Preface

This book is the first to be devoted entirely to the potential theory of the heat (or diffusion) equation

$$\sum_{i=1}^{n} \frac{\partial^2 u}{\partial x_i^2} = \frac{\partial u}{\partial t}$$

in Euclidean space $\mathbb{R}^{n+1} = \mathbb{R}^n \times \mathbb{R}$. It thus deals with time dependent potential theory. Its purpose is to give a logical, coherent introduction to a subject that has been approached in several conflicting ways.

The subject has had an unusual history. Some of the elementary results have appeared only in the twenty-first century, but some other results were included in an axiomatic theory in the nineteen sixties. I make no reference to the axiomatic theory in the text, but the definition and treatment of subtemperatures in Chapter 3 is designed to reconcile the harmonic spaces definition with the one that I have used in my researches. This approach is very recent, having first appeared in 2008.

Most results in the heat potential theory have been modelled on the classical results for Laplace's equation. However, after a great deal of thought, I decided to make no mention of the classical theory in the text. I wanted to write a book in which heat potential theory stands as a subject in its own right, free from the clutter of perpetual references to the classical case. Many of the proofs in heat potential theory are similar to those in classical potential theory, so if the classical case is covered first, then either the classical proofs have to be duplicated, or the proofs for the case of the heat equation have to be replaced by a claim that the proofs are similar to the classical case. Many times I have seen such claims in print, but in a substantial number of cases the claim has proved to be false. There is no substitute for writing out all the details of a proof. One could, of course, deduce the results of classical potential theory from those of heat potential theory, as the former is the special case of the latter in which nothing depends on time. But I don't think that would be of much interest. Such an approach to the classical case would be unnecessarily tortuous, and there are already some excellent texts that deal solely with classical potential theory.

The reader already familiar with the heat equation, may be surprised to find that the explicit Poisson integral representation of solutions of the heat equation on a rectangular domain, is not mentioned in the text. I have found it unnecessary, and so have been delighted to omit it because of its complication, which may even have deterred some mathematicians from researching on the heat equation. Its place has been taken by a caloric measure interpretation of the representation of solutions on a circular cylinder.

I have reworked the entire content of this book, including not only most of the individual proofs but also the overall approach. Despite this, the only essentially

new material here is that on caloric measure in Chapter 8, a few minor theorems, and a few examples. Otherwise, there is just a great deal of fine-tuning, including a different approach to the Riesz decomposition theorem in Chapter 6. As befits an introduction, I have treated the subject in as elementary a way as I could. I have not attempted a complete coverage, and in particular have made no mention of the probabilistic approach. The prerequisites for understanding all the proofs are a good background in the calculus of functions of several variables, in the limiting processes and inequalities of analysis, in measure theory, and in general topology for Chapter 9. Some general analytic results that are not easy to find elsewhere, or are not necessarily presented in the most suitable form elsewhere, have been included where they are needed.

The material is presented in logical order, which differs from the chronological order in which the results were first discovered. Chapter 1 deals with elementary issues, although the results are more recent than might be expected, and those in Section 1.6 first appeared in 2002. Chapter 2 presents the classical existence theory for temperatures on a circular cylinder, which is much older than the results in Chapter 1. Subtemperatures are introduced in Chapter 3, using a definition based on the representation theorem in Chapter 2. However, heat balls and modified heat balls are essential to our treatment, and Chapter 3 contains some necessary existence theory for temperatures on such, and other, domains. This introduces the PWB approach to the Dirichlet problem without the added complications of arbitrary open sets and arbitrary boundary functions. Chapters 4 and 5 deal with temperatures and subtemperatures, respectively, on domains of the form $\mathbb{R}^n \times \,]0, a[$, where $0 < a \leq +\infty$. Although potentials are used as early as Chapter 2, the general theory of heat potentials begins in earnest in Chapter 6, where Green functions and heat potentials are introduced, and a distributional approach to the Riesz decomposition theorem is taken. Chapter 7 deals with polar sets, reductions, and thermal capacity. In Chapter 8 we consider the generalized Dirichlet problem, where the open sets and boundary functions are arbitrary. This chapter includes a new treatment of caloric measure for such sets. Finally, in Chapter 9 we discuss the thermal fine topology, which gives us an insight into the continuity properties of subtemperatures, and thus enables us to improve upon some earlier results. Each chapter concludes with bibliographical notes and comments, which include mention of matters not covered in the text, and of open questions even in Chapter 1. They do not provide a detailed historical account of the theory, except for the more recent results. They contain very few references to the axiomatic approach, and none to the probabilistic approach, mainly because those approaches have very different starting points and linking them with the main text would take too much space.

It is a pleasure to acknowledge my debt to those who have written earlier books on potential theory. In particular, L. L. Helms' book *Introduction to Potential Theory* [**33**] first inspired me to take up the subject, with the eventual goal of writing a book such as this one. J. L. Doob's *Classical Potential Theory and its Probabilistic Counterpart* [**14**] has been a source of ideas for clever proofs, and for most of the material in Chapter 9, but is definitely not for the novice. D. H. Armitage & S. J. Gardiner's *Classical Potential Theory* [**3**] has provided many ideas for clever proofs, which I have been able to adapt to the case of the heat equation. The material in Chapter 2, on the existence of solutions to the Dirichlet problem on a circular cylinder, is based on the treatment in E. M. Landis' *Second Order Equations of*

Elliptic and Parabolic Type [**49**].

The reader who wants to look at an axiomatic approach to potential theory that includes heat potential theory, could consult H. Bauer's *Harmonische Räume und ihre Potentialtheorie* [**5**], C. Constantinescu & A. Cornea's *Potential Theory on Harmonic Spaces* [**12**], or J. Bliedtner & W. Hansen's *Potential Theory: An Analytic and Probabilistic Approach to Balayage* [**7**].

Notation and Terminology

We summarize here our basic notations and conventions. Most notation will be explained as it is introduced during the course of the book, and is indexed.

We say that a number or function f is *positive* if $f > 0$, *negative* if $f < 0$, *nonnegative* if $f \geq 0$, and *nonpositive* if $f \leq 0$. However, we say that a real function f is *increasing* if $f(a) \leq f(b)$ whenever $a \leq b$, and that it is *decreasing* if $f(a) \geq f(b)$ whenever $a \leq b$. Similarly, for sequences of numbers or functions, the terms *increasing* and *decreasing* are used in the wide sense.

We use $\mathbb{R}^n$ to denote real Euclidean space of dimension n, with $n \geq 1$, but often omit the superscript if $n = 1$. We also denote the set of positive integers by $\mathbb{N}$, and the set of rational numbers by $\mathbb{Q}$. A typical point of $\mathbb{R}^n$ is denoted by $x = (x_1, ..., x_n)$, and we write $|x|$ for the Euclidean norm $\left(\sum_{i=1}^{n} x_i^2\right)^{1/2}$ of x. The inner product $\sum_{i=1}^{n} x_i y_i$ of two points x and y in $\mathbb{R}^n$ is written $\langle x, y \rangle$. Most of the material is presented in the context of $\mathbb{R}^{n+1} = \mathbb{R}^n \times \mathbb{R} = \{(x, t) : x \in \mathbb{R}^n, t \in \mathbb{R}\}$, where the variables $x = (x_1, ..., x_n)$ are called the spatial variables and t is called the temporal variable. Where there is no need to specify these variables separately, we use p or q to denote a typical point of $\mathbb{R}^{n+1}$, reserving x and y for points of $\mathbb{R}^n$. We denote the Euclidean norm of a point p in $\mathbb{R}^{n+1}$ by $|p|$, leaving the notation for a point to distinguish between $|p|$ and $|x|$. Similarly, the open balls of radius r in $\mathbb{R}^n$ and $\mathbb{R}^{n+1}$ are denoted by $B(x, r)$ and $B(p, r)$, respectively; thus $B(x, r) = \{y \in \mathbb{R}^n : |x - y| < r\}$ and $B(p, r) = \{q \in \mathbb{R}^{n+1} : |p - q| < r\}$. The points x and p are called the centres of the respective balls. A unit ball is a ball of radius 1, and a unit sphere is its boundary. The origin of Euclidean space is denoted by 0, regardless of the dimension of the space.

All topological concepts are relative to the Euclidean topology of $\mathbb{R}^{n+1}$, unless otherwise stated. The symbol E denotes an open set in $\mathbb{R}^{n+1}$, which is always assumed to be nonempty. For any set S in $\mathbb{R}^{n+1}$, we denote its boundary by ∂S and its closure by $\overline{S}$, although we denote the closure of a ball by $\overline{B}(p, r)$ rather than $\overline{B(p, r)}$, and similarly for other sets that depend on listed parameters. The boundary of a set is taken with respect to the one-point compactification of $\mathbb{R}^{n+1}$, so that the point at infinity is included if the set is unbounded. The interior of S is denoted by S°. The connected components of a set are referred to simply as its *components*, and a nonempty connected open set is called a *domain*. By a *hyperplane*, we mean a set of the form $\{p \in \mathbb{R}^{n+1} : \langle p, q \rangle = a\}$ for some $q \in \mathbb{R}^{n+1}$ and $a \in \mathbb{R}$. If A and B are two sets, we put $A \backslash B = \{p \in A : p \notin B\}$. A set is called a G_δ set if it can be expressed as a countable intersection of open sets, and an F_σ set if it can be expressed as a countable union of closed sets.

All of our functions are extended real-valued, that is, their values are in $\mathbb{R}$ or are $\pm\infty$. This necessitates a limited arithmetic with $\pm\infty$ when we add or multiply

functions, so we adopt the following conventions, in which $t \in \mathbb{R}$:

$$(\pm\infty) + (\pm\infty) = \pm\infty = t + (\pm\infty) = (\pm\infty) + t,$$

$$(\pm\infty).(\pm\infty) = +\infty, \qquad (\pm\infty).(\mp\infty) = -\infty,$$

$$t.(\pm\infty) = (\pm\infty).t = \begin{cases} \pm\infty & \text{if} \quad t > 0, \\ 0 & \text{if} \quad t = 0, \\ \mp\infty & \text{if} \quad t < 0. \end{cases}$$

Other expressions, such as $(\pm\infty) + (\mp\infty)$, are left undefined. We put $\inf \emptyset = +\infty$ and $\sup \emptyset = -\infty$.

If f is an extended real-valued function defined on a set $S \subseteq \mathbb{R}^{n+1}$, q is a limit point of S in some topology, and $\mathcal{N}_q$ is the collection of neighbourhoods of q in that topology, then we define

$$\liminf_{p \to q,\, p \in S} f(p) = \sup_{N \in \mathcal{N}_q} \left(\inf_{p \in N \cap S \setminus \{q\}} f(p) \right)$$

and

$$\limsup_{p \to q,\, p \in S} f(p) = \inf_{N \in \mathcal{N}_q} \left(\sup_{p \in N \cap S \setminus \{q\}} f(p) \right).$$

We say that $\lim_{p \to q,\, p \in S} f(p)$ exists if $\liminf_{p \to q,\, p \in S} f(p) = \limsup_{p \to q,\, p \in S} f(p)$, and if that common value is l we write $\lim_{p \to q,\, p \in S} f(p) = l$. Here l may be a real number or $\pm\infty$. If S is the domain of definition of f, or if $S \in \mathcal{N}_q$, then we may omit the qualification "$p \in S$". We say that f is continuous at q if f is defined at q and $\lim_{p \to q} f(p) = f(q)$, regardless of whether $f(q) \in \mathbb{R}$.

If u and v are extended real-valued functions defined on the same set, we use $u \vee v$ to denote $\max\{u, v\}$ and $u \wedge v$ to denote $\min\{u, v\}$. We also put $u^+ = u \vee 0$ and $u^- = -(u \wedge 0)$, thus obtaining the identities $u = u^+ - u^-$ and $|u| = u^+ + u^-$. If S is a subset of the domain of definition of u, and $u(p) \leq M$ for all $p \in S$ and some real number M, then we say that u is *upper bounded* on S. Similarly, if $u(p) \geq m$ for all $p \in S$ and some $m \in \mathbb{R}$, then we say that u is *lower bounded* on S. If u is both upper bounded on S and lower bounded on S, we say that u is *bounded* on S. If u is bounded on K for each compact subset K of S, then we say that u is *locally bounded* on S. We define *locally upper bounded* and *locally lower bounded* analogously. If D is the domain of definition of u, we define the *support* of u to be the set $D \setminus \{p \in D : u = 0 \text{ on } D \cap B(p, r) \text{ for some } r > 0\}$.

A family $\mathcal{F}$ of functions defined on a set S is said to be *uniformly bounded* on S if there is a real number M such that $|u(p)| \leq M$ for all $u \in \mathcal{F}$ and all $p \in S$. The family $\mathcal{F}$ is said to be *locally uniformly bounded* on S if it is uniformly bounded on K for each compact subset K of S. We define the phrases *uniformly upper bounded*, *uniformly lower bounded*, *locally uniformly upper bounded*, and *locally uniformly lower bounded*, analogously. A sequence $\{u_j\}$ is said to converge *locally uniformly* on S if it converges uniformly on each compact subset of S.

Let X be a subset of the one-point compactification of $\mathbb{R}^{n+1}$. The class $\mathcal{B}$ of *Borel* subsets of X is the smallest σ-algebra to contain the open subsets of X. We say that an extended real-valued function u on X is *Borel measurable* if the set $\{p \in X : u(p) > a\}$ belongs to $\mathcal{B}$ for every real number a. Continuous functions are Borel measurable. A *nonnegative (Borel) measure* on X is a countably additive set

function μ, defined on a σ-algebra that contains $\mathcal{B}$, taking nonnegative extended-real values, such that $\mu(\emptyset) = 0$ and $\mu(K) < +\infty$ for every compact subset K of X. Such a measure is regular, in the sense that

$$\mu(S) = \inf\{\mu(E) : S \subseteq E, \ E \text{ is open}\} = \sup\{\mu(K) : K \subseteq S, \ K \text{ is compact}\}.$$

The *support* of a nonnegative measure μ is the set of points $p \in X$ such that $\mu(N) > 0$ for every open neighbourhood N of p. It is the smallest closed set F such that $\mu(X \backslash F) = 0$. If S belongs to the σ-algebra upon which μ is defined, we say that S is *μ-measurable*, and define the *restriction* of μ to S by $\mu_S(T) = \mu(T \cap S)$ for all μ-measurable sets T. If $X \subseteq Y$ and $X \neq Y$, we define the restriction of μ to S *as a nonnegative measure on Y* by adding the condition $\mu_S(Y \backslash X) = 0$. We say that an extended real-valued function u on X is *μ-measurable* if the set $\{p \in X : u(p) > a\}$ is μ-measurable for every $a \in \mathbb{R}$. In the case where μ is Lebesgue measure, we omit the prefix μ-. A μ-measurable function u on X is said to be *μ-integrable* on X if $\int_X |u| \, d\mu < +\infty$, and *locally μ-integrable* on X if $\int_K |u| \, d\mu < +\infty$ for every compact subset K of X. The prefix μ- is omitted if μ is Lebesgue measure. When writing integrals with respect to Lebesgue measure, we usually use the traditional notation $\int_X u(p) \, dp$. A relation which holds on a μ-measurable set Y such that $\mu(X \backslash Y) = 0$, is said to hold *μ-almost everywhere* on X, and again the prefix μ- is omitted if μ is Lebesgue measure.

A *signed measure* on X is a countably additive set function ν, defined on a σ-algebra that contains $\mathcal{B}$, taking only real values, such that $\nu(\emptyset) = 0$. (Some relaxation of the finiteness is described in Chapter 4.) In view of the Hahn-Jordan decomposition theorem, there are disjoint ν-measurable sets P and N such that $P \cup N = X$, and nonnegative finite measures ν^+ and ν^- on X, such that for all ν-measurable subsets S of X we have $\nu^+(S) = \nu(S \cap P)$ and $\nu^-(S) = \nu(S \cap N)$. Then ν has the decomposition $\nu = \nu^+ - \nu^-$, and the nonnegative measure $|\nu| = \nu^+ + \nu^-$ is called the *total variation* of ν.