Chapter 1

Preamble

The mathematical formulation of *fractional calculus* and the underpinnings of the *physics behind diffusion* both have their origins in the first half of the nineteenth century, but it is only in the last half of the twentieth century that these two topics have been seen to be related. The goal in this preamble is to indicate why these two ideas coalesced.

1.1. Diffusion and Brownian motion

One way to understand diffusion processes is provided by the theory of random walks. Perhaps the simplest, certainly the most well-known, example here is the so-called Brownian motion in which the dynamics is governed by an uncorrelated, Markovian, Gaussian stochastic process. In this case, the dynamics can be described at a macroscopic level using the diffusion equation. This connection, one that dates from at least the early 1900s, provides a transparent view of the diffusion model and allows effective solution by means of a well-understood partial differential equation (PDE). However, as we shall see, it also points out its limitations. In particular, when the random walk involves correlations, non-Gaussian statistics, or non-Markovian "memory" effects, the diffusion equation cannot describe the macroscopic limiting process. We shall see that one way to extend the classical diffusion paradigm is to extend the idea of a derivative, an idea that in itself has has a long mathematical history.

The roots of modern theory of diffusion processes are to be found in the early decades of the nineteenth century. The work of Fourier in formulating his famous heat equation comes to mind. Fourier's arguments [**108**] were from a macroscopic viewpoint but from a particle-diffusion perspective.

The first systematic study was by the Scottish chemist, Thomas Graham. Graham was the inventor of dialysis through a method of separation of materials by diffusion through a membrane. His research work on diffusion in gases was performed beginning in 1828 and finally published in 1833 [**118**]. He observed that, "gases of different nature, when brought into contact, do not arrange themselves according to their density, the heaviest undermost, and the lighter uppermost, but they spontaneously diffuse, mutually and equally, through each other, and so remain in the intimate state of mixture for any length of time." Graham did not only perform the first quantitative experiments on diffusion, but was able to make reliable measurements of the coefficient of diffusion.

At about the same time period, in what turns out to be parallel work, is the celebrated *Brownian motion* which comes from the investigations of another Scottish scientist, the botanist Robert Brown in 1827 [**38**]. His initial observations were of the seemingly chaotic movement of pollen grains suspended in water when viewed under a microscope. He also observed a similar motion in particles of inorganic matter ruling out the possibility that the effect was life-related. However, Brown did not provide a mathematical framework for the phenomenon.

In fact, amongst the first to address the problem of diffusion from a differential-equation viewpoint was the German physiologist Adolf Fick [**105**]. Following on from the work of Graham and also aware of Fourier's heat equation formulation, Fick was investigating the way that fluids and nutrients travel through membranes in living organisms. His eponymous law postulates that the diffusive flux would flow from regions of high concentration to regions of low. Mathematically, this means that if $u(x,t)$ is the concentration at point x at time t, then J, the amount of material per unit volume per unit time, will satisfy $J \propto \nabla u$ or $J = -D\nabla u$, where D is the diffusivity. Combining this with the conservation law $u_t = -\operatorname{div} J$ (the rate of change of concentration in time is proportional to the net amount of flux leaving the region and where we have chosen the units to obtain the proportionality constant unity) gives the heat equation $u_t = D\operatorname{div}(\nabla u)$ or $u_t = Du_{xx}$ in one space dimension. Of course, this was much less of a *first principles* approach than it was an analogy with Fourier's derivation of the heat equation for temperature flow some 50 years earlier.

Einstein's 1905 paper on the topic [**92**] provided the now generally accepted explanation for Brownian motion and had far-reaching consequences for physics. Indeed, he published four famous papers that year, the others being on special relativity [**94**], the correlation between matter and energy (the famous $E = mc^2$) [**91**], and on the photoelectric effect [**93**]. While the work on the first two are certainly the most celebrated—and the 1921 Nobel prize was awarded for the photoelectric paper—the most often-cited is the one on Brownian motion. There are two key pieces to this work: first, the assumption that a change in the direction of motion of a particle is random and that the mean-squared displacement over many changes is proportional to time; second, he combined this with the Boltzmann distribution for a system in thermal equilibrium to get a value on the proportionality, the *diffusivity* D. Thus $\langle x^2 \rangle = 2Dt,\;\; D = \frac{RT}{6N\pi\eta a}$, where T is the temperature, R is the universal gas constant, N is Avogadro's number, and $\gamma = 6\pi\eta a$ where a is the particle radius and η is the viscosity, is Stokes's relation for the viscous drag coefficient. With this he was able to predict the properties of the unceasing motion of Brownian particles in terms of collisions with surrounding liquid molecules.

The French physicist Jean Baptiste Perrin, did a series of experiments beginning in 1908, one of which verified Einstein's equations and led to acceptance of the atomic and molecular-kinetic theory. Perrin received the Nobel prize for this work in 1926.

In the same year that Einstein was formulating the Brownian motion concept into a mathematical model, the British statistician Karl Pearson, in his letter to *Nature* [**263**], was asking the question:

> A man starts from the point O and walks ℓ yards in a straight line; he then turns through any angle whatever and walks another ℓ yards in a second straight line. He repeats this process n times. I require the probability that after n stretches he is at a distance between r and $r + \delta$ from his starting point O.

A solution was provided almost immediately by Lord Rayleigh (who had been studying similar questions since 1880):

> If n be very great, the probability sought is $\frac{2}{n}e^{-r^2/n}\, r\, dr$.

This prompted the retort by Pearson:

> The lesson of Lord Rayleigh's solution is that in open country the most probable place to find a drunken man who is at all capable of keeping on his feet is somewhere near his starting point!

From this one can infer the origins of the term "drunkard's walk" as an alternative to "random walk".

Similar results have also been verified in numerous experiments including Perrin's measurements of mean square displacements to determine Avogadro's number. Thus, classical Brownian motion can be viewed as a random walk in which the dynamics is governed by an uncorrelated, Markovian, Gaussian stochastic process. On the other hand, when the random walk involves correlations, non-Gaussian statistics, or a non-Markovian process (for example, due to *memory* effects), the diffusion equation will fail to describe the macroscopic limit.

Indeed, for an increasing number of processes the overall motion of an object is better described by steps that are not independent, identically distributed copies of each other and that can take different times to perform. In addition, the inclusion of memory effects (thus removing the assumption of a Markov process) can be particularly important. These considerations lead to so-called *anomalous diffusion* processes and the physically motivated examples are numerous. These include the foraging of animals: think of the random walk that asks one to minimize the time taken to go from a given starting place to an unknown region where food is available; as Rayleigh noted, the search path is definitely not Brownian.

Actually, our issue with the fixed paradigms of Brownian motion should not lie only with the underlying assumptions, we should ask for the outcomes of the model to be satisfactory from a physical perspective. Rayleigh's observation that in Brownian motion the particles have a high probability of being near their starting position can be seen in terms of the probability density function given by the Gaussian; a relatively slow diffusion initially but a very rapid decay of the plume in space. It has certainly been observed that many processes do not exhibit this effect.

1.2. Early fractional integrals and their inverses

There is the well-known story of L'Hôpital asking Leibniz about the possibility that n be a fraction in the formula $\frac{d^n}{x^n}$ and the reply in 1695, "It will lead to a paradox." But he added prophetically, "From this apparent paradox, one day useful consequences will be drawn" [**209**]. The idea of such a fractional derivative had already occurred to Leibnitz through the observation that the basic formula $\frac{d}{dx}x^n = nx^{n-1}$ makes sense for noninteger values (and specifically in Leibniz's case with $\alpha = \frac{1}{2}$).

There were many attempts at such generalisations most of which were based on replacing formulae for integer order derivatives by a noninteger value. Mathematicians frequently ask if a definition can be extended to a broader set, and one of the earliest successes in the direction of an extended derivative was Euler's generalisation of the factorial function to the Gamma function in 1725. For this purpose he used the representation

$$n! = \prod_{k=1}^{n} k = \int_0^1 \left(-\log(x)\right)^n dx \tag{1.1}$$

to obtain the formula (in modern notation)

$$\frac{d^\alpha x^\beta}{dx^\alpha} = \frac{\Gamma(\beta+1)}{\Gamma(\beta-\alpha+1)} x^{\beta-\alpha}, \tag{1.2}$$

valid for all real α and β and in complete analogy to the case when α is an integer. Of course generalising this beyond powers of x was another matter entirely and there turned out to be more complexity and indeed obstacles, not the least of which is the fact that the fractional derivative of a constant will not be zero. Indeed the original work by Euler and others in describing the Gamma function itself really only came to full fruition with the tools of complex analysis. We will see a similar pattern for the generalisation of the derivative to noninteger values.

The next significant mention of a derivative of noninteger order appears in a long textbook by the French mathematician, S. F. Lacroix [**206**] that in its day had widespread impact. Lacroix also noted the formula

$$\frac{d^n x^m}{dx^n} = \frac{n!}{(n-m)!} x^{n-m}$$

as being valid for integers n and m. Again, using the then available formula for the Gamma function extension of the factorial [**208**], replacing m by $\frac{1}{2}$ and n by any real number a, he obtained the formal expression

$$\frac{d^{\frac{1}{2}}}{dx^{\frac{1}{2}}} x^a = \frac{\Gamma(a+1)}{\Gamma\left(a+\frac{1}{2}\right)} x^{a-\frac{1}{2}}.$$

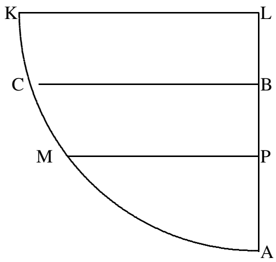
He gives the explicit case for $n = 1$, $y = x$ then as

$$\frac{d^{\frac{1}{2}}}{dx^{\frac{1}{2}}} x = \frac{2}{\sqrt{\pi}} x^{\frac{1}{2}}$$

since $\Gamma\left(\frac{3}{2}\right) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{1}{2}\sqrt{\pi}$. This turns out to agree with the more rigorous formula obtained some years later by a variety of authors.

The key to the next phase included an application, namely the *tautochrone problem* of finding the curve down which a bead placed anywhere will fall to the bottom in the same amount of time. The solution was originally found by Huygens [**154**] in a formal way, but in 1823 Abel [**1**, **2**] provided the rigorous mathematical solution to the general tautochrone problem:[1]

[1] The below is taken directly from [**273**] which is in turn taken from the translations [**3**].



Suppose that CB is a horizontal line, A is a setpoint, AB is perpendicular to BC, AM is a curve with rectangular coordinates AP= x, PM= y. Moreover, AB= a, AM= s. It is known that as a body moves along an arc CA, when the initial velocity is zero, that the time T, which is necessary for the passage, depends on the shape of the curve, and on a. One has to find the definition of a curve KCA, for which the time T is a given function of a, for example $\psi(a)$.

The picture above appeared in the French translation in [**3**] and Abel obtains the equation

$$\psi(a) = \int_0^a \frac{ds}{\sqrt{a-s}},$$

and he then continues:[2]

> Instead of solving this equation, I will show how one can derive s from the more general equation
>
> $$\psi(a) = \int_0^a \frac{ds}{(a-s)^n} \tag{1.3}$$
>
> where n has to be less than 1 to prevent the integral between the two limits being infinite; $\psi(a)$ is an arbitrary function that is not infinite, when $a = 0$.

Abel seeks the unknown function $s(x)$ in the form of a power series, and after term-by-term operations using Legendre's then-recently discovered properties of the Gamma function, he arrives at the solution of equation (1.3)

$$s(x) = \frac{\sin(n\pi)}{\pi} x^n \int_0^1 \frac{\psi(xt)\,dt}{(1-t)^{1-n}}. \tag{1.4}$$

As remarkable as this theorem is, [**1**] further explores these representations to obtain, in his notation,

$$s(x) = \frac{1}{\Gamma(1-n)} \int^n \psi x.dx^n = \frac{1}{\Gamma(1-n)} \frac{d^{-n}\psi x}{dx^{-n}}. \tag{1.5}$$

Thus Abel understood that he had unified the notions of integration and differentiation and their extension to noninteger orders. Equation (1.3) can be rewritten as

$$\psi(t) = \int_0^t \frac{s'(x)\,dx}{(t-x)^n}. \tag{1.6}$$

[2] We added our equation numbers for later reference.

Thus we have the main ingredient: the Abel integral operator,

$$I^\alpha f = \frac{1}{\Gamma(\alpha)} \int_0^t \frac{f(\tau)\,d\tau}{(t-\tau)^{1-\alpha}}, \qquad \alpha > 0. \tag{1.7}$$

The reader will later recognise (1.6) again in Chapter 4 as the key element in the two most commonly used "modern" fractional derivatives for applications, the so-called Riemann–Liouville and Djrbashian–Caputo derivatives.

Equally remarkable to the mathematical formulations is the fact the entire structure was based on the solution of an important application rather than generalisation for its own sake. This aspect would to a certain extent be subordinated to purely mathematical development until the middle of the twentieth century when the use of fractional operators with their inherent nonlocal nature became a building block of a wide range of quite diverse applications.

Liouville published three long memoirs in 1832 and several more in the next two decades. His starting point is the known result for derivatives of integral order $D^m\, e^{ax} = a^m e^{ax}$ which he attempted to extend in a natural way to $D^\alpha\, e^{ax} = a^\alpha e^{ax}$. He expanded the function $f(x)$ in the series $f(x) = \sum_{k=0}^\infty c_k e^{a_k x}$ and defined the derivative of order α to be

$$D^\alpha f(x) = \sum_{k=0}^\infty c_k a_k^\alpha e^{a_k x}. \tag{1.8}$$

This has the disadvantage that the allowed values of α have to be restricted to those for which the series in (1.8) converged.

As a second method applied to negative powers $f(x) = x^{-a}$ where $a > 0$, he started with the Laplace transform formula

$$x^{-a} = \frac{1}{\Gamma(a)} \int_0^\infty t^{a-1} e^{-xt}\, dt, \tag{1.9}$$

then used equation (1.8) to obtain

$$D^\alpha x^{-a} = (-1)^\alpha \frac{\Gamma(\alpha + a)}{\Gamma(a)} x^{-a-\alpha}. \tag{1.10}$$

Clearly, this formula also has limited applicability. However, if it is formally generalised to $\beta = -a$ and disregarding the existence of the resulting integral, one obtains the formula

$$\frac{d^\alpha x^\beta}{dx^\alpha} = \frac{(-1)^\alpha \Gamma(-\beta + \alpha)}{\Gamma(-\beta)} x^{\beta - \alpha}. \tag{1.11}$$

This formula is reminiscent of (1.2) and is in fact identical to it for the case of integral α. For nonintegral α things are different: taking $\alpha = \frac{1}{2}$ and

$\beta = -\frac{1}{2}$ gives

$$\frac{\Gamma(\frac{1}{2})}{\Gamma(0)}\frac{1}{x} = 0 \neq \frac{i}{\sqrt{\pi}}\frac{1}{x} = \frac{(-1)^{1/2}\Gamma(1)}{\Gamma(\frac{1}{2})}\frac{1}{x}. \tag{1.12}$$

In fact by further analysis (see [**146**]) it can be shown that the definitions of Euler and Liouville differed in their limits of integration. In fact the necessity of the correct limits of integration is central (and subsequent definitions realised that the definition of a fractional derivative *required* a starting point—as indeed Abel was aware and in fact used).

Liouville recast his formula (1.8) as a fractional integral of order α

$$\int^{\alpha} f(x)\,dx^{\alpha} = \frac{1}{(-1)^{\alpha}\Gamma(\alpha)}\int_0^{\infty} f(x+y)y^{\alpha-1}\,dy \tag{1.13}$$

and then derived a formula for fractional differentiation

$$\frac{d^{\alpha}f}{dx^{\alpha}} = \frac{(-1)^{n-\alpha}}{\Gamma(n-\alpha)}\int_0^{\infty}\frac{d^n f(x+y)}{dx^n}y^{n-\alpha-1}\,dy, \qquad n-1<\alpha<n, \tag{1.14}$$

where f is defined by his exponential formula (1.8) with $\lambda_n > 0$ and $f(-\infty) = 0$. See [**220**]. He further inserted the representation (1.8) directly into the above to obtain the series representation

$$\frac{d^{\alpha}f}{dx^{\alpha}} = \lim_{h\to 0}\left[\frac{(-1)^{\alpha}}{h^{\alpha}}\sum_{k=0}^{\infty}(-1)^k\binom{\alpha}{k}f(x+kh)\right], \tag{1.15}$$

$$\text{where } \binom{\alpha}{k} = \frac{\Gamma(\alpha-1)\Gamma(k-1)}{\Gamma(\alpha+k-1)}.$$

Riemann's work here stemmed from his student days and was published only posthumously in 1875 [**287**]. The definition taken was more Abelian in character: the fractional integral of order α was

$$I^{\alpha}f(x) = \frac{1}{\Gamma(\alpha)}\int_c^x (x-t)^{\alpha-1}f(t)\,dt + \psi(x). \tag{1.16}$$

Here c is a given constant and $\psi(x)$ represents the *constant of integration* or *complementary function* needed in some form. The latter caused considerable confusion, one representation being

$$\psi(x) = \sum_{k=1}^{\infty} K_k \frac{x^{-\alpha-k}}{\Gamma(-k-\alpha+1)}, \tag{1.17}$$

where K_k are finite constants.

As we will see, the modern resolution of this is to view the fractional integral and the derivative derived from it as being defined for a fixed starting point c which is then part of the definition.

With this Riemann suggested that the fractional derivative of order α be defined by taking the ordinary derivative of the Abel integral of order $1-\alpha$,

$$[D^\alpha f](x) = \frac{d}{dx}[I^{1-\alpha} f](x). \tag{1.18}$$

More modern notation would write this as ${}^{\mathrm{RL}}D^\alpha f$ to incorporate the contribution of Liouville in the notation. If the historical record had been more accurate, this might be more appropriately be written ${}^{\mathrm{A}}D^\alpha$.

Despite the above drawbacks in the approaches of both Liouville and Riemann, their names formed the nomenclature from the late nineteenth century: the fractional integral is a weakly singular integral of Abel type and the fractional derivative of order α is the regular derivative of the fractional integral of order $1-\alpha$.

There were other nineteenth century attempts at defining the concept of a fractional derivative that have survived the passage of time. Foremost amongst these is the so-called Grünwald–Letnikov derivative.

The idea of Grünwald was to extend the basic difference limit for the derivative to higher orders,

$$f'(x) = \lim_{h\to 0} \frac{f(x+h)-f(x)}{h},$$
$$f''(x) = \lim_{h\to 0} \frac{f(x+2h)-2f(x+h)+f(x)}{h^2}, \quad \ldots;$$

the general case uses the binomial theorem to obtain

$$f^{(n)}(x) = \lim_{h\to 0} \frac{1}{h^n} \sum_{k=0}^{n} (-1)^k \binom{n}{k} f\big(x+(n-k)h\big). \tag{1.19}$$

Removing the restriction that n be an integer in (1.19) gives

$$D^\alpha f(x) = \lim_{h\to 0} \frac{1}{h^\alpha} \sum_{k=0}^{\infty} (-1)^k \binom{\alpha}{k} f\big(x+(\alpha-k)h\big). \tag{1.20}$$

This idea of an infinite sum of difference quotients was certainly inspired by the work of Liouville. Note that the computation of this derivative requires knowledge of the entire past history of the system due to the summation operator and thus, as in the case of the Abel formulation, it must be viewed as a nonlocal operator. The above ideas were taken further into what is now known as the Grünwald–Letnikov derivative.

We now take the theme into the twentieth century. The motivation for the work of Weyl on fractional integrals was due to the fact that the Abel integral does not preserve periodicity; f being periodic of, say, period 2π does not guarantee that integrals $I^\alpha f$ have the same period or even that they

are periodic at all. This is simply a statement that this fractional derivative is better suited to integration of functions defined by power series than by Fourier series. Functions periodic on the unit circle $f = \sum_{-\infty}^{\infty} c_k e^{2\pi ik}$ for which the integral vanishes ($c_0 = 0$) will, when integrated, be again periodic after choosing the constant of integration appropriately. Thus if $f \in L^p(0, 2\pi)$ with $1 \le p < \infty$ such that f is periodic of period 2π and the integral of f vanishes, then the Weyl fractional integral of order α [**346**] is
(1.21)

$$[^W I^\alpha_\pm f](x) = \frac{1}{2\pi}\int_0^{2\pi} \Psi^\alpha_\pm(x-y) f(y)\,dy, \qquad \Psi^\alpha_\pm(z) = \sum_{k=-\infty, k\neq 0}^{\infty} \frac{e^{ikz}}{\pm ik^\alpha}.$$

The Weyl fractional integral is identical to the Abel fractional integral when the function f is 2π periodic and its integral over a period vanishes, that is with $c_0 = 0$ in the above. As in the Riemann–Liouville case, the Weyl fractional derivative of order α is defined as

$$[^W D^\alpha_\pm f](x) = \pm\frac{d}{dx}[^W I^{1-\alpha}_\pm f](x). \tag{1.22}$$

The Abel, Riemann–Liouville, and Weyl derivatives and integrals are all one-sided as they require a fixed lower endpoint and an upper value to either the left or right of this. The Riesz derivative [**288, 290**] was designed to be more symmetric and suitable for boundary value as opposed to initial value problems. Riesz defined $I^\alpha f$ for $f \in L^1_{\mathrm{Loc}}(\mathbb{R})$ by

$$[I^\alpha f](x) = \frac{1}{2\cos(\alpha\pi/2)}\Big([_W I^\alpha_+ f](x) + [_W I^\alpha_- f]\Big)(x). \tag{1.23}$$

The Riesz derivative is then defined in an analogous way to the Weyl derivative.

Based on the work of Riesz, Hardy and Littlewood investigated the mapping properties of the Riemann–Liouville derivative in Hölder spaces. A function f defined in an open set $\Omega \subset \mathbb{R}^n$ is said to be Hölder continuous of order β, $0 < \beta \le 1$, if $|f(x) - f(y)| \le C\|x - y\|^\beta$. The Hölder space $C^{k,\beta}(\Omega)$ consists of those functions on Ω having continuous derivatives up to order k and such that the kth partial derivatives are Hölder continuous with exponent β. The parameter β allows an interpolation between the integer derivatives, and one might ask, "Is that not what fractional derivatives do and what is the connection?" That such a connection exists is the result of the Hardy–Littlewood theorem which (roughly) states for $f \in C^{k,\beta}(\Omega)$ the αth fractional derivative of f lies in $C^{k,\beta-\alpha}(\Omega)$ [**138**, page 587]. There is a generalisation to functions in the Sobolev spaces $H^{m,p}(\Omega)$—the Hardy–Littlewood–Sobolev theorem.

At this stage the concept of a fractional derivative seemed reasonably complete and precise, but there were still many fractional calculus topics to be explored. Such a derivative had the properties originally laid out as essential: it should be a linear operation and defined for as broad a class of functions as possible while for integer values it should equal the usual integer derivative. The Riemann–Liouville version met this criteria but at a cost. The fractional derivative of a constant was nonzero—in fact it was a singular function. From a purely mathematical perspective this doesn't matter, but for applications it most certainly does. Newton's original motivation for the subject would no longer make sense for such definitions.

There is a simple solution. Simply reverse the orders of fractional integration and integer derivatives in the Riemann–Liouville derivative. The cost here is to narrow the class of applicable functions to which the derivative then applies. Such a version of the Abel derivative was studied extensively beginning in the 1950s by the Armenian mathematician Mkhitar Djrbashian and culminating in his extensive research monograph on the topic published in 1966 [**87**], but only available in Russian. The English translation appeared in 1993 [**81**].

In 1967 Caputo realised that using standard derivatives in modeling many physical systems showed that the decay was exponential irrespective of frequency, and this simply did not correspond to observations. He pioneered the *derivative first* version for the Abel integral as a superior model, and the work of many others followed suit. For this reason the version is often referred to the *Caputo derivative*, although its definition dates back to 1823 and the work of Djrbashian was almost unknown in the west.

The basic function of ordinary differential equations (ODE) is the exponential: the solution to $\frac{dy}{dx} = \lambda y$ is the function $y = e^{\lambda x}$. What is the equivalent for fractional derivatives? The answer turns out to be another special function of entire type: the eponymous function

$$E_{\alpha,\beta}(z) = \sum_{k=0}^{\infty} \frac{z^k}{\Gamma(\alpha k + \beta)}, \quad z \in \mathbb{C}, \tag{1.24}$$

was studied in a series of papers by Mittag-Leffler from 1903–1905 [**250, 251**]. It forms a critical component of all of fractional calculus.

The exponential function is also the key component of the fundamental solution of the heat equation, but replacing the time derivative u_t by one of fractional type, $D_t^\alpha u$ requires another special function of entire type: the Wright function [**354–357**]. This is originally due to the number theorist E. M. Wright in connection with his investigations in the asymptotic theory of partitions, but it has found many, seemingly unconnected, applications within several areas of mathematics.

The notion of fractional derivatives takes on a new complexity when one considers differential operators acting on functions of several variables, for example, the Laplacian $-\triangle$. Certainly, the second power $\triangle^2$ leads to the biharmonic equation $\triangle^2 u = f$ which has wide applicability, but what about fractional powers or, in greater generality, if $-\triangle$ is replaced by a general second order elliptic operator $\mathbb{L}$? This introduces a rich area for applications and, over the last ten years, this has become a vibrant area of research.

1.3. A short overview on inverse problems

Not surprisingly, there are many examples of inverse-type problems in mathematics that go far back in history. We are of course most interested in those that involve differential equations, and one of the most quoted, and from Section 1.2 quite relevant to our specific topic, is the tautochrone problem. The invention of the pendulum clock around 1656 by Christiaan Huygens was inspired by investigations of Galileo Galilei a half century before and resulted in more than a thousandfold increase in accuracy over previous mechanisms. Although a good approximation to a harmonic oscillator, the actual resulting equation is nonlinear and accuracy deteriorates with amplitude of oscillation. Thus further solutions were required. Huygens again proposed that the pendulum be constrained by placing two evolutes on either side, and the question was then what shape they should take to make the pendulum isochronous. His answer, in 1673, was to make the pendulum swing in an arc of a cycloid by choosing the evolute to be an inverted cycloid in shape. Unfortunately, friction against the evolute causes a greater error than that corrected by the cycloidal path.

In the early part of the seventeenth century Kepler published his laws of planetary motion having found them by analyzing the astronomical observations of his mentor Tycho Brahe. They gave remarkable accuracy but were ad hoc and derived by essentially *data fitting* rather than from firm principles. Newton provided the answer: the gravitational attraction between the planets and the sun coming from a central, radial force. The further question was what precise form this force should take? The choice was the inverse square law that allowed the direct derivation of Kepler's equations from this assumption alone. In retrospect this can be rephrased as, "What gravitational force law would give rise to Kepler's equations", and the solution to this inverse problem constituted one of the greatest scientific triumphs of all time.

The solution of Poisson's equation $-\triangle u = f$ for u given a function f is of course one of the most basic problems in differential equations. The *inverse gravimetry problem* of recovering f given the measured values of ∇u on the boundary $\partial\Omega$ was known to Laplace. But the first simple solutions were obtained only after another hundred years by Stokes and some fifty years later by Herglotz. An excellent exposition of this problem can be found in [**160**].

Despite these and later subsequent triumphs that might be classified as an inverse problem, the study of inverse problems for differential equations did not take place until the twentieth century and its systematic study not until the last half. Two early examples are noteworthy.

In 1911 Hermann Weyl [**345**] showed the asymptotic behavior of eigenvalues of the Laplace–Beltrami operator in a domain $\Omega \subset \mathbb{R}^d$: the number, $N(\lambda)$ of the Dirichlet eigenvalues (counting their multiplicities) less than or equal to λ is

$$N(\lambda) = (2\pi)^{-d}\lambda^{d/2}\omega_d \mathrm{vol}(\Omega) \mp \frac{1}{4}(2\pi)^{1-d}\omega_{d-1}\lambda^{(d-1)/2}\mathrm{area}(\partial\Omega) + o(\lambda^{(d-1)/2}),$$

where ω_d is the volume of the unit ball in $\mathbb{R}^d$; see [**345**]. Based on this, one can recover the volume and surface area of Ω from spectral data. This led to questions about whether further geometrical quantities of a region Ω could be determined from such spectral information epitomized by Marc Kac's famous article, "Can you hear the shape of a drum?" [**172**].

Following similar spectral question lines, in the late 1920s Ambartsumian in his study of atomic structure and the energy levels arising from the Schrödinger equation asked the question of whether the potential arising from the force field could be determined from spectral measurements which directly correlate to the eigenvalues of the equation. He in fact gave a solution in a particular case, and the complete solution was provided by Borg in 1946 [**36**]. This showed that the potential $q(x)$ in $-u_{xx}+q(x)u = \lambda_n u$ on a finite interval could be uniquely determined from two complete spectral series $\{\lambda_n\}_{n=1}^{\infty}$ and $\{\mu_n\}_{n=1}^{\infty}$ corresponding to different endpoint boundary conditions. This work was continued in 1951 by Gel'fand and Levitan using a much shorter and ingenious method that subsequently has been much copied and expanded beyond the original scope of the problem; see [**111**]. It also included a constructive algorithm for $q(x)$. This was later improved to the point where the inverse problem of recovering $q(x)$ from its spectral data is solved significantly faster than computing even a few eigenvalues given $q(x)$, and it also expanded to uniqueness and reconstruction algorithms for more general equations $-\big(p(x)u'\big)' + q(x)u = \lambda\rho(x)u$ and boundary conditions [**57, 300**].

Although PDEs have a history dating to at least the middle of the eighteenth century and they were extensively used as models to describe physical processes throughout this period, the formulation and solution methods were less rigorous. The early twentieth century saw a considerable shift, not just in their analysis but in their relationship to physics; the birth of modern physics was accomplished by increased sophistication in mathematics. In all these periods of intense discovery there is a tendency to draw lines, and one famous example is due to Hadamard. Hadamard took the point of view that every mathematical problem corresponding to some physical or technological problem must be *well-posed.* His definition of this was that the mathematical problem had to have a solution, this solution was unique, and further, it had to depend continuously on the data. The third was the most controversial, and the reasoning behind it was that an arbitrary small change in the data should not lead to large changes in the solution. The most explicit versions of these statements were in his 1923 book [**127**] on PDEs, which was extremely influential, but there were similar statements much earlier [**125**, **126**]. Of course several already well-known problems such as the backwards heat equation and the Cauchy problem for Laplace's equation fell into the *ill-posed* category, as it became known. Alternative names were equally pejorative: "incorrectly set" or "ill-conditioned" problems.

The result was that for several decades such classic inverse problems as noted above or the recovery of parameters within the equation were viewed in mathematically negative terms—despite the fact that there were critical applications to the contrary. The rationale for this negativity was based on the instability arising from even small measurement errors and hence a resulting lack of reliability on interpreting the solution.

The 1960s saw the tide turn with a series of methods aimed at quantifying and ameliorating the level of ill-conditioning—*regularisation methods* that sought to temper the effect. We show some of these in Chapter 8 and indeed they are a fundamental tool in almost all inverse problems involving differential equations.

The same decade saw the beginning of a systematic study in the determination of unknown coefficients occurring in partial differential operators from extra information measurements. Frequently, the latter consisted of additional boundary measurements. For example, if the direct problem partial differential equation (PDE) is of elliptic type and subject to a Dirichlet boundary condition, then one might additionally measure Neumann conditions. If the equation is of parabolic type, one might use the values of the solution $u(x, t)$ for a fixed value of x and t ranging over an interval. We will refer to this as "time trace data" throughout the book.

An alternative situation is when part of the boundary is hidden and one wants to recover solution values $u(x)$ there from additional measurements at accessible parts of the boundary. Examples here include the abovementioned Cauchy problem for Laplace's equation and the *backwards heat equation* whereby one is able to measure $u(x, t)$ at a later time $t = T$ and hopes to recover the initial configuration of temperature $u(x, 0)$. These very classic problems were known in their basic form from at least the nineteenth century. They have unique solutions but the dependence of the unknowns on the data is highly unstable.

Another important problem is that of inverse scattering. A common paradigm is to convert the wave equation to frequency domain and obtain the Helmholtz equation $\triangle u + k^2 u = 0$. In the simplest situation, we have an impenetrable scatterer D; a plane wave with direction d is fired at D and the resulting amplitude is measured in all directions on a large sphere (which could have infinite radius) the so-called far-field pattern. The uniqueness question is whether these measurements allow recovery of ∂D. Further questions include allowing a penetrable obstacle and the equation $\triangle u + k^2 n(x) u = 0$, where $n(x)$ is the interior refraction index of D, but additionally providing the far-field pattern arising from a complete set of incident directions. Problems such as these are at the heart of imaging modalities and the determination of material parameters where the only information that can be measured is exterior to the region. The applications are ubiquitous in science and engineering.

In the above models we have classical PDEs of all the three main types and in each case the derivatives are the usual ones of integer order. One of the main purposes of this book is to answer the following questions: What changes if some of the integer order derivatives are replaced by ones of fractional type? Do we we still retain uniqueness? Are there cases where the integer order situation leads to an underdetermined problem but uniqueness is restored under fractional derivatives? Does the degree of ill-conditioning depend on the fractional order α, and if so, to what extent? There is also the additional question of determining the fractional operator itself, for this might depend on several fractional exponents as well as coupling coefficients.

Let us finally point out that for some classical examples, such as the backwards heat equation, it is immediate what a physically meaningful fractional counterpart should be, namely replacing the first order time derivative by a fractional one. This is much less clear for some others, like inverse scattering, where space derivatives in higher dimensions play a central role. Thus we focus on inherently one-dimensional derivative concepts and predominantly time derivatives. These come with a decisive nonlocality and directionality that in some inverse problems strongly influence the degree of its ill-posedness. Going to higher (space) dimensions also requires mathematical tools that are very different from those used here. For these reasons, some of the recent highly productive areas of inverse problems for fractional PDEs, such as those involving fractional version of the Laplacian, are only touched upon in an outlook in the final chapter.