000), pp. 1094–1106. DOI: 10.1109/83.846251.

[12] M. I. Belishev. "An approach to multidimensional inverse problems for the wave equation". In: *Dokl. Akad. Nauk SSSR* 297.3 (1987), pp. 524–527.

[13] M. I. Belishev and Y. V. Kurylev. "To the reconstruction of a Riemannian manifold via its spectral data (BC–Method)". In: *Communications in partial differential equations* 17.5–6 (1992), pp. 767–804.

[14] D. ben-Avraham and S. Havlin. *Diffusion and reactions in fractals and disordered systems*. Cambridge University Press, Cambridge, 2000, pp. xiv+316. DOI: 10.1017/CBO9780511605826.

[15] B. Berkowitz, A. Cortis, M. Dentz, and H. Scher. "Modeling non-Fickian transport in geological formations as a continuous time random walk". In: *Reviews of Geophysics* 44.2 (2006).

[16] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah. "Julia: A fresh approach to numerical computing". In: *SIAM Rev.* 59.1 (2017), pp. 65–98. DOI: 10.1137/141000671.

[17] C. Bucur and E. Valdinoci. *Nonlocal diffusion and applications*. Vol. 20. Lecture Notes of the Unione Matematica Italiana. Springer, 2016.

[18] C. Bucur and E. Valdinoci. *Nonlocal diffusion and applications*. Vol. 20. Lecture Notes of the Unione Matematica Italiana. Springer, [Cham]; Unione Matematica Italiana, Bologna, 2016, pp. xii+155. DOI: 10.1007/978-3-319-28739-3.

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

[20] B. Canuto and O. Kavian. "Determining coefficients in a class of heat equations via boundary measurements". In: *SIAM J. Math. Anal.* 32.5 (2001), pp. 963–986. DOI: 10.1137/S003614109936525X.

[21] M. Caputo. "Linear models of dissipation whose Q is almost frequency independent—II". In: *Geophysical Journal International* 13.5 (1967), pp. 529–539.

[22] B. Carreras, V. Lynch, and G. Zaslavsky. "Anomalous diffusion and exit time distribution of particle tracers in plasma turbulence model". In: *Physics of Plasmas* 8.12 (2001), pp. 5096–5103.

[23] Carroll and Dickinson. "Construction of neural nets using the Radon transform". In: *International 1989 Joint Conference on Neural Networks*. 1989, 607–611 vol.1. DOI: 10.1109/IJCNN.1989.118639.

[24] E. Cartlidge. "Optical Neural Networks". In: *Optics and Photonics News* 31.6 (2020), pp. 32–39. DOI: 10.1364/OPN.31.6.000032.

[25] J. Cheng and M. Yamamoto. "Identification of convection term in a parabolic equation with a single measurement". In: *Nonlinear Anal.* 50.2, Ser. A: Theory Methods (2002), pp. 163–171. DOI: 10.1016/S0362-546X(01)00742-8.

[26] W. W. Chow, S. W. Koch, and M. I. Sargent. *Semiconductor-laser physics*. Springer Science & Business Media, 2012.

[27] A. M. Cormack. "Representation of a Function by Its Line Integrals, with Some Radiological Applications". In: *Journal of Applied Physics* 34.9 (1963), pp. 2722–2727. DOI: 10.1063/1.1729798.

[28] G. Covi. "Inverse problems for a fractional conductivity equation". In: *Nonlinear Anal.* 193 (2020), pp. 111418, 18. DOI: 10.1016/j.na.2019.01.008.

[29] D. Cox, J. Little, and D. O'Shea. *Ideals, varieties, and algorithms*. Second. Undergraduate Texts in Mathematics. An introduction to computational algebraic geometry and commutative algebra. Springer-Verlag, New York, 1997, pp. xiv+536.

[30] S. R. Deans. *The Radon transform and some of its applications*. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1983, pp. xi+289.

[31] D. del-Castillo-Negrete, B. Carreras, and V. Lynch. "Nondiffusive transport in plasma turbulence: a fractional diffusion approach". In: *Physical review letters* 94.6 (2005), p. 065003.

[32] K. Diethelm. *The analysis of fractional differential equations*. Vol. 2004. Lecture Notes in Mathematics. An application-oriented exposition using differential operators of Caputo type. Springer-Verlag, Berlin, 2010, pp. viii+247. DOI: 10.1007/978-3-642-14574-2.

[33] A. Einstein. "Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen". In: *Annalen der physik* 4 (1905).

[34] P. Elbau, M. Ritsch-Marte, O. Scherzer, and D. Schmutz. "Motion reconstruction for optical tomography of trapped objects". In: *Inverse Problems* 36.4 (2020), pp. 044004, 25. DOI: 10.1088/1361-6420/ab67db.

[35] T. Erneux and P. Glorieux. *Laser Dynamics*. Cambridge University Press, 2010.

[36] D. Finch, S. K. Patch, and Rakesh. "Determining a function from its mean values over a family of spheres". In: *SIAM J. Math. Anal.* 35.5 (2004), pp. 1213–1240. DOI: 10.1137/S0036141002417814.

[37] P. Funk. *Über Flächen mit lauter geschlossenen geodätischen Linien*. Vol. 74. 1913, pp. 278–300.

[38] T. Ghosh, Y.-H. Lin, and J. Xiao. "The Calderón problem for variable coefficients nonlocal elliptic operators". In: *Comm. Partial Differential Equations* 42.12 (2017), pp. 1923–1961. DOI: 10.1080/03605302.2017.1390681.

[39] T. Ghosh, A. Rüland, M. Salo, and G. Uhlmann. "Uniqueness and reconstruction for the fractional Calderón problem with a single measurement". In: *J. Funct. Anal.* 279.1 (2020), pp. 108505, 42. DOI: 10.1016/j.jfa.2020.108505.

[40] T. Ghosh, M. Salo, and G. Uhlmann. "The Calderón problem for the fractional Schrödinger equation". In: *Anal. PDE* 13.2 (2020), pp. 455–475. DOI: 10.2140/apde.2020.13.455.

[41] A. B. Goncharov. "Methods of integral geometry and recovering a function with compact support from its projections in unknown directions". In: *Acta Appl. Math.* 11.3 (1988), pp. 213–222. DOI: 10.1007/BF00140119.

[42] I. Goodfellow, Y. Bengio, and A. Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016.

[43] R. Gorenflo, A. A. Kilbas, F. Mainardi, and S. V. Rogosin. *Mittag-Leffler functions, related topics and applications*. Springer Monographs in Mathematics. Springer, Heidelberg, 2014, pp. xiv+443. DOI: 10.1007/978-3-662-43930-2.

[44] A. Grigor'yan. *Heat kernel and analysis on manifolds*. Vol. 47. AMS/IP Studies in Advanced Mathematics. American Mathematical Society, Providence, RI; International Press, Boston, MA, 2009, pp. xviii+482. DOI: 10.1090/amsip/047.

[45] N. Guberman. *On Complex Valued Convolutional Neural Networks*. 2016. DOI: 10.48550/ARXIV.1602.09046.

[46] H. Haken. "Analogy between higher instabilities in fluids and lasers". In: *Physics Letters A* 53.1 (1975), pp. 77–78.

[47] B. Hanin. "Universal Function Approximation by Deep Neural Nets with Bounded Width and ReLU Activations". In: *Mathematics* 7.10 (2019). DOI: 10.3390/math7100992.

[48] M. Hedley and H. Yan. "Motion artifact suppression: a review of post-processing techniques". In: *Magnetic resonance imaging* 10.4 (1992), pp. 627–635.

[49] S. Helgason. *Integral geometry and Radon transforms*. Springer, New York, 2011, pp. xiv+301. DOI: 10.1007/978-1-4419-6055-9.

[50] T. Helin, M. Lassas, L. Oksanen, and T. Saksala. "Correlation based passive imaging with a white noise source". In: *J. Math. Pures Appl. (9)* 116 (2018), pp. 132–160. DOI: 10.1016/j.matpur.2018.05.001.

[51] T. Helin, M. Lassas, L. Ylinen, and Z. Zhang. "Inverse problems for heat equation and space–time fractional diffusion equation with one measurement". In: *J. Differential Equations* 269.9 (2020), pp. 7498–7528. DOI: 10.1016/j.jde.2020.05.022.

[52] B. I. Henry, T. A. Langlands, and P. Straka. "An introduction to fractional diffusion". In: *Complex Physical, Biophysical and Econophysical Systems*. World Scientific, 2010, pp. 37–89.

[53] A. Hirose. *Complex-valued neural networks*. 2nd ed. Vol. 400. Springer-Verlag, Berlin, 2012.

[54] K. Hornik. "Some new results on neural network approximation". In: *Neural Networks* 6.8 (1993), pp. 1069–1072.

[55] K. Hornik, M. Stinchcombe, and H. White. "Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks". In: *Neural Networks* 3.5 (1990), pp. 551–560. DOI: https://doi.org/10.1016/0893-6080(90)90005-6.

[56] C. Huygens. *Œeuvres complètes de Christiaan Huygens, Correspondance 1654–1665*. (letters; no. 1333 of 24 February 1665, no. 1335 of 26 February 1665, no. 1345 of 6 March 1665). Societe Hollandaise Des Sciences, Martinus Nijho, La Haye, 1893.

[57] Y. Ito. "Differentiable approximation by means of the Radon transformation and its applications to neural networks". In: *Journal of Computational and Applied Mathematics* 55.1 (1994), pp. 31–50. DOI: https://doi.org/10.1016/0377-0427(94)90183-X.

[58] J. Janno and N. Kinash. "Reconstruction of an order of derivative and a source term in a fractional diffusion equation from final measurements". In: *Inverse Problems* 34.2 (2018), pp. 025007, 19. DOI: 10.1088/1361-6420/aaa0f0.

[59] D. Jiang, Z. Li, Y. Liu, and M. Yamamoto. "Weak unique continuation property and a related inverse source problem for time-fractional diffusion-advection equations". In: *Inverse Problems* 33.5 (2017), pp. 055013, 22. DOI: 10.1088/1361-6420/aa58d1.

[60] B. Jin and W. Rundell. "A tutorial on inverse problems for anomalous diffusion processes". In: *Inverse Problems* 31.3 (2015), pp. 035003, 40. DOI: 10.1088/0266-5611/31/3/035003.

[61] F. John. *Plane waves and spherical means applied to partial differential equations*. Interscience Publishers, New York-London, 1955, pp. viii+172.

[62] A. C. Kak and M. Slaney. *Principles of computerized tomographic imaging*. Vol. 33. Classics in Applied Mathematics. Reprint of the 1988 original. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2001, pp. xiv+327. DOI: 10.1137/1.9780898719277.

[63] M. Kalke and S. Siltanen. "Sinogram interpolation method for sparse-angle tomography". In: *Applied Mathematics* 2014 (2014). DOI: 10.4236/am.2014.53043.

[64] A. Katchalov, Y. Kurylev, M. Lassas, and N. Mandache. "Equivalence of time-domain inverse problems and boundary spectral problems". In: *Inverse Problems* 20.2 (2004), pp. 419–436. DOI: 10.1088/0266-5611/20/2/007.

[65] A. Katchalov, Y. Kurylev, and M. Lassas. *Inverse boundary spectral problems*. Vol. 123. Chapman & Hall/CRC Monographs and Surveys in Pure and Applied Mathematics. Chapman & Hall/CRC, Boca Raton, FL, 2001, pp. xx+290. DOI: 10.1201/9781420036220.

[66] J. Ketola and L. Lamberg. "An algorithm for recovering unknown projection orientations and shifts in 3-D tomography". In: *Inverse Probl. Imaging* 5.1 (2011), pp. 75–93. DOI: 10.3934/ipi.2011.5.75.

[67] Y. Khanin. *Principles of Laser Dynamics*. Elsevier Science, 2012.

[68] Y. Kian, L. Oksanen, E. Soccorsi, and M. Yamamoto. "Global uniqueness in an inverse problem for time fractional diffusion equations". In: *J. Differential Equations* 264.2 (2018), pp. 1146–1170. DOI: 10.1016/j.jde.2017.09.032.

[69] Y. Kian. *Simultaneous determination of coefficients, internal sources and an obstacle of a diffusion equation from a single measurement*. 2020. DOI: 10.48550/ARXIV.2007.08947.

[70] Y. Kian, Z. Li, Y. Liu, and M. Yamamoto. "The uniqueness of inverse problems for a fractional equation with a single measurement". In: *Math. Ann.* 380.3-4 (2021), pp. 1465–1495. DOI: 10.1007/s00208-020-02027-z.

[71] Y. Kian, E. Soccorsi, Q. Xue, and M. Yamamoto. *Identification of time-varying source term in time-fractional diffusion equations*. 2019. DOI: 10.48550/ARXIV.1911.09951.

[72] Y. Kian, E. Soccorsi, and M. Yamamoto. "On time-fractional diffusion equations with space-dependent variable order". In: *Ann. Henri Poincaré* 19.12 (2018), pp. 3855–3881. DOI: 10.1007/s00023-018-0734-y.

[73] Y. Kian and M. Yamamoto. "Reconstruction and stable recovery of source terms and coefficients appearing in diffusion equations". In: *Inverse Problems* 35.11 (2019), pp. 115006, 24. DOI: 10.1088/1361-6420/ab2d42.

[74] A. A. Kilbas, H. M. Srivastava, and J. J. Trujillo. *Theory and applications of fractional differential equations*. Vol. 204. North-Holland Mathematics Studies. Elsevier Science B.V., Amsterdam, 2006, pp. xvi+523.

[75] A. Kubica, K. Ryszewska, and M. Yamamoto. *Time-fractional differential equations—a theoretical introduction*. SpringerBriefs in Mathematics. Springer, Singapore, 2020, pp. x+134. DOI: 10.1007/978-981-15-9066-5.

[76] M. Kwaśnicki. "Ten equivalent definitions of the fractional Laplace operator". In: *Fract. Calc. Appl. Anal.* 20.1 (2017), pp. 7–51. DOI: 10.1515/fca-2017-0002.

[77] L. Lamberg. "Unique recovery of unknown projection orientations in three-dimensional tomography". In: *Inverse Probl. Imaging* 2.4 (2008), pp. 547–575. DOI: 10.3934/ipi.2008.2.547.

[78] L. Lamberg and L. Ylinen. "Two-dimensional tomography with unknown view angles". In: *Inverse Probl. Imaging* 1.4 (2007), pp. 623–642. DOI: 10.3934/ipi.2007.1.623.

[79] E. K. Lau, L. J. Wong, and M. C. Wu. "Enhanced modulation characteristics of optical injection-locked lasers: A tutorial". In: *IEEE Journal of Selected Topics in Quantum Electronics* 15.3 (2009), pp. 618–633.

[80] Y. LeCun, Y. Bengio, and G. Hinton. "Deep learning". In: *Nature* 521.7553 (2015), pp. 436–444.

[81] J. M. Lee and G. Uhlmann. "Determining anisotropic real-analytic conductivities by boundary measurements". In: *Comm. Pure Appl. Math.* 42.8 (1989), pp. 1097–1112. DOI: 10.1002/cpa.3160420804.

[82] T. von Lerber, M. Lassas, V. S. Lyubopytov, L. Ylinen, A. Chipouline, K. Hofmann, and F. Küppers. "All-optical majority gate based on an injection-locked laser". In: *Scientific reports* 9.1 (2019), pp. 1–7. DOI: 10.1038/s41598-019-51025-y.

[83] M. Leshno, V. Y. Lin, A. Pinkus, and S. Schocken. "Multilayer feedforward networks with a nonpolynomial activation function can approximate any function". In: *Neural Networks* 6.6 (1993), pp. 861–867.

[84] Y. Liu, W. Rundell, and M. Yamamoto. "Strong maximum principle for fractional diffusion equations and an application to an inverse source problem". In: *Fract. Calc. Appl. Anal.* 19.4 (2016), pp. 888–906. DOI: 10.1515/fca-2016-0048.

[85] Z. Liu and R. Slavík. "Optical injection locking: From principle to applications". In: *Journal of Lightwave Technology* 38.1 (2019), pp. 43–59.

[86] E. N. Lorenz. "Deterministic nonperiodic flow". In: *J. Atmospheric Sci.* 20.2 (1963), pp. 130–141. DOI: 10.1175/1520-0469(1963)020<0130:DNF>2.0.CO;2.

[87] Z. Lu, H. Pu, F. Wang, Z. Hu, and L. Wang. "The expressive power of neural networks: a view from the width". In: *Proceedings of the 31st International Conference on Neural Information Processing Systems*. 2017, pp. 6232–6240.

[88] R. L. Magin. "Fractional calculus models of complex dynamics in biological tissues". In: *Comput. Math. Appl.* 59.5 (2010), pp. 1586–1593. DOI: 10.1016/j.camwa.2009.08.039.

[89] J. Martin-Regalado, F. Prati, M. San Miguel, and N. B. Abraham. "Polarization properties of vertical-cavity surface-emitting lasers". In: *IEEE Journal of Quantum Electronics* 33.5 (1997), pp. 765–783.

[90] W. S. McCulloch and W. Pitts. “A logical calculus of the ideas immanent in nervous activity”. In: *The bulletin of mathematical biophysics* 5.4 (1943), pp. 115–133.

[91] H. N. Mhaskar. “Neural Networks for Optimal Approximation of Smooth and Analytic Functions”. In: *Neural Computation* 8.1 (Jan. 1996), pp. 164–177. DOI: 10.1162/neco.1996.8.1.164.

[92] F. Natterer. *The mathematics of computerized tomography*. B. G. Teubner, Stuttgart; John Wiley & Sons, Ltd., Chichester, 1986, pp. x+222.

[93] M. A. Nielsen. *Neural networks and deep learning*. Vol. 25. Determination press San Francisco, CA, 2015.

[94] L. Perko. *Differential equations and dynamical systems*. Third. Vol. 7. Texts in Applied Mathematics. Springer-Verlag, New York, 2001, pp. xiv+553. DOI: 10.1007/978-1-4613-0003-8.

[95] J. Perrin. “Mouvement brownien et réalité moléculaire”. In: *Ann. de Chim. et de Phys.* 18 (1909), pp. 1–114.

[96] A. Pikovsky, J. Kurths, M. Rosenblum, and J. Kurths. *Synchronization: a universal concept in nonlinear sciences*. 12. Cambridge university press, 2003.

[97] A. Pinkus. “Approximation theory of the MLP model in neural networks”. In: *Acta Numerica* 8 (1999), pp. 143–195. DOI: 10.1017/S0962492900002919.

[98] C. Rackauckas and Q. Nie. “DifferentialEquations.jl–A performant and feature-rich ecosystem for solving differential equations in Julia”. In: *Journal of Open Research Software* 5.1 (2017). DOI: 10.5334/jors.151.

[99] J. Radon. “On the determination of functions from their integral values along certain manifolds”. In: *IEEE transactions on medical imaging* 5.4 (1986), pp. 170–176.

[100] J. Radon. “Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten”. In: *Berichte über die Verhandlungen der Königlich-Sächsischen Akademie der Wissenschaften zu Leipzig, Mathematisch-Physische Klasse* 69.4 (1917), pp. 262–277.

[101] Rakesh and W. W. Symes. “Uniqueness for an inverse problem for the wave equation”. In: *Comm. Partial Differential Equations* 13.1 (1988), pp. 87–96. DOI: 10.1080/03605308808820539.

[102] M. Rantala, S. Vänskä, S. Järvenpää, M. Kalke, M. Lassas, J. Moberg, and S. Siltanen. “Wavelet-based reconstruction for limited-angle X-ray tomography”. In: *IEEE Transactions on Medical Imaging* 25.2 (2006), pp. 210–217. DOI: 10.1109/TMI.2005.862206.

[103] X. Ros-Oton. “Nonlocal elliptic equations in bounded domains: a survey”. In: *Publ. Mat.* 60.1 (2016), pp. 3–26.

[104] W. Rudin. *Functional analysis*. Second. International Series in Pure and Applied Mathematics. McGraw-Hill, Inc., New York, 1991, pp. xviii+424.

[105] J. Sacher, D. Baums, P. Panknin, W. Elsässer, and E. O. Göbel. "Intensity instabilities of semiconductor lasers under current modulation, external light injection, and delayed feedback". In: *Physical Review A* 45.3 (1992), p. 1893.

[106] D. B. Salzman. "A method of general moments for orienting 2D projections of unknown 3D objects". In: *Computer vision, graphics, and image processing* 50.2 (1990), pp. 129–156.

[107] S. G. Samko, A. A. Kilbas, and O. I. Marichev. *Fractional integrals and derivatives*. Theory and applications, Edited and with a foreword by S. M. Nikol′skiĭ, Translated from the 1987 Russian original, Revised by the authors. Gordon and Breach Science Publishers, Yverdon, 1993, pp. xxxvi+976.

[108] M. San Miguel, Q. Feng, and J. V. Moloney. "Light-polarization dynamics in surface-emitting semiconductor lasers". In: *Physical Review A* 52.2 (1995), p. 1728.

[109] M. J. Saxton. "A biological interpretation of transient anomalous subdiffusion. I. Qualitative model". In: *Biophysical journal* 92.4 (2007), pp. 1178–1191.

[110] F. Scarselli and A. C. Tsoi. "Universal approximation using feedforward neural networks: A survey of some existing methods, and some new results". In: *Neural networks* 11.1 (1998), pp. 15–37.

[111] M. Schatz, E. Orlova, P. Dube, H. Stark, F. Zemlin, and M. Van Heel. "Angular reconstitution in three-dimensional electron microscopy: practical and technical aspects". In: *Scanning Microscopy* 11 (1997), pp. 179–193.

[112] J. Schmidhuber. "Deep learning in neural networks: An overview". In: *Neural Networks* 61 (2015), pp. 85–117. DOI: 10.1016/j.neunet.2014.09.003.

[113] V. A. Sharafutdinov. *Integral geometry of tensor fields*. Inverse and Ill-posed Problems Series. VSP, Utrecht, 1994, p. 271. DOI: 10.1515/9783110900095.

[114] T. Shiota. "An inverse problem for the wave equation with first order perturbation". In: *Amer. J. Math.* 107.1 (1985), pp. 241–251. DOI: 10.2307/2374463.

[115] W. A. Shurcliff. *Polarized light: Production and use*. Harvard University Press, 1962.

[116] S. Siltanen, V. Kolehmainen, S. Järvenpää, J. P. Kaipio, P. Koistinen, M. Lassas, J. Pirttilä, and E. Somersalo. "Statistical inversion for medical x-ray tomography with few radiographs: I. General theory". In: *Physics in Medicine and Biology* 48.10 (May 2003), pp. 1437–1463. DOI: 10.1088/0031-9155/48/10/314.

[117] A. Singer and Y. Shkolnisky. "Three-dimensional structure determination from common lines in cryo-EM by eigenvectors and semidefinite programming". In: *SIAM J. Imaging Sci.* 4.2 (2011), pp. 543–572. DOI: 10.1137/090767777.

[118] D. R. Solli and B. Jalali. "Analog optical computing". In: *Nature Photonics* 9.11 (2015), pp. 704–706.

[119] H. Stover and W. Steier. "Locking of laser oscillators by light injection". In: *applied physics letters* 8.4 (1966), pp. 91–93.

[120] S. H. Strogatz. *Nonlinear dynamics and chaos*. Second. With applications to physics, biology, chemistry, and engineering. Westview Press, Boulder, CO, 2015, pp. xiii+513.

[121] H. Sun, W. Chen, and Y. Chen. "Variable-order fractional differential operators in anomalous diffusion modeling". In: *Physica A: Statistical Mechanics and its Applications* 388.21 (2009), pp. 4586–4592.

[122] J. Sylvester and G. Uhlmann. "A global uniqueness theorem for an inverse boundary value problem". In: *Ann. of Math. (2)* 125.1 (1987), pp. 153–169. DOI: 10.2307/1971291.

[123] G. Thalhammer, R. Steiger, M. Meinschad, M. Hill, S. Bernet, and M. Ritsch-Marte. "Combined acoustic and optical trapping". In: *Biomedical optics express* 2.10 (2011), pp. 2859–2870.

[124] J. Touch, A.-H. Badawy, and V. J. Sorger. "Optical computing". In: *Nanophotonics* 6.3 (2017), pp. 503–505. DOI: 10.1515/nanoph-2016-0185.

[125] C. Trabelsi, O. Bilaniuk, Y. Zhang, D. Serdyuk, S. Subramanian, J. F. Santos, S. Mehri, N. Rostamzadeh, Y. Bengio, and C. J. Pal. "Deep Complex Networks". In: *International Conference on Learning Representations*. 2018.

[126] E. Valdinoci. "From the long jump random walk to the fractional Laplacian". In: *Bol. Soc. Esp. Mat. Apl. SeMA* 49 (2009), pp. 33–44.

[127] M. Van Heel, E. Orlova, G. Harauz, H. Stark, P. Dube, F. Zemlin, and M. Schatz. "Angular reconstitution in three-dimensional electron microscopy: historical and theoretical aspects". In: *Scanning Microscopy* 11 (1997), pp. 195–210.

[128] M. Van Heel. "Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction". In: *Ultramicroscopy* 21.2 (1987), pp. 111–123.

[129] P. Virtue, S. X. Yu, and M. Lustig. "Better than real: Complex-valued neural nets for MRI fingerprinting". In: *2017 IEEE International Conference on Image Processing (ICIP)*. 2017, pp. 3953–3957. DOI: 10.1109/ICIP.2017.8297024.

[130] F. Voigtlaender. *The universal approximation theorem for complex-valued neural networks*. 2020. DOI: 10.48550/ARXIV.2012.03351.

[131] S. Wieczorek, B. Krauskopf, T. B. Simpson, and D. Lenstra. "The dynamical complexity of optically injected semiconductor lasers". In: *Physics Reports* 416.1-2 (2005), pp. 1–128.

[132] S. Wiggins. *Introduction to applied nonlinear dynamical systems and chaos*. Second. Vol. 2. Texts in Applied Mathematics. Springer-Verlag, New York, 2003, pp. xxx+843.

[133] M. Wolter and A. Yao. "Complex Gated Recurrent Neural Networks". In: *Advances in Neural Information Processing Systems*. Ed. by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett. Vol. 31. Curran Associates, Inc., 2018.

[134] L. Ylinen, T. von Lerber, F. Küppers, and M. Lassas. "Analysis of a dynamical system modeling lasers and applications for optical neural networks". In: *SIAM J. Appl. Dyn. Syst.* 21.2 (2022), pp. 840–878. DOI: 10.1137/21M1405976.

[135] L. Zalcman. "Uniqueness and nonuniqueness for the Radon transform". In: *Bull. London Math. Soc.* 14.3 (1982), pp. 241–245. DOI: 10.1112/blms/14.3.241.

[136] Y. Zhang, D. A. Benson, M. M. Meerschaert, and E. M. LaBolle. "Space-fractional advection-dispersion equations with variable parameters: Diverse formulas, numerical solutions, and application to the Macrodispersion Experiment site data". In: *Water Resources Research* 43.5 (2007).