Chapter 1

Introduction: Examples, Background, and Perspectives

1 Orientation

1.1 Geometry as a Variable

The central object of this book[1] is the *geometry* as a variable. As in the theory of functions of real variables, we need a differential calculus, spaces of geometries, evolution equations, and other familiar concepts in analysis when the variable is no longer a scalar, a vector, or a function, but is a geometric domain. This is motivated by many important problems in science and engineering that involve the geometry as a modeling, design, or control variable. In general the geometric objects we shall consider will not be parametrized or structured. Yet we are not starting from scratch, and several building blocks are already available from many fields: geometric measure theory, physics of continuous media, free boundary problems, the parametrization of geometries by functions, the set derivative as the inverse of the integral, the parametrization of functions by geometries, the Pompéiu–Hausdorff metric, and so on.

As is often the case in mathematics, spaces of geometries and notions of derivatives with respect to the geometry are built from well-established elements of functional analysis and differential calculus. There are many ways to structure families of geometries. For instance, a domain can be made variable by considering

[1] The *numbering* of equations, theorems, lemmas, corollaries, definitions, examples, and remarks is by chapter. When a reference to another chapter is necessary it is always followed by the words *in Chapter* and the *number of the chapter*. For instance, "equation (2.18) in Chapter 9." The text of theorems, lemmas, and corollaries is slanted; the text of definitions, examples, and remarks is normal shape and ended by a square □. This makes it possible to aesthetically emphasize certain words especially in definitions. The bibliography is by author in alphabetical order. For each author or group of coauthors, there is a numbering in square brackets starting with [**1**]. A reference to an item by a single author is of the form J. Dieudonné [**3**] and a reference to an item with several coauthors S. Agmon, A. Douglis, and L. Nirenberg [**2**]. *Boxed formulae* or *statements* are used in some chapters for two distinct purposes. First, they emphasize certain important definitions, results, or identities; second, in long proofs of some theorems, lemmas, or corollaries, they isolate key intermediary results which will be necessary to more easily follow the subsequent steps of the proof.

the images of a fixed domain by a family of diffeomorphisms that belong to some function space over a fixed domain. This naturally occurs in physics and mechanics, where the deformations of a continuous body or medium are smooth, or in the numerical analysis of optimal design problems when working on a fixed grid. This construction naturally leads to a group structure induced by the composition of the diffeomorphisms. The underlying spaces are no longer topological vector spaces but groups that can be endowed with a nice complete metric space structure by introducing the *Courant metric*. The practitioner might or might not want to use the underlying mathematical structure associated with his or her constructions, but it is there and it contains information that might guide the theory and influence the choice of the numerical methods used in the solution of the problem at hand.

The parametrization of a fixed domain by a fixed family of diffeomorphisms obviously limits the family of variable domains. The topology of the images is similar to the topology of the fixed domain. Singularities that were not already present there cannot be created in the images. Other constructions make it possible to considerably enlarge the family of variable geometries and possibly open the doors to pathological geometries that are no longer open sets with a nice boundary. Instead of parametrizing the domains by functions or diffeomorphisms, certain families of functions can be parametrized by sets. A single function completely specifies a set or at least an equivalence class of sets. This includes the distance functions and the characteristic function, but also the support function from convex analysis. Perhaps the best known example of that construction is the Pompéiu–Hausdorff metric topology. This is a very weak topology that does not preserve the volume of a set. When the volume, the perimeter, or the curvatures are important, such functions must be able to yield relaxed definitions of volume, perimeter, or curvatures. The characteristic function that preserves the volume has many applications. It played a fundamental role in the integration theory of Henri Lebesgue at the beginning of the 20th century. It was also used in the 1950s by E. De Giorgi to define a relaxed notion of perimeter in the *theory of minimal surfaces*.

Another technique that has been used successfully in *free* or *moving boundary problems*, such as motion by mean curvature, shock waves, or detonation theory, is the use of level sets of a function to describe a free or moving boundary. Such functions are often the solution of a system of partial differential equations. This is another way to build new tools from functional analysis. The choice of families of function parametrized sets or of families of set parametrized functions, or other appropriate constructions, is obviously problem dependent, much like the choice of function spaces of solutions in the theory of partial differential equations or optimization problems. This is one aspect of the geometry as a variable. Another aspect is to build the equivalent of a differential calculus and the computational and analytical tools that are essential in the characterization and computation of geometries. Again, we are not starting from scratch and many building blocks are already available, but many questions and issues remain open.

This book aims at covering a small but fundamental part of that program. We had to make difficult choices and refer the reader to appropriate books and references for *background material* such as geometric measure theory and *specialized topics* such as homogenization theory and microstructures which are available in excellent books in English. It was unfortunately not possible to include references to the considerable literature on numerical methods, free and moving boundary problems, and optimization.

1.2 Outline of the Introductory Chapter

We first give a series of generic examples where the shape or the geometry is the modeling, control, or optimization variable. They will be used in the subsequent chapters to illustrate the many ways such problems can be formulated. The first example is the celebrated problem of the optimal shape of a column formulated by Lagrange in 1770 to prevent buckling. The extremization of the eigenvalues has also received considerable attention in the engineering literature. The free interface between two regions with different physical or mechanical properties is another generic problem that can lead in some cases to a *mixing* or a *microstructure*. Two typical problems arising from applications to condition the thermal environment of satellites are described in sections 7 and 8. The first one is the design of a thermal diffuser of minimal weight subject to an inequality constraint on the output thermal power flux. The second one is the design of a thermal radiator to effectively radiate large amounts of thermal power to space. The geometry is a volume of revolution around an axis that is completely specified by its height and the function which specifies its lateral boundary. Finally, we give a glimpse at *image segmentation*, which is an example of shape/geometric identification problems. Many chapters of this book are of direct interest to *imaging sciences*.

Section 10 presents some background and perspectives. A fundamental issue is to find *tractable* and preferably analytical representations of a geometry as a variable that are compatible with the problems at hand. The generic examples suggest two types of representations: the ones where the *geometry is parametrized by functions* and the ones where a family of *functions is parametrized by the geometry.* As is always the case, the choice is very much problem dependent. In the first case, the topology of the variable sets is fixed; in the second case the families of sets are much larger and topological changes are included. The book presents the two points of view. Finally, section 11 sketches the material in the second edition of the book.

2 A Simple One-Dimensional Example

A general feature of minimization problems with respect to a shape or a geometry subject to a state equation constraint is that they are generally not convex and that, when they have a solution, it is generally not unique. This is illustrated in the following simple example from J. CÉA [**2**]: minimize the *objective function*

$$J(a) \stackrel{\text{def}}{=} \int_0^a |y_a(x) - 1|^2 \, dx,$$

where $a \geq 0$ and y_a is the solution of the boundary value problem (*state equation*)

$$\frac{d^2 y_a}{dx^2}(x) = -2 \text{ in } \Omega_a \stackrel{\text{def}}{=} (0, a), \quad \frac{dy_a}{dx}(0) = 0, \quad y_a(a) = 0. \tag{2.1}$$

Here the one-dimensional geometric domain $\Omega_a =]0, a[$ is the minimizing variable. We recognize the classical structure of a control problem, except that the minimizing variable is no longer under the integral sign but in the limits of the integral sign. One consequence of this difference is that even the simplest problems will usually not be convex or convexifiables. They will require a special analysis.

In this example it is easy to check that the solution of the state equation is

$$y_a(x) = a^2 - x^2 \quad \text{and} \quad J(a) = \frac{8}{15}a^5 - \frac{4}{3}a^3 + a.$$

The graph of J, shown in Figure 1.1, is not the graph of a convex function. Its global minimum in $a_0 = 0$, local maximum in a_1, and local minimum in a_2,

$$a_1 \stackrel{\text{def}}{=} \sqrt{\frac{3}{4}\left(1 - \frac{1}{\sqrt{3}}\right)}, \quad a_2 \stackrel{\text{def}}{=} \sqrt{\frac{3}{4}\left(1 + \frac{1}{\sqrt{3}}\right)},$$

are all different.

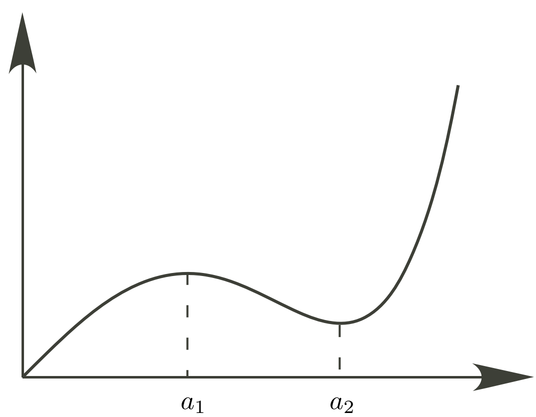

Figure 1.1. *Graph of $J(a)$.*

To avoid a trivial solution, a strictly positive lower bound must be put on a. A unique minimizing solution is obtained for $a \geq a_1$ where the gradient of J is zero. For $0 < a < a_2$, the minimum will occur at the preset lower or upper bound on a.

3 Buckling of Columns

The next example illustrates the fact that even simple problems can be nondifferentiable with respect to the geometry. This is generic of all eigenvalue problems when the eigenvalue is not simple.

One of the early optimal design problems was formulated by J. L. Lagrange [**1**] in 1770 (cf. I. Todhunter and K. Pearson [**1**]) and later studied by the Danish mathematician and astronomer T. Clausen [**1**] in 1849. It consists in finding the best profile of a vertical column of fixed volume to prevent buckling.

It turns out that this problem is in fact a hidden maximization of an eigenvalue. Many incorrect solutions had been published until 1992. This problem and other problems related to columns have been revisited in a series of papers by S. J. Cox [**1**], S. J. Cox and M. L. Overton [**1**], S. J. Cox [**2**], and S. J. Cox and C. M. McCarthy [**1**]. Since Lagrange many authors have proposed solutions, but a complete theoretical and numerical solution for the buckling of a column was given only in 1992 by S. J. Cox and M. L. Overton [**1**]. The difficulty was that the eigenvalue is not simple and hence not differentiable with respect to the geometry.

Consider a normalized column of unit height and unit volume (see Figure 1.2). Denote by ℓ the *magnitude of the normalized axial load* and by u the resulting transverse displacement. Assume that the potential energy is the sum of the bending and elongation energies

$$\int_0^1 EI\,|u''|^2\,dx - \ell \int_0^1 |u'|^2\,dx,$$

where I is the second moment of area of the column's cross section and E is its Young's modulus. For sufficiently small load ℓ the minimum of this potential energy with respect to all admissible u is zero. *Euler's buckling load* λ of the column is the largest ℓ for which this minimum is zero. This is equivalent to finding the following minimum:

$$\lambda \stackrel{\text{def}}{=} \inf_{0\neq u\in V} \frac{\int_0^1 EI\,|u''|^2\,dx}{\int_0^1 |u'|^2\,dx}, \tag{3.1}$$

where $V = H_0^2(0,1)$ corresponds to the clamped case, but other types of boundary conditions can be contemplated. This is an eigenvalue problem with a special Rayleigh quotient.

Assume that E is constant and that the second moment of area $I(x)$ of the column's cross section at the height x, $0 \le x \le 1$, is equal to a constant c times its

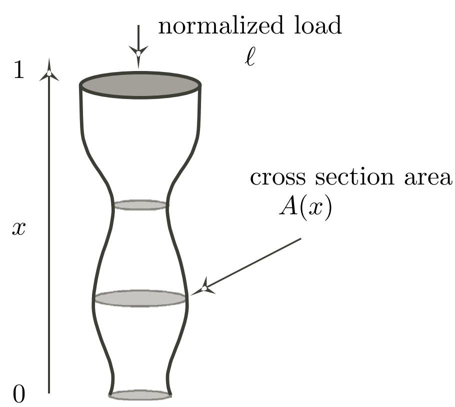

Figure 1.2. *Column of height one and cross section area A under the load ℓ.*

cross-sectional area $A(x)$,

$$I(x) = c\,A(x) \quad \text{and} \quad \int_0^1 A(x)\,dx = 1.$$

Normalizing λ by cE and taking into account the engineering constraints

$$\exists 0 < A_0 < A_1,\ \forall x \in [0,1], \quad 0 < A_0 \le A(x) \le A_1,$$

we finally get

$$\sup_{A \in \mathcal{A}} \lambda(A), \quad \lambda(A) \stackrel{\text{def}}{=} \inf_{0 \ne u \in V} \frac{\int_0^1 A\,|u''|^2\,dx}{\int_0^1 |u'|^2\,dx}, \tag{3.2}$$

$$\mathcal{A} \stackrel{\text{def}}{=} \left\{ A \in L^2(0,1) \,:\, A_0 \le A \le A_1 \text{ and} \int_0^1 A(x)\,dx = 1 \right\}. \tag{3.3}$$

4 Eigenvalue Problems

Let D be a bounded open Lipschitzian domain in $\mathbf{R}^{\mathrm{N}}$ and $A \in L^\infty(D; \mathcal{L}(\mathbf{R}^{\mathrm{N}}, \mathbf{R}^{\mathrm{N}}))$ be a matrix function defined on D such that

$$^*\!A = A \quad \text{and} \quad \alpha I \le A \le \beta I \tag{4.1}$$

for some coercivity and continuity constants $0 < \alpha \le \beta$ and $^*\!A$ is the transpose of A. Consider the minimization or the maximization of the first eigenvalue

$$\left.\begin{matrix} \sup\limits_{\Omega \in \mathcal{A}(D)} \lambda^A(\Omega) \\ \inf\limits_{\Omega \in \mathcal{A}(D)} \lambda^A(\Omega) \end{matrix}\right| \quad \lambda^A(\Omega) \stackrel{\text{def}}{=} \inf_{0 \ne \varphi \in H_0^1(\Omega)} \frac{\int_\Omega A\nabla\varphi \cdot \nabla\varphi\,dx}{\int_\Omega |\varphi|^2\,dx}, \tag{4.2}$$

where $\mathcal{A}(D)$ is a family of admissible open subsets of D (cf., for instance, sections 2, 7, and 9 of Chapter 8).

In the vectorial case, consider the following *linear elasticity* problem: find $U \in H_0^1(\Omega)^3$ such that

$$\forall W \in H_0^1(\Omega)^3, \quad \int_\Omega C\varepsilon(U) \cdot\cdot\, \varepsilon(W)\,dx = \int_\Omega F \cdot W\,dx \tag{4.3}$$

for some distributed loading $F \in L^2(\Omega)^3$ and a *constitutive law* C which is a bilinear symmetric transformation of

$$\mathrm{Sym}_3 \stackrel{\text{def}}{=} \left\{ \tau \in \mathcal{L}(\mathbf{R}^3; \mathbf{R}^3) \colon {}^*\tau = \tau \right\}, \quad \sigma \cdot\cdot\, \tau \stackrel{\text{def}}{=} \sum_{1 \le i,j \le 3} \sigma_{ij}\,\tau_{ij}$$

($\mathcal{L}(\mathbf{R}^3; \mathbf{R}^3)$ is the space of all linear transformations of $\mathbf{R}^3$ or 3×3-matrices) under the following assumption.

Assumption 4.1.
The *constitutive law* is a transformation $C \in \mathrm{Sym}_3$ for which there exists a constant $\alpha > 0$ such that $C\tau \cdot\cdot\, \tau \ge \alpha\, \tau \cdot\cdot\, \tau$ for all $\tau \in \mathrm{Sym}_3$. □

For instance, for the Lamé constants $\mu > 0$ and $\lambda \geq 0$, the special *constitutive law* $C\tau = 2\mu\,\tau + \lambda\,\mathrm{tr}\,\tau\, I$ satisfies Assumption 4.1 with $\alpha = 2\mu$.

The associated bilinear form is

$$a_\Omega(U, W) \stackrel{\text{def}}{=} \int_\Omega C\varepsilon(U) \cdot\cdot\, \varepsilon(W)\, dx,$$

where U is a vector function, $D(U)$ is the Jacobian matrix of U, and

$$\varepsilon(U) \stackrel{\text{def}}{=} \frac{1}{2}\left(D(U) + {}^*D(U)\right)$$

is the *strain tensor*. The first eigenvalue is the minimum of the *Rayleigh quotient*

$$\lambda(\Omega) = \inf\left\{\frac{a_\Omega(U, U)}{\int_\Omega |U|^2\, dx} : \forall U \in H_0^1(\Omega)^3,\, U \neq 0\right\}.$$

A typical problem is to find the sensitivity of the first eigenvalue with respect to the shape of the domain Ω. In 1907, J. Hadamard [**1**] used displacements along the normal to the boundary Γ of a C^∞-domain to compute the derivative of the first eigenvalue of the clamped plate. As in the case of the column, this problem is not differentiable with respect to the geometry when the eigenvalue is not simple.

5 Optimal Triangular Meshing

The *shape calculus* that will be developed in Chapters 9 and 10 for problems governed by partial differential equations (the *continuous model*) will be readily applicable to their *discrete model* as in the finite element discretization of elliptic boundary value problems. However, some care has to be exerted in the choice of the formula for the gradient, since the solution of a finite element discretization problem is usually less smooth than the solution of its continuous counterpart.

Most shape objective functionals will have two basic formulas for their shape gradient: a *boundary expression* and a *volume expression*. The boundary expression is always nicer and more compact but can be applied only when the solution of the underlying partial differential equation is smooth and in most cases *smoother* than the finite element solution. This leads to serious computational errors. The right formula to use is the less attractive volume expression that requires only the same smoothness as the finite element solution. Numerous computational experiments confirm that fact (cf., for instance, E. J. Haug and J. S. Arora [**1**] or E. J. Haug, K. K. Choi, and V. Komkov [**1**]). With the volume expression, the gradient of the objective function with respect to internal and boundary nodes can be readily obtained by plugging in the right velocity field.

A large class of linear elliptic boundary value problems can be expressed as the minimum of a quadratic function over some Hilbert space. For instance, let Ω be a bounded open domain in $\mathbf{R}^N$ with a smooth boundary Γ. The solution u of the boundary value problem

$$-\Delta u = f \text{ in } \Omega, \quad u = 0 \text{ on } \Gamma$$

is the minimizing element in the Sobolev space $H_0^1(\Omega)$ of the energy functional

$$E(v,\Omega) \stackrel{\text{def}}{=} \int_\Omega |\nabla v|^2 - 2f\,v\,dx,$$

$$J(\Omega) \stackrel{\text{def}}{=} \inf_{v\in H_0^1(\Omega)} E(v,\Omega) = E(u,\Omega) = -\int_\Omega |\nabla u|^2\,dx.$$

The elements of this problem are a Hilbert space V, a continuous symmetrical coercive bilinear form on V, and a continuous linear form ℓ on V. With this notation

$$\exists u\in V, \quad E(u) = \inf_{v\in V} E(v), \quad E(v) \stackrel{\text{def}}{=} a(v,v) - 2\,\ell(v)$$

and u is the unique solution of the variational equation

$$\exists u \in U, \forall v\in V, \quad a(u,v) = \ell(v).$$

In the finite element approximation of the solution u, a finite-dimensional subspace V_h of V is used for some small mesh parameter h. The solution of the approximate problem is given by

$$\exists u_h\in V_h, \quad E(u_h) = \inf_{v_h\in V_h} E(v_h), \quad \Rightarrow \exists u_h\in U_h, \forall v_h\in V_h, \quad a(u_h,v_h) = \ell(v_h).$$

It is easy to show that the error can be expressed as follows:

$$a(u-u_h, u-u_h) = \|u-u_h\|_V^2 = 2\left[E(u_h) - E(u)\right].$$

Assume that Ω is a polygonal domain in $\mathbf{R}^N$. In the finite element method, the domain is partitioned into a set τ_h of small triangles by introducing nodes in $\bar\Omega$

$$\begin{matrix} M \stackrel{\text{def}}{=} \{M_i\in\Omega : 1\le i\le p\} \\ \partial M \stackrel{\text{def}}{=} \{M_i\in\partial\Omega : p+1\le i\le p+q\}\end{matrix} \quad \text{and} \quad \overline{M} \stackrel{\text{def}}{=} M\cup\partial M$$

for some integers $p\ge N+1$ and $q\ge 1$ (see Figure 1.3). Therefore the triangularization $\tau_h = \tau_h(\overline{M})$, the solution space $V_h = V_h(\overline{M})$, and the solution $u_h = u_h(\overline{M})$ are functions of the positions of the nodes of the set $\overline{M}$. Assuming that the total

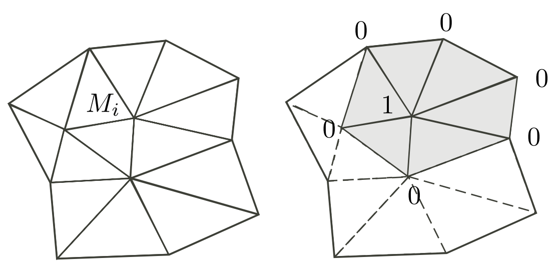

Figure 1.3. *Triangulation and basis function associated with node M_i.*

number of nodes is fixed, consider the following optimal triangularization problem:

$$\inf_{\overline{M}} j(\overline{M}), \quad j(\overline{M}) \stackrel{\text{def}}{=} E(u_h(\tau_h(\overline{M})), \Omega) = \inf_{v_h \in V_h(\overline{M})} E(v_h, \Omega),$$

$$\|u - u_h\|_V^2 = \int_\Omega |\nabla(u - u_h)|^2\, dx = 2\, \left[E(u_h, \Omega(\tau_h(\overline{M}))) - E(u, \Omega)\right]$$
$$= 2\, \left[J(\Omega(\tau_h(\overline{M}))) - J(\Omega)\right],$$

$$J(\Omega(\tau_h(\overline{M}))) \stackrel{\text{def}}{=} \inf_{v_h \in V_h} E(v_h, \Omega(\tau_h(\overline{M}))) = E(u_h, \Omega(\tau_h(\overline{M}))) = -\int_\Omega |\nabla u_h|^2\, dx.$$

The objective is to compute the partial derivative of $j(\overline{M})$ with respect to the ℓth component $(M_i)^\ell$ of the node M_i:

$$\frac{\partial j}{\partial (M_i)^\ell}\, (\overline{M}).$$

This partial derivative can be computed by using the velocity method for the special velocity field (cf. M. C. Delfour, G. Payre, and J.-P. Zolésio [**3**])

$$V_{i\ell}(x) = b_{M_i}(x)\, \vec{e}_\ell,$$

where $b_{M_i} \in V_h$ is the (piecewise P^1) basis function associated with the node M_i: $b_{M_i}(M_j) = \delta_{ij}$ for all i, j. In that method each point X of the plane is moved according to the solution of the vector differential equation

$$\frac{dx}{dt}(t) = V(x(t)), \quad x(0) = X.$$

This yields a transformation $X \mapsto T_t(X) \stackrel{\text{def}}{=} x(t; X) : \mathbf{R}^2 \to \mathbf{R}^2$ of the plane, and it is natural to introduce the following notion of semiderivative:

$$dJ(\Omega; V) \stackrel{\text{def}}{=} \lim_{t \searrow 0} \frac{J(T_t(\Omega)) - J(\Omega)}{t}.$$

For $t \geq 0$ small, the velocity field must be chosen in such a way that triangles are moved onto triangles and the point M_i is moved in the direction $\vec{e}_\ell$:

$$M_i \to M_{it} = M_i + t\, \vec{e}_\ell.$$

This is achieved by choosing the following velocity field:

$$V_{i\ell}(t, x) = b_{M_{it}}(x)\, \vec{e}_\ell,$$

where $b_{M_{it}}$ is the piecewise P^1 basis function associated with node M_{it}: $b_{M_{it}}(M_j) = \delta_{ij}$ for all i, j. This yields the family of transformations

$$T_t(x) = x + t\, b_{M_i}(x)\, \vec{e}_\ell$$

which moves the node M_i to $M_i + t\, \vec{e}_\ell$ and hence

$$\frac{\partial j}{\partial (M_i)^\ell}(\overline{M}) = dJ(\Omega; V_{i\ell}).$$

Going back to our original example, introduce the shape functional

$$J(\Omega) \stackrel{\text{def}}{=} \inf_{v\in H_0^1(\Omega)} E(\Omega, v) = -\int_\Omega |\nabla u|^2\, dx, \quad E(\Omega, v) = \int_\Omega |\nabla v|^2 - 2\, f\, v\, dx.$$

In Chapter 9, we shall show that we have the following boundary and volume expressions for the derivative of $J(\Omega)$:

$$dJ(\Omega; V) = -\int_\Gamma \left|\frac{\partial u}{\partial n}\right|^2 V\cdot n\, d\Gamma,$$

$$dJ(\Omega; V) = \int_\Omega A'(0)\nabla u\cdot\nabla u - 2\,[\operatorname{div} V(0) f + \nabla f\cdot V(0)]\, u\, dx,$$

$$A'(0) = \operatorname{div} V(0)\, I - {}^*DV(0) - DV(0).$$

For a P^1-approximation

$$V_h \stackrel{\text{def}}{=} \left\{ v\in C^0(\bar\Omega)\, :\, v|_K \in P^1(K),\, \forall K\in\tau_h\right\}$$

and the trace of the normal derivative on Γ is not defined. Thus, it is necessary to use the volume expression. For the velocity field $V_{i\ell}$

$$DV_{i\ell} = \vec e_\ell\, {}^*\nabla b_{M_i}, \quad \operatorname{div} DV_{i\ell} = \vec e_\ell\cdot\nabla b_{M_i},$$

$$A'(0) = \vec e_\ell\cdot\nabla b_{M_i}\, I - \vec e_\ell\, {}^*\nabla b_{M_i} - \nabla b_{M_i}\, {}^*\vec e_\ell.$$

Since

$$\frac{\partial j}{\partial (M_i)^\ell}\,(\overline{M}) = dJ(\Omega; V_{i\ell}),$$

we finally obtain the formula for the derivative of the function $j(\overline{M})$ with respect to node M_i in the direction $\vec e_\ell$:

$$\begin{aligned}\frac{\partial j}{\partial (M_i)^\ell}\,(\overline{M}) = \int_\Omega & [\vec e_\ell\cdot\nabla b_{M_i}\, I - \vec e_\ell\, {}^*\nabla b_{M_i} - \nabla b_{M_i}\, {}^*\vec e_\ell]\, \nabla; u_h\cdot\nabla u_h \\ & - 2\,[\vec e_\ell\cdot\nabla b_{M_i}\, f + \nabla f\cdot\vec e_\ell\, b_{M_i}]\, u_h\, dx.\end{aligned}$$

Since the support of b_{M_i} consists of the triangles having M_i as a vertex, the gradient with respect to the nodes can be constructed piece by piece by visiting each node.

6 Modeling Free Boundary Problems

The first step towards the solution of a shape optimization is the mathematical modeling of the problem. Physical phenomena are often modeled on relatively smooth or nice geometries. Adding an objective functional to the model will usually push the system towards rougher geometries or even microstructures. For instance, in the optimal design of plates the optimization of the profile of a plate led to highly oscillating profiles that looked like a comb with abrupt variations ranging from zero to maximum thickness. The phenomenon began to be understood in 1975 with the paper of N. Olhoff [**1**] for circular plates with the introduction of the mechanical notion of *stiffeners*. The optimal plate was a *virtual* plate, a microstructure, that is a homogenized geometry. Another example is the Plateau problem of minimal surfaces that experimentally exhibits surfaces with singularities. In both cases, it is mathematically natural to replace the geometry by a characteristic function, a function that is equal to 1 on the set and 0 outside the set. Instead of optimizing over a restricted family of geometries, the problem is relaxed to the optimization over a set of measurable characteristic functions that contains a much larger family of geometries, including the ones with boundary singularities and/or an arbitrary number of holes.

6.1 Free Interface between Two Materials

Consider the optimal design problem studied by J. Céa and K. Malanowski [**1**] in 1970, where the optimization variable is the distribution of two materials with different physical characteristics within a fixed domain D. It cannot a priori be assumed that the two regions are separated by a smooth interface and that each region is connected. This problem will be covered in more details in section 4 of Chapter 5.

Let $D \subset \mathbf{R}^N$ be a bounded open domain with Lipschitzian boundary ∂D. Assume for the moment that the domain D is partitioned into two subdomains Ω_1 and Ω_2 separated by a smooth interface $\partial\Omega_1 \cap \partial\Omega_2$ as illustrated in Figure 1.4. Domain Ω_1 (resp., Ω_2) is made up of a material characterized by a constant $k_1 > 0$ (resp., $k_2 > 0$). Let y be the solution of the *transmission problem*

$$\begin{cases} -k_1\triangle y = f \text{ in } \Omega_1 \quad \text{and} \quad -k_2\triangle y = f \text{ in } \Omega_2, \\ y = 0 \text{ on } \partial D \quad \text{and} \quad k_1\dfrac{\partial y}{\partial n_1} + k_2\dfrac{\partial y}{\partial n_2} = 0 \text{ on } \overline{\Omega}_1 \cap \overline{\Omega}_2, \end{cases} \tag{6.1}$$

where n_1 (resp., n_2) is the unit outward normal to Ω_1 (resp., Ω_2) and f is a given function in $L^2(D)$. Assume that $k_1 > k_2$. The objective is to maximize the equivalent of the *compliance*

$$J(\Omega_1) = -\int_D f y \, dx \tag{6.2}$$

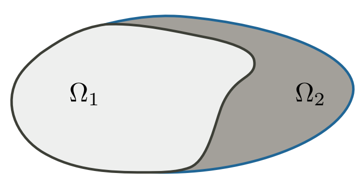

Figure 1.4. *Fixed domain D and its partition into Ω_1 and Ω_2.*

over all domains Ω_1 in D subject to the following constraint on the volume of material k_1 which occupies the part Ω_1 of D:

$$\boxed{\mathrm{m}(\Omega_1) \le \alpha, \quad 0 < \alpha < \mathrm{m}(D)} \tag{6.3}$$

for some constant α.

If χ denotes the characteristic function of the domain Ω_1,

$$\chi(x) = 1 \text{ if } x \in \Omega_1 \text{ and } 0 \text{ if } x \notin \Omega_1,$$

the compliance $J(\chi) = J(\Omega_1)$ can be expressed as the infimum over the Sobolev space $H_0^1(D)$ of an energy functional defined on the fixed set D:

$$J(\chi) = \min_{\varphi \in H_0^1(D)} E(\chi, \varphi), \tag{6.4}$$

$$E(\chi, \varphi) \overset{\text{def}}{=} \int_D (k_1\, \chi + k_2\, (1 - \chi))\, |\nabla \varphi|^2 - 2\, \chi f \varphi\, dx. \tag{6.5}$$

$J(\chi)$ can be minimized or maximized over some appropriate family of characteristic functions or with respect to their relaxation to functions between 0 and 1 that would correspond to microstructures. As in the eigenvalue problem, the objective function is an infimum, but here the infimum is over a space that does not depend on the function χ that specifies the geometric domain. This will be handled by the special techniques of Chapter 10 for the differentiation of the minimum of a functional.

6.2 Minimal Surfaces

The celebrated Plateau's problem, named after the Belgian physicist and professor J. A. F. Plateau [**1**] (1801–1883), who did experimental observations on the geometry of soap films around 1873, also provides a nice example where the geometry is a variable. It consists in finding the surface of least area among those bounded by a given curve. One of the difficulties in studying the *minimal surface problem* is the description of such surfaces in the usual language of differential geometry. For instance, the set of possible singularities is not known.

Measure theoretic methods such as k-currents (k-dim surfaces) were used by E. R. Reifenberg [**1**, **2**, **3**, **4**] around 1960, H. Federer and W. H. Fleming [**1**] in 1960 (normals and integral currents), F. J. Almgren, Jr. [**1**] in 1965 (varifolds), and H. Federer [**5**] in 1969.

In the early 1950s, E. De Giorgi [**1**, **2**, **3**] and R. Caccioppoli [**1**] considered a hypersurface in the N-dimensional Euclidean space $\mathbf{R}^{\mathrm{N}}$ as the boundary of a set. In order to obtain a *boundary measure*, they restricted their attention to sets whose characteristic function is of bounded variation. Their key property is an associated natural notion of *perimeter* that extends the classical surface measure of the boundary of a smooth set to the larger family of *Caccioppoli sets* named after the celebrated Neapolitan mathematician Renato Caccioppoli.[2]

[2] In 1992 his tormented personality was remembered in a film directed by Mario Martone, *The Death of a Neapolitan Mathematician* (Morte di un matematico napoletano).

Caccioppoli sets occur in many shape optimization problems (or free boundary problems), where a surface tension is present on the (free) boundary, such as in the free interface water/soil in a dam (C. Baiocchi, V. Comincioli, E. Magenes, and G. A. Pozzi [**1**]) in 1973 and in the free boundary of a water wave (M. Souli and J.-P. Zolésio [**1**, **2**, **3**, **4**, **5**]) in 1988. More details will be given in Chapter 5.

7 Design of a Thermal Diffuser

Shape optimization problems are everywhere in engineering, physics, and medicine. We choose two illustrative examples that were proposed by the Canadian Space Program in the 1980s. The first one is the design of a thermal diffuser to condition the thermal environment of electronic devices in communication satellites; the second one is the design of a thermal radiator that will be described in the next section. There are more and more design and control problems coming from medicine. For instance, the design of endoprotheses such as valves, stents, and coils in blood vessels or left ventricular assistance devices (cardiac pumps) in interventional cardiology helps to improve the health of patients and minimize the consequences and costs of therapeutical interventions by going to mini-invasive procedures.

7.1 Description of the Physical Problem

This problem arises in connection with the use of high-power solid-state devices (HPSSD) in communication satellites (cf. M. C. Delfour, G. Payre, and J.-P. Zolésio [**1**]). An HPSSD dissipates a large amount of thermal power (typ. >50 W) over a relatively small mounting surface (typ. 1.25 cm^2). Yet, its junction temperature is required to be kept moderately low (typ. 110°C). The thermal resistance from the junction to the mounting surface is known for any particular HPSSD (typ. 1°C/W), so that the mounting surface is required to be kept at a lower temperature than the junction (typ. 60°C). In a space application the thermal power must ultimately be dissipated to the environment by the mechanism of radiation. However, to radiate large amounts of thermal power at moderately low temperatures, correspondingly large radiating areas are required. Thus we have the requirement to efficiently spread the high thermal power flux (TPF) at the HPSSD source (typ. 40 W/cm^2) to a low TPF at the radiator (typ. 0.04 W/cm^2) so that the source temperature is maintained at an acceptably low level (typ. <60°C) at the mounting surface. The efficient spreading task is best accomplished using heatpipes, but the snag in the scheme is that heatpipes can accept only a limited maximum TPF from a source (typ. max 4 W/cm^2).

Hence we are led to the requirement for a thermal diffuser. This device is inserted between the HPSSD and the heatpipes and reduces the TPF at the source (typ. >40 W/cm^2) to a level acceptable to the heatpipes (typ. >4 W/cm^2). The heatpipes then sufficiently spread the heat over large space radiators, reducing the TPF from a level at the diffuser (typ. 4 W/cm^2) to that at the radiator (typ. 0.04 W/cm^2). This scheme of heat spreading is depicted in Figure 1.5.

It is the design of the thermal diffuser which is the problem at hand. We may assume that the HPSSD presents a uniform thermal power flux to the diffuser

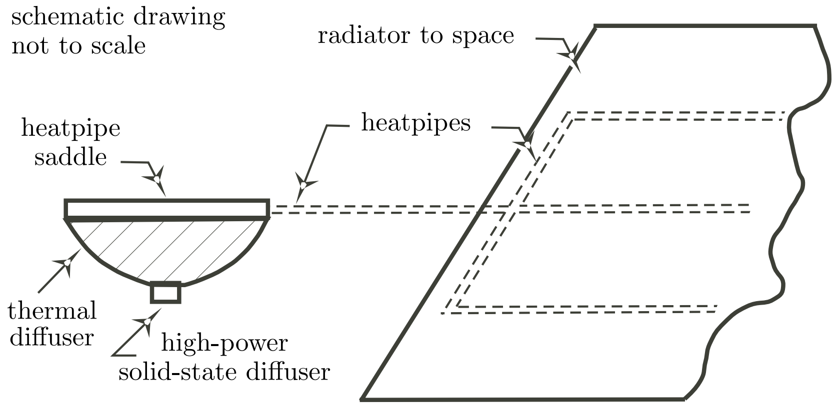

Figure 1.5. *Heat spreading scheme for high-power solid-state devices.*

at the HPSSD/diffuser interface. Heatpipes are essentially isothermalizing devices, and we may assume that the diffuser/heatpipe saddle interface is indeed isothermal. Any other surfaces of the diffuser may be treated as adiabatic.

7.2 Statement of the Problem

Assume that the thermal diffuser is a volume Ω symmetrical about the z-axis (cf. Figure 1.6 (A)) whose boundary surface is made up of three regular pieces: the mounting surface Σ_1 (a disk perpendicular to the z-axis with center in $(r, z) = (0, 0)$), the lateral adiabatic surface Σ_2, and the interface Σ_3 between the diffuser and the heatpipe saddle (a disk perpendicular to the z-axis with center in $(r, z) = (0, L)$).

The temperature distribution over this volume Ω is the solution of the stationary heat equation $k\Delta T = 0$ (ΔT, the Laplacian of T) with the following boundary conditions on the surface $\Sigma = \Sigma_1 \cup \Sigma_2 \cup \Sigma_3$ (the boundary of Ω):

$$k\,\frac{\partial T}{\partial n} = q_{in} \text{ on } \Sigma_1, \quad k\,\frac{\partial T}{\partial n} = 0 \text{ on } \Sigma_2, \quad T = T_3 \text{ (constant) } \text{ on } \Sigma_3, \tag{7.1}$$

where n always denotes the outward unit normal to the boundary surface Σ and $\partial T/\partial n$ is the normal derivative to the boundary surface Σ,

$$\frac{\partial T}{\partial n} = \nabla T \cdot n \quad (\nabla T = \text{ the gradient of } T). \tag{7.2}$$

The parameters appearing in (7.1) are

k = thermal conductivity (typ. 1.8W/cm×°C),

q_{in} = uniform inward thermal power flux at the source (positive constant).

The radius R_0 of the mounting surface Σ_1 is fixed so that the boundary surface Σ_1 is already given in the design problem.

For practical considerations, we assume that the diffuser is solid without interior hollows or cutouts. The class of shapes for the diffuser is characterized by the design parameter $L > 0$ and the positive function $R(z)$, $0 < z \le L$, with $R(0) = R_0 > 0$. They are volumes of revolution Ω about the z-axis generated by the surface A between the z-axis and the function $R(z)$ (cf. Figure 1.6 (B)), that is,

$$\Omega \stackrel{\text{def}}{=} \left\{(x,y,z)\,:\, 0 < z < L,\, x^2 + y^2 < R(z)^2\right\}. \tag{7.3}$$

So the shape of Ω is completely specified by the length $L > 0$ and the function $R(z) > 0$ on the interval $[0, L]$.

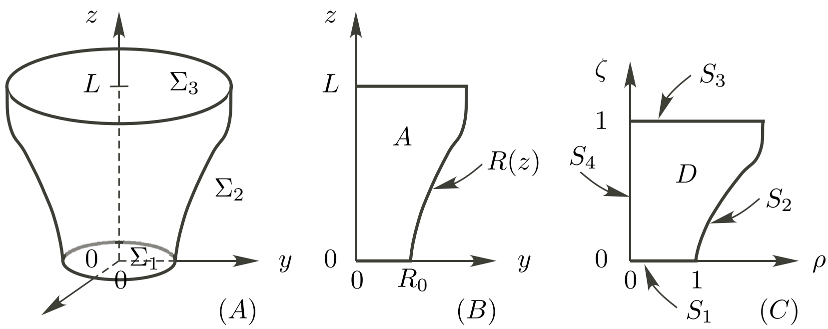

Figure 1.6. (A) Volume Ω and its boundary Σ; (B) Surface A generating Ω; (C) Surface D generating $\widetilde{\Omega}$.

Assuming that the diffuser is made up of a homogeneous material of uniform density (no hollow) the design objective is to minimize the volume

$$J(\Omega) \stackrel{\text{def}}{=} \int_\Omega dx = \pi \int_0^L R(z)^2\, dz \tag{7.4}$$

subject to a uniform constraint on the outward thermal power flux at the interface Σ_3 between the diffuser and the heatpipe saddle:

$$\sup_{p\in\Sigma_3} -k\,\frac{\partial T}{\partial z}(p) \le q_{out} \text{ or } k\,\frac{\partial T}{\partial n} + q_{out} \ge 0 \text{ on } \Sigma_3, \tag{7.5}$$

where q_{out} is a specified positive constant.

It is readily seen that the minimization problem (7.4) subject to the constraint (7.5) (where T is the solution of the heat equation with the boundary conditions (7.1)) is independent of the fixed temperature T_3 on the boundary Σ_3. In other words the optimal shape Ω, if it exists, is independent of T_3. As a result, from now on we set T_3 equal to 0.

7.3 Reformulation of the Problem

In a shape optimization problem the formulation is important from both the theoretical and the numerical viewpoints. In particular condition (7.5) is difficult to numerically handle since it involves the pointwise evaluation of the normal derivative on the piece of boundary Σ_3. This problem can be reformulated as the minimization of T on Σ_3, where T is now the solution of a variational inequality. Consider the following minimization problem over the subspace of functions that are positive or zero on Σ_3:

$$V^+(\Omega) \stackrel{\text{def}}{=} \left\{v \in H^1(\Omega) \, : \, v|_{\Sigma_3} \geq 0\right\}, \tag{7.6}$$

$$\inf_{v \in V^+(\Omega)} \int_\Omega \frac{1}{2}|\nabla v|^2 \, dx - \int_{\Sigma_1} q_{in} \, v \, d\Sigma + \int_{\Sigma_3} q_{out} \, v \, d\Sigma. \tag{7.7}$$

$H^1(\Omega)$ is the usual Sobolev space on the domain Ω, and the inequality on Σ_3 has to be interpreted *quasi-everywhere* in the capacity sense. Leaving aside those technicalities, the minimizing solution of (7.7) is characterized by

$$\begin{aligned} -k \, \Delta T = 0 \text{ in } \Omega, \quad k\frac{\partial T}{\partial n} = q_{in} \text{ on } \Sigma_1, \quad k\frac{\partial T}{\partial n} = 0 \text{ on } \Sigma_2, \\ T \geq 0, \quad \left(k\frac{\partial T}{\partial n} + q_{out}\right) \geq 0, \quad T\left(k\frac{\partial T}{\partial n} + q_{out}\right) = 0 \text{ on } \Sigma_3. \end{aligned} \tag{7.8}$$

The former constraint (7.5) is verified and replaced by the new constraint

$$T = 0 \text{ on } \Sigma_3. \tag{7.9}$$

If there exists a nonempty domain Ω of the form (7.3) such that $T = 0$ on Σ_3, the problem is feasible.

In this formulation the pointwise constraint on the normal derivative of the temperature on Σ_3 has been replaced by a pointwise constraint on the less demanding trace of the temperature on Σ_3. Yet, we now have to solve a variational inequality instead of a variational equation for the temperature T.

7.4 Scaling of the Problem

In the above formulations the *shape parameter* L and the *shape function* R are not independent of each other since the function R is defined on the interval $[0, L]$. This motivates the following changes of variables and the introduction of the dimensionless temperature y:

$$x \mapsto \xi_1 = \frac{x}{R_0}, \quad y \mapsto \xi_2 = \frac{y}{R_0}, \quad z \mapsto \zeta = \frac{z}{L}, \quad 0 \leq \zeta \leq 1,$$

$$\tilde{L} = \frac{L}{R_0}, \quad \tilde{R}(\zeta) = \frac{R(L\zeta)}{R_0},$$

$$y(\xi_1, \xi_2, \zeta) = \frac{k}{Lq_{in}} T(R_0\xi_1, R_0\xi_2, L\zeta),$$

$$D \stackrel{\text{def}}{=} \left\{(\xi_1, \xi_2, \zeta) \, : \, 0 < \zeta < 1, \, \xi_1^2 + \xi_2^2 < \tilde{R}(z)^2\right\}.$$

The parameter $\tilde{L}$ now appears as a coefficient in the partial differential equation

$$\tilde{L}^2\left(\frac{\partial^2 y}{\partial \xi_1^2}+\frac{\partial^2 y}{\partial \xi_2^2}\right)+\frac{\partial^2 y}{\partial \zeta^2}=0 \text{ in } D$$

with the following boundary conditions on the boundary $S = S_1 \cup S_2 \cup S_3$ of D:

$$\frac{\partial y}{\partial \nu_A}=1 \text{ on } S_1, \quad \frac{\partial y}{\partial \nu_A}=0 \text{ on } S_2, \quad y=0 \text{ on } S_3, \tag{7.10}$$

where ν denotes the outward normal to the boundary surface S and $\partial y/\partial \nu_A$ is the conormal derivative to the boundary surface S,

$$\frac{\partial y}{\partial \nu_A}=\tilde{L}^2\left(\nu_1\frac{\partial y}{\partial \xi_1}+\nu_2\frac{\partial y}{\partial \xi_2}\right)+\nu_3\frac{\partial y}{\partial \zeta}.$$

Finally, the optimal design problem depends only on the ratio $q = q_{out}/q_{in}$ through the constraint

$$\frac{\partial y}{\partial \nu_A}+\frac{q_{out}}{q_{in}} \geq 0 \text{ on } S_3.$$

The *design variables* are the parameter $\tilde{L} > 0$ and the function $\tilde{R} > 0$ now defined on the fixed interval $[0, 1]$.

7.5 Design Problem

The fact that this specific design problem can be reduced to finding a parameter and a function gives the false or unfounded impression that it can now be solved by standard mathematical programming and numerical methods. Early work on such problems revealed a different reality, such as oscillating boundaries and convergence towards nonphysical designs. Clearly, the geometry refused to be handled by standard methods without a better understanding of the underlying physics and its inception in the modeling of the geometric variable.

At the theoretical level, the existence of solution requires a concept of continuity with respect to the geometry of the solution of either the heat equation with an inequality constraint on the TPF or the variational inequality with an equality constraint on the temperature. The other element is the lower semicontinuity of the objective functional that is not too problematic for the volume functional as long as the chosen topology on the geometry preserves the continuity of the volume functional. For instance, the classical Hausdorff metric topology does not preserve the volume. In the context of fluid mechanics (cf., for instance, O. Pironneau [**1**]), it means that a drag minimizing sequence of sets with constant volume may converge to a set with twice the volume (cf. Example 4.1 in Chapter 6). A wine making industry exploiting the convergence in the Hausdorff metric topology could yield miraculous profits.

Other serious issues are, for instance, the lack of differentiability of the solution of a variational inequality at the continuous level that will inadvertently affect the differentiability or the evolution of a gradient method at the discrete level. We shall see that there is not only one topology for shapes but a whole range that selectively preserve some but not all of the geometrical features. Again the right choice is problem dependent, much like the choice of the right Sobolev space in the theory of partial differential equations.

8 Design of a Thermal Radiator

Current trends indicate that future communications satellites and spacecrafts will grow ever larger, consume ever more electrical power, and dissipate larger amounts of thermal energy. Various techniques and devices can be deployed to condition the thermal environment for payload boxes within a spacecraft, but it is desirable to employ those which offer good performance for low cost, low weight, and high reliability. A thermal radiator (or radiating fin) which accepts a given TPF from a payload box and radiates it directly to space can offer good performance and high reliability at low cost. However, without careful design, such a radiator can be unnecessarily bulky and heavy. It is the mass-optimized design of the thermal radiator which is the problem at hand (cf. M. C. Delfour, G. Payre, and J.-P. Zolésio [**2**]). We may assume that the payload box presents a uniform TPF (typ. 0.1 to 1.0 W/cm^2) into the radiator at the box/radiator interface. The radiating surface is a second surface mirror which consists of a sheet of glass whose inner surface has silver coating. We may assume that the TPF out of the radiator/space interface is governed by the T^4 radiation law, although we must account also for a constant TPF (typ. 0.01 W/cm^2) into this interface from the sun. Any other surfaces of the radiator may be treated as adiabatic. Two constraints restrict freedom in the design of the thermal radiator:

(i) the maximum temperature at the box/radiator interface is not to exceed some constant (typ. 50°); and

(ii) no part of the radiator is to be thinner than some constant (typ. 1 mm).

8.1 Statement of the Problem

Assume that the radiator is a volume Ω symmetrical about the z-axis (cf. Figure 1.7) whose boundary surface is made up of three regular pieces: the contact surface Σ_1 (a disk perpendicular to the z-axis with center at the point $(r,z) = (0,0)$), the lateral adiabatic surface Σ_2, and the radiating surface Σ_3 (a disk perpendicular to the z-axis with center at $(r,z) = (0,L)$). More precisely

$$\begin{aligned}\Sigma_1 &= \left\{(x,y,z) \,:\, z=0 \text{ and } x^2+y^2 \le R_0^2\right\},\\ \Sigma_2 &= \left\{(x,y,z) \,:\, x^2+y^2 = R(z)^2,\ 0\le z\le L\right\},\\ \Sigma_3 &= \left\{(x,y,z) \,:\, z=L \text{ and } x^2+y^2 \le R(L)^2\right\},\end{aligned} \tag{8.1}$$

where the radius $R_0 > 0$ (typ. 10 cm), the length $L > 0$, and the function

$$R \,:\, [0,L] \to \mathbf{R}, \quad R(0) = R_0, \quad R(z) > 0,\ 0 \le z \le L, \tag{8.2}$$

are given ($\mathbf{R}$, the field of real numbers).

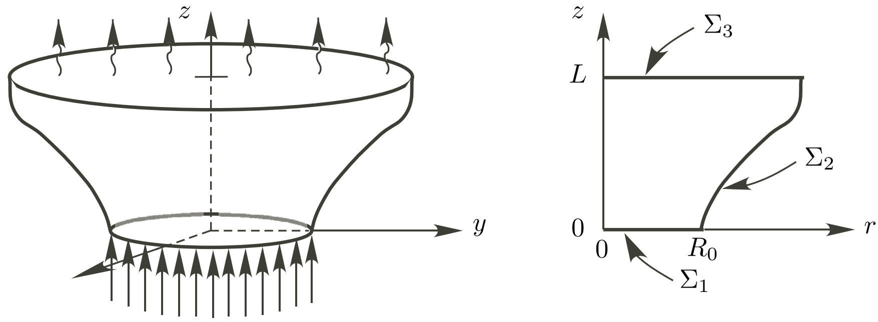

Figure 1.7. *Volume Ω and its cross section.*

The temperature distribution (in Kelvin degrees) over the volume Ω is the solution of the stationary heat equation

$$\Delta T = 0 \quad \text{(the Laplacian of } T) \tag{8.3}$$

with the following boundary conditions on the surface $\Sigma = \Sigma_1 \cup \Sigma_2 \cup \Sigma_3$ (the boundary of Ω):

$$k\frac{\partial T}{\partial n} = q_{in} \text{ on } \Sigma_1, \quad k\frac{\partial T}{\partial n} = 0 \text{ on } \Sigma_2, \quad k\frac{\partial T}{\partial n} = -\sigma\varepsilon T\,|T|^3 + q_s \text{ on } \Sigma_3, \tag{8.4}$$

where n denotes the outward normal to the boundary surface Σ, $\partial T/\partial n$ is the normal derivative on the boundary surface Σ, and

$$\frac{\partial T}{\partial n} = \nabla T \cdot n \quad (\nabla T = \text{ the gradient of } T).$$

The parameters appearing in (8.1)–(8.4) are

k = thermal conductivity (typ. 1.8 W/cm×°C),

q_{in} = uniform inward thermal power flux at the source (typ. 0.1 to 1.0 W/cm^2),

σ = Boltzmann's constant (5.67×10^{-8} W/m^{2}K^4),

ε = surface emissivity (typ. 0.8),

q_s = solar inward thermal power flux (0.01 W/cm^2).

The optimal design problem consists in minimizing the volume

$$J(\Omega) \stackrel{\text{def}}{=} \pi \int_0^L R(z)^2\, dz \tag{8.5}$$

over all length $L > 0$ and shape function R subject to the constraint

$$T(x, y, z) \le T_f \quad \text{(typ. 50°C)}, \quad \forall (x, y, z) \in \Sigma_1. \tag{8.6}$$

In this analysis we shall drop the requirement (ii) in the introduction.

8.2 Scaling of the Problem

As in the case of the diffuser, it is convenient to introduce the following dimensionless coordinates and temperature y (see Figure 1.8):

$$x \mapsto \xi_1 = \frac{x}{R_0}, \quad y \mapsto \xi_2 = \frac{y}{R_0}, \quad z \mapsto \zeta = \frac{z}{L}, \quad 0 \le \zeta \le 1,$$
$$\tilde{L} = \frac{L}{R_0}, \quad \tilde{R}(\zeta) = \frac{R(L\zeta)}{R_0},$$
$$y(\xi_1, \xi_2, \zeta) = \left(\frac{\sigma\varepsilon R_0}{k}\right)^{1/3} T(R_0\xi_1, R_0\xi_2, L\zeta),$$
$$D \overset{\text{def}}{=} \left\{(\xi_1, \xi_2, \zeta) \,:\, 0 < \zeta < 1,\, \xi_1^2 + \xi_2^2 < \tilde{R}(z)^2\right\}.$$

The parameter $\tilde{L}$ now appears as a coefficient in the partial differential equation

$$\tilde{L}^2 \left(\frac{\partial^2 y}{\partial \xi_1^2} + \frac{\partial^2 y}{\partial \xi_2^2}\right) + \frac{\partial^2 y}{\partial \zeta^2} = 0 \text{ in } D$$

with the following boundary conditions on the boundary $S = S_1 \cup S_2 \cup S_3$ of D:

$$\begin{aligned} \frac{\partial y}{\partial \nu_A} &= \tilde{L}\left[\left(\frac{\sigma\varepsilon R_0}{k}\right)^{1/3} \frac{q_{in}}{k} R_0\right] && \text{on } S_1, \\ \frac{\partial y}{\partial \nu_A} &= 0 && \text{on } S_2, \\ \frac{\partial y}{\partial \nu_A} + \tilde{L} y |y|^3 &= \tilde{L}\left[\left(\frac{\sigma\varepsilon R_0}{k}\right)^{1/3} \frac{q_{in}}{k} R_0\right] \frac{q_s}{q_{in}} && \text{on } S_3, \end{aligned} \tag{8.7}$$

where ν denotes the outward normal to the boundary surface S and $\partial y / \partial \nu_A$ is the conormal derivative to the boundary surface Σ,

$$\frac{\partial y}{\partial \nu_A} = \tilde{L}^2 \left(\nu_1 \frac{\partial y}{\partial \xi_1} + \nu_2 \frac{\partial y}{\partial \xi_2}\right) + \nu_3 \frac{\partial y}{\partial \zeta}.$$

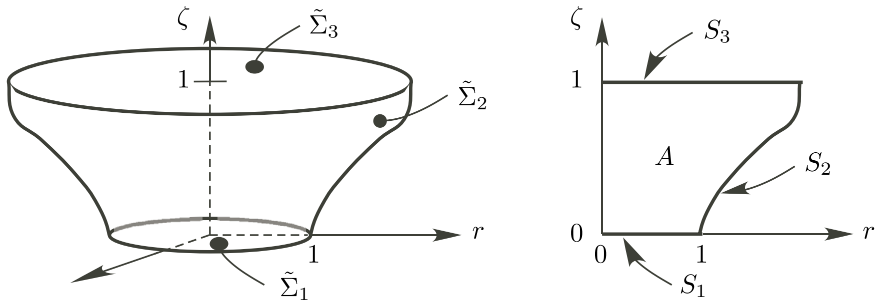

Figure 1.8. *Volume $\widetilde{\Omega}$ and its generating surface A.*

9 A Glimpse into Segmentation of Images

The study of the problem of linguistic or visual perceptions was initiated by several pioneering authors, such as H. Blum [**1**] in 1967, D. Marr and E. Hildreth [**1**] in 1980, and D. Marr [**1**] in 1982. It involves specialists of psychology, artificial intelligence, and experimentalists such as D. H. Hubel and T. N. Wiesel [**1**] in 1962 and F. W. Campbell and J. G. Robson [**1**] in 1968.

In the first part of this section, we revisit the pioneering work of D. Marr and E. Hildreth [**1**] in 1980 on the smoothing of the image by convolution with a sufficiently differentiable normalized function as a function of the *scaling parameter*. We extend the *space-frequency uncertainty principle* to N-dimensional images. It is the analogue of the *Heisenberg uncertainty principle* of quantum mechanics. We revisit the *Laplacian filter* and we generalize the *linearity assumption* of D. Marr and E. Hildreth [**1**] from linear to curved contours.

In the second part, we show how shape analysis methods and the shape and tangential calculuses can be applied to objective functionals defined on the whole contour of an image. We anticipate Chapters 9 and 10 on shape derivatives by the velocity method and show how they can be applied to snakes, active geodesic contours, and level sets (cf., for instance, the book of S. Osher and N. Paragios [**1**]). It shows that the Eulerian shape semiderivative is the basic ingredient behind those representations, including the case of the oriented distance function. In all cases, the evolution equation for the continuous gradient descent method is shown to have the same structure. For more material along the same lines, the reader is referred to M. Dehaes [**1**] and M. Dehaes and M. C. Delfour [**1**].

9.1 Automatic Image Processing

The first level of image processing is the detection of the contours or the boundaries of the objects in the image. For an ideal image $I : D \to \mathbf{R}$ defined in an open two-dimensional frame D with values in an interval of greys continuously ranging from white to black (see Figure 1.9), the edges of an object correspond to the loci of discontinuity of the image I (cf. D. Marr and E. Hildreth [**1**]), also called "step edges" by D. Marr [**1**]. As can be seen from Figure 1.9, the loci of discontinuity may only reveal part of an object hidden by another one and a subsequent and different level of processing is required. A more difficult case is the detection of black curves or cracks in a white frame (cf. Figure 1.10), where the function I becomes a measure supported by the curves rather than a function.

In practice, the frame D of the image is divided into periodically spaced cells P (square, hexagon, diamond) with a quantized value or pixel from 256 grey levels. For small squares a piecewise linear continuous interpolation or higher degree C^1-interpolation can be used to remove the discontinuities at the intercell boundaries. In addition, observations and measurements introduce noise or perturbations and the interpolated image I needs to be further *smoothed* or *filtered*. When the characteristics of the noise are known, an appropriate filter can do the job (see, for instance, the use of *low pass filters* in A. Rosenfeld and M. Thurston [**1**], A. P. Witkin [**1**], and A. L. Yuille and T. Poggio [**1**]).

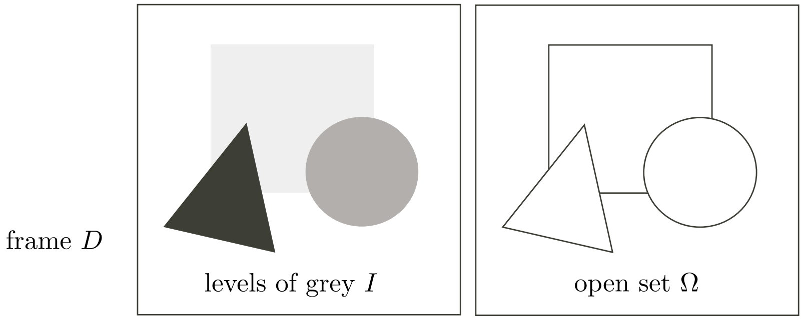

Figure 1.9. *Image I of objects and their segmentation in the frame D.*

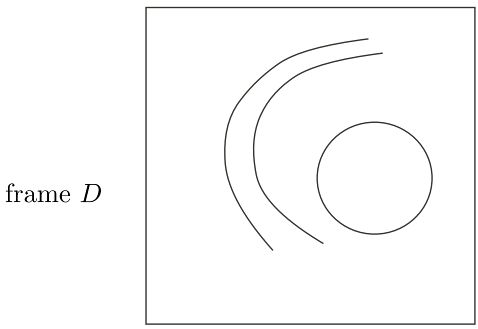

Figure 1.10. *Image I containing black curves or cracks in the frame D.*

Given an ideal *grey level image I* defined in a fixed bounded open *frame D* we want to identify the *edges* or boundaries of the objects contained in the image as shown in Figure 1.9. The *edge identification* or *segmentation* problem is a low-level processing of a *real* image which should also involve higher-level processings or tasks (cf. M. Kass, M. Witkin, and D. Terzopoulos [**1**]). Intuitively the edges coincide with the loci of discontinuities of the image function I. At such points the norm $|\nabla I|$ of the gradient ∇I is infinite. Roughly speaking the segmentation of an ideal image I consists in finding the loci of discontinuity of the function I. The idea is now to first smooth I by convolution with a sufficiently differentiable function. This operation will be followed and/or combined with the use of an *edge detector* as the zero crossings of the Laplacian. In the literature the term filter often applies to both the filter and the detector. In this section we shall make the distinction between the two operations.

9.2 Image Smoothing/Filtering by Convolution and Edge Detectors

In this section we revisit the work of D. Marr and E. Hildreth [**1**] in 1980 on the smoothing of the image I by convolution with a sufficiently differentiable normalized function ρ as a function of the *scaling parameter* $\varepsilon > 0$. In the second part of this section we revisit the *Laplacian filter* and we generalize the *linearity assumption* of D. Marr and E. Hildreth [**1**] from linear to curved contours.

9.2.1 Construction of the Convolution of I

Let $\rho : \mathbf{R}^{\mathrm{N}} \to \mathbf{R}$, $N \geq 1$, be a sufficiently smooth function such that

$$\rho \geq 0 \qquad \text{and} \qquad \int_{\mathbf{R}^{\mathrm{N}}} \rho(x)\, dx = 1.$$

Associate with the image I, ρ, and a *scaling parameter* $\varepsilon > 0$ the normalized convolution

$$I_\varepsilon(x) \overset{\text{def}}{=} (I * \rho_\varepsilon)(x) = \frac{1}{\varepsilon^N} \int_{\mathbf{R}^{\mathrm{N}}} I(y)\, \rho\left(\frac{x-y}{\varepsilon}\right) dy, \tag{9.1}$$

where for $x \in \mathbf{R}^{\mathrm{N}}$

$$\rho_\varepsilon(x) \overset{\text{def}}{=} \frac{1}{\varepsilon^N} \rho\left(\frac{x}{\varepsilon}\right) \qquad \text{and} \qquad \int_{\mathbf{R}^{\mathrm{N}}} \rho_\varepsilon(x)\, dx = 1. \tag{9.2}$$

The function ρ_ε plays the role of a probability density and $\rho_\varepsilon(x)\, dx$ of a probability measure. Under appropriate conditions I_ε converges to the original image I as $\varepsilon \to 0$. For a sufficiently small $\varepsilon > 0$ the loci of discontinuity of the function I are transformed into loci of strong variation of the gradient of the convolution I_ε. When ρ has compact support, the convolution acts locally around each point in a neighborhood whose size is of order ε. A popular choice for ρ is the *Gaussian* with integral normalized to one in $\mathbf{R}^{\mathrm{N}}$ defined by

$$G^N(x) \overset{\text{def}}{=} \frac{1}{(\sqrt{2\pi})^N}\, e^{-\frac{1}{2}|x|^2} \qquad \text{and} \qquad \int_{\mathbf{R}^{\mathrm{N}}} G^N(x)\, dx = 1$$

and the *normalized Gaussian of variance* ε

$$G^N_\varepsilon(x) = \frac{1}{\varepsilon^N} \frac{1}{(\sqrt{2\pi})^N}\, e^{-\frac{1}{2}\left|\frac{x}{\varepsilon}\right|^2} \qquad \text{and} \qquad \int_{\mathbf{R}^{\mathrm{N}}} G^N_\varepsilon(x)\, dx = 1. \tag{9.3}$$

As noted in L. Alvarez, P.-L. Lions, and J.-M. Morel [**1**] in dimension $N = 2$, the function $u(t) = G^2_{\sqrt{2t}} * I$ is the solution of the parabolic equation

$$\frac{\partial u}{\partial t}(t,x) = \Delta u(t,x), \quad u(0,x) = I(x).$$

9.2.2 Space-Frequency Uncertainty Relationship

It is interesting to compute the Fourier transform $\mathcal{F}(G^N_\varepsilon)$ of G^N_ε to make explicit the relationship between the mean square deviations of G^N_ε and its Fourier transform $\mathcal{F}(G^N_\varepsilon)$. For $\omega \in \mathbf{R}^{\mathrm{N}}$, define the *Fourier transform* of a function f by

$$\mathcal{F}(f)(\omega) \overset{\text{def}}{=} \frac{1}{(\sqrt{2\pi})^N} \int_{\mathbf{R}^{\mathrm{N}}} f(x)\, e^{-i\,\omega\cdot x}. \tag{9.4}$$

In applications the integral of the square of f often corresponds to an energy. We shall refer to the L^2-norm of f as the *energy norm* and the function $f^2(x)$ as the *energy density.* The Fourier transform (9.4) of the normalized Gaussian (9.3) is

$$\mathcal{F}(G_\varepsilon^N)(\omega) \stackrel{\text{def}}{=} \frac{1}{(\sqrt{2\pi})^N} \int_{\mathbf{R}^N} G_\varepsilon^N(x)\, e^{-i\,\omega\cdot x} dx = \frac{1}{(\sqrt{2\pi})^N}\, e^{-\frac{1}{2}|\varepsilon\omega|^2}. \tag{9.5}$$

D. Marr and E. Hildreth [**1**] notice that there is an uncertainty relationship between the mean square deviation with respect to the energy density $f(x)^2$ and the mean square deviation with respect to the energy density $\mathcal{F}(f)(\omega)^2$ of its Fourier transform in dimension 1 using a result of R. Bracewell [**1**, pp. 160–161], where he uses the constant $1/(2\pi)$ instead of $1/\sqrt{2\pi}$ in the definition of the Fourier transform that yields $1/(4\pi)$ instead of $1/2$ as a lower bound to the product of the two mean square deviations.

This *uncertainty relationship* generalizes to dimension N. First define the notions of *centroïd* and *variance* with respect to the energy density $f(x)^2$. Given a function $f \in L^1(\mathbf{R}^N) \cap L^2(\mathbf{R}^N)$ and $x \in \mathbf{R}^N$, define the *centroïd* as

$$\overline{x} \stackrel{\text{def}}{=} \frac{\int_{\mathbf{R}^N} x\, f(x)^2\, dx}{\int_{\mathbf{R}^N} f(x)^2\, dx} \tag{9.6}$$

and the *variance* as

$$\langle \Delta x\rangle^2 \stackrel{\text{def}}{=} \langle x - \overline{x}\rangle^2 = \frac{\int_{\mathbf{R}^N} |x|^2\, f(x)^2\, dx}{\int_{\mathbf{R}^N} f(x)^2\, dx} - |\overline{x}|^2.$$

Theorem 9.1 (Uncertainty relationship)**.** *Given $N \geq 1$ and $f \in W^{1,1}(\mathbf{R}^N) \cap W^{1,2}(\mathbf{R}^N)$ such that $-ix\, f \in L^1(\mathbf{R}^N) \cap L^2(\mathbf{R}^N)$,*

$$\boxed{\langle \Delta x\rangle\, \langle \Delta \omega\rangle\, \geq\, \frac{N}{2}.} \tag{9.7}$$

For $f = G_\varepsilon^N$

$$f(x) = G_\varepsilon^N(x) = \frac{1}{\varepsilon^N}\frac{1}{(\sqrt{2\pi})^N}\, e^{-\frac{1}{2}\left|\frac{x}{\varepsilon}\right|^2} \quad \Rightarrow \langle \Delta x\rangle^2 \,=\, \frac{N\varepsilon^2}{2}, \tag{9.8}$$

$$\hat{f}(\omega) = \mathcal{F}(G_\varepsilon^N)(\omega) = \frac{1}{(\sqrt{2\pi})^N}\, e^{-\frac{1}{2}|\varepsilon\omega|^2} \quad \Rightarrow \langle \Delta \omega\rangle^2 \,=\, \frac{N}{2\,\varepsilon^2} \tag{9.9}$$

$$\Rightarrow \boxed{\langle \Delta x\rangle\langle \Delta \omega\rangle \,=\, \frac{N}{2} \qquad \text{for } G_\varepsilon^N \text{ and } \mathcal{F}(G_\varepsilon^N).} \tag{9.10}$$

So the *normalized Gaussian filter* is indeed an *optimal* filter since it achieves the lower bound in all dimensions as stated by D. Marr and E. Hildreth [**1**] in dimension 1 in 1980. In the context of quantum mechanics, this relationship is the analogue of the *Heisenberg uncertainty principle.*

9.2.3 Laplacian Detector

One way to detect the edges of a regular object is to start from the convolution-smoothed image I_ε. Given a direction v, $|v| = 1$, and a point $x \in \mathbf{R}^2$ an *edge point* will correspond to a local minimum or maximum of the directional derivative $f(t) \stackrel{\text{def}}{=} \nabla I_\varepsilon(x+tv)\cdot v$ with respect to t. Denote by $\hat{t}$ such a point. Then a necessary condition for an extremal is

$$D^2 I_\varepsilon(x+\hat{t}v)v\cdot v = 0.$$

The point $\hat{x} = x + \hat{t}v$ is a *zero crossing* following the terminology of D. Marr and E. Hildreth [**1**] of the second-order directional derivative in the direction v. Thus we are looking for the pairs $(\hat{x}, \hat{v})$ verifying the necessary condition

$$D^2 I_\varepsilon(\hat{x})\hat{v}\cdot\hat{v} = 0 \tag{9.11}$$

and, more precisely, lines or curves $\mathcal{C}$ such that

$$\forall x \in \mathcal{C},\ \exists v(x),\ |v(x)| = 1, \quad D^2 I_\varepsilon(x)v(x)\cdot v(x) = 0. \tag{9.12}$$

This condition is necessary in order that, in each point $x \in \mathcal{C}$, there exists a direction $v(x)$ such that $\nabla I_\varepsilon(x)\cdot v$ is extremal.

In order to limit the search to points x rather than to pairs (x, v), the Laplacian detector was introduced in D. Marr and E. Hildreth [**1**] under two assumptions: the linear variation of the intensity along the edges $\mathcal{C}$ of the object and the condition of zero crossing of the second-order derivative in the direction normal to $\mathcal{C}$.

Assumption 9.1 (Linear variation condition)**.**
The intensity I_ε in a neighborhood of lines parallel to $\mathcal{C}$ is locally linear (affine). □

Assumption 9.2.
The zero-crossing condition is verified in all points of $\mathcal{C}$ in the direction of the normal n at the point, that is,

$$\boxed{D^2 I_\varepsilon\, n\cdot n = 0 \text{ on } \mathcal{C}.} \tag{9.13}$$

□

Under those two assumptions and for a line $\mathcal{C}$, it is easy to show that the points of $\mathcal{C}$ verify the necessary condition

$$\Delta I_\varepsilon = 0 \text{ on } \mathcal{C}. \tag{9.14}$$

Such conditions can be investigated for edges $\mathcal{C}$ that are the boundary Γ of a smooth domain $\Omega \subset R^N$ of class C^2 by using the *tangential calculus* developed in section 5.2 of Chapter 9 and M. C. Delfour [**7**, **8**] for objects in $\mathbf{R}^{\mathrm{N}}$, $N \geq 1$, and not just in dimension $N = 2$. Indeed we want to find points $x \in \Gamma$ such that the function $\nabla I_\varepsilon(x)\cdot n(x)$ is an extremal of the function $f(t) = \nabla I_\varepsilon(x+t\,\nabla b_\Omega(x))\cdot\nabla b_\Omega(x)$ in $t = 0$. This yields the *local necessary condition* $D^2 I_\varepsilon \nabla b_\Omega \cdot \nabla b_\Omega = D^2 I_\varepsilon\, n\cdot n = 0$ on Γ of Assumption 9.2. As for Assumption 9.1, its analogue is the following.

Assumption 9.3.
The restriction to the curve $\mathcal{C}$ of the gradient of the intensity I_ε is a constant vector c, that is,

$$\nabla I_\varepsilon = c, \quad c \text{ is a constant vector on } \mathcal{C}. \tag{9.15}$$

□

Indeed the Laplacian of I_ε on $\mathcal{C}$ can be decomposed as follows (cf. section 5.2 of Chapter 9):

$$\Delta I_\varepsilon = \operatorname{div}_\Gamma \nabla I_\varepsilon + D^2 I_\varepsilon \, n \cdot n,$$

where $\operatorname{div}_\Gamma \nabla I_\varepsilon$ is the *tangential divergence* of ∇I_ε that can be defined as follows:

$$\operatorname{div}_\Gamma \nabla I_\varepsilon = \operatorname{div}(\nabla I_\varepsilon \circ p_{\mathcal{C}})|_{\mathcal{C}}$$

and $p_{\mathcal{C}}$ is the *projection* onto $\mathcal{C}$. Hence, under Assumption 9.3, $\operatorname{div}_\Gamma(\nabla I_\varepsilon) = 0$ on $\mathcal{C}$ since

$$\operatorname{div}_\Gamma(\nabla I_\varepsilon) = \operatorname{div}(\nabla I_\varepsilon \circ p_\Gamma)|_\Gamma = \operatorname{div}(c)|_\Gamma = 0,$$

and, in view of (9.13) of Assumption 9.2, $D^2 I_\varepsilon \, n \cdot n = 0$ on $\mathcal{C}$. Therefore

$$\Delta I_\varepsilon = \operatorname{div}_\Gamma(\nabla I_\varepsilon) + D^2 I_\varepsilon \, n \cdot n = 0 \text{ on } \mathcal{C},$$

and we obtain the necessary condition

$$\boxed{\Delta I_\varepsilon = 0 \text{ on } \mathcal{C}.} \tag{9.16}$$

9.3 Objective Functions Defined on the Whole Edge

With the pioneering work of M. Kass, M. Witkin, and D. Terzopoulos [**1**] in 1988 we go from a local necessary condition at a point of the edge to a global necessary condition by introducing objective functionals defined on the entire edge of an object. Here many computations and analytical studies can be simplified by adopting the point of view that a closed curve in the plane is the boundary of a set and using the whole machinery developed for shape and geometric analysis and the tangential and shape calculuses in Chapter 9.

9.3.1 Eulerian Shape Semiderivative

In this section we briefly summarize the main elements of the *velocity method*.

Definition 9.1.
Given D, $\varnothing \neq D \subset \mathbf{R}^N$, consider the set $\mathcal{P}(D) = \{\Omega : \Omega \subset D\}$ of subsets of D. The set D is the *holdall* or the *universe*. A *shape functional* is a map $J : \mathcal{A} \to E$ from an *admissible family* $\mathcal{A}$ in $\mathcal{P}(D)$ with values in a topological vector space E (usually $\mathbf{R}$). □

Given a *velocity field* $V\colon [0,\tau] \times \mathbf{R}^N \to \mathbf{R}^N$ (the notation $V(t)(x) = V(t,x)$ will often be used), consider the transformations

$$T(t,X) \overset{\text{def}}{=} x(t,X), \qquad t \geq 0,\ X \in \mathbf{R}^N, \tag{9.17}$$

where $x(t,X) = x(t)$ is defined as the *flow* of the differential equation

$$\frac{dx}{dt}(t) = V(t,x(t)),\ t \geq 0, \quad x(0,X) = X \tag{9.18}$$

(here the notation $x \mapsto T_t(x) = T(t,x) : \mathbf{R}^N \to \mathbf{R}^N$ will be used). The *Eulerian shape semiderivative* of J in Ω in the direction V is defined as

$$dJ(\Omega;V) \overset{\text{def}}{=} \lim_{t \searrow 0} \frac{J(\Omega_t) - J(\Omega)}{t} \tag{9.19}$$

(when the limit exists in E), where $\Omega_t = T_t(\Omega) = \{T_t(x) : x \in \Omega\}$. Under appropriate assumptions on the family $\{V(t)\}$, the transformations $\{T_t\}$ are homeomorphisms that transport the boundary Γ of Ω onto the boundary Γ_t of Ω_t and the interior Ω onto the interior of Ω_t.

9.3.2 From Local to Global Conditions on the Edge

For simplicity, drop the subscript ε of the convolution-smoothed image I_ε and assume that the edge of the object is the boundary Γ of an open domain Ω of class C^2. Following V. Caselles, R. Kimmel, and G. Sapiro [**1**], it is important to choose objective functionals that are intrinsically defined and do not depend on an arbitrary parametrization of the boundary. For instance, given a *frame* $D =]0,a[\times]0,b[$ and a *smoothed image* $I : D \to \mathbf{R}$, find an extremum of the objective functional

$$E(\Omega) \overset{\text{def}}{=} \int_\Gamma \frac{\partial I}{\partial n}\, d\Gamma, \tag{9.20}$$

where the integrand is the normal derivative of I. Using the velocity method the *Eulerian shape directional semiderivative* is given by the expression (cf. Chapter 9)

$$dE(\Omega;V) = \int_\Gamma \left[H \frac{\partial I}{\partial n} + \frac{\partial}{\partial n}\left(\frac{\partial I}{\partial n}\right)\right] V \cdot n\, d\Gamma, \tag{9.21}$$

where $H = \Delta b_\Omega$ is the mean curvature and $n = \nabla b_\Omega$ is the outward unit normal. Proceeding in a formal way a necessary condition would be

$$\begin{aligned} H \frac{\partial I}{\partial n} + \frac{\partial}{\partial n}\left(\frac{\partial I}{\partial n}\right) &= 0 \text{ on } \Gamma \\ \Rightarrow \Delta b_\Omega \nabla I \cdot \nabla b_\Omega + \nabla(\nabla I \cdot \nabla b_\Omega) \cdot \nabla b_\Omega &= 0 \\ \Rightarrow \Delta b_\Omega \nabla I \cdot \nabla b_\Omega + D^2 I \nabla b_\Omega \cdot \nabla b_\Omega + D^2 b_\Omega \nabla I \cdot \nabla b_\Omega &= 0 \\ \Rightarrow \boxed{D^2 I\, n \cdot n + H \frac{\partial I}{\partial n} = 0 \text{ on } \Gamma.} \end{aligned} \tag{9.22}$$

This *global condition* is to be compared with the *local condition* (9.13). It can also be expressed in terms of the Laplacian and the tangential Laplacian of I as

$$\boxed{\Delta I - \Delta_\Gamma I = 0 \text{ on } \Gamma,} \tag{9.23}$$

which can be compared with the local condition (9.16), $\Delta I = 0$. This arises from the decomposition of the Laplacian of I with respect to Γ using the following identity for a smooth vector function U:

$$\operatorname{div} U = \operatorname{div}_\Gamma U + DU\, n \cdot n, \tag{9.24}$$

where the *tangential divergence* of U is defined as

$$\operatorname{div}_\Gamma U = \operatorname{div} (U \circ p_\Gamma)|_\Gamma$$

and p_Γ is the projection onto Γ. Applying this to $U = \nabla I$ and recalling the definition of the tangential Laplacian

$$\Delta I = \operatorname{div}_\Gamma \nabla I + D^2 I\, n \cdot n = \Delta_\Gamma I + H \frac{\partial I}{\partial n} + D^2 I\, n \cdot n,$$

where the tangential gradient $\nabla_\Gamma I$ and the *Laplace–Beltrami* operator $\Delta_\Gamma I$ are defined as

$$\nabla_\Gamma I = \nabla (I \circ p_\Gamma)|_\Gamma \quad \text{and} \quad \Delta_\Gamma I = \operatorname{div}_\Gamma (\nabla_\Gamma I).$$

9.4 Snakes, Geodesic Active Contours, and Level Sets

9.4.1 Objective Functions Defined on the Contours

In the literature the *objective* or *energy functional* is generally made up of two terms: one (image energy) that depends on the image and one (internal energy) that specifies the smoothness of Γ. A general form of objective functional is

$$E(\Omega) \stackrel{\text{def}}{=} \int_\Gamma g(I)\, d\Gamma, \quad g(I) \text{ is a function of } I. \tag{9.25}$$

The directional semiderivative with respect to a velocity field V is given by

$$dE(\Omega; V) = \int_\Gamma \left[H g(I) + \frac{\partial}{\partial n} g(I) \right] n \cdot V\, d\Gamma. \tag{9.26}$$

This gradient will make the *snakes* move and will *activate* the contours.

9.4.2 Snakes and Geodesic Active Contours

If a *gradient descent method* is used to minimize (9.25) starting from an initial curve $\mathcal{C}_0 = \mathcal{C}$, the iterative process is equivalent to following the evolution $\mathcal{C}_t$ (boundary of the smooth domain Ω_t) of the closed curve $\mathcal{C}$ given by the equation

$$\frac{\partial \mathcal{C}_t}{\partial t} = -\left[H_t\, g_I + (\nabla g_I \cdot n_t) \right] n_t \text{ on } \mathcal{C}_t, \tag{9.27}$$

where H_t is the mean curvature, n_t is the unit exterior normal, and the right-hand side of the equation is formally the "derivative" of (9.25) given by (9.26). Equation (9.27) is referred to as the *geodesic flow.* For $g_I = 1$ it is the *motion by mean curvature*

$$\frac{\partial \mathcal{C}_t}{\partial t} = -H_t\, n_t \text{ on } \mathcal{C}_t. \tag{9.28}$$

9.4.3 Level Set Method

The idea to represent the contours $\mathcal{C}_t$ by the zero-level set of a function $\varphi_t(x) = \varphi(t,x)$ for a function $\varphi : [0,\tau] \times \mathbf{R}^N \to \mathbf{R}$ by setting

$$\mathcal{C}_t = \left\{x \in \mathbf{R}^N : \varphi(t,x) = 0\right\} = \varphi_t^{-1}\{0\}$$

and to replace (9.27) by an equation for φ is due to S. Osher and J. A. Sethian [**1**] in 1988. This approach seems to have been simultaneously introduced in image processing by V. Caselles, F. Catté, T. Coll, and F. Dibos [**1**] in 1993 under the name "geometric partial differential equations" with, in addition to the mean curvature term, a "transport" term and by R. Malladi, J. A. Sethian, and B. C. Vemuri [**1**] in 1995 under the name "level set approach" combined with the notion of "extension velocity."

Let $(t,x) \mapsto \varphi(t,x) : [0,\tau] \times \mathbf{R}^N \to \mathbf{R}$ be a smooth function and Ω be a subset of $\mathbf{R}^N$ of boundary $\Gamma = \overline{\Omega} \cap \overline{\complement\Omega}$ such that

$$\operatorname{int}\Omega = \left\{x \in \mathbf{R}^N : \varphi(0,x) < 0\right\} \quad \text{and} \quad \Gamma = \left\{x \in \mathbf{R}^N : \varphi(0,x) = 0\right\}. \tag{9.29}$$

Let $V : [0,\tau] \times \mathbf{R}^N \to \mathbf{R}^N$ be a sufficiently smooth velocity field so that the transformations $\{T_t\}$ are *diffeomorphisms.* Moreover, assume that the images $\Omega_t = T_t(\Omega)$ verify the following properties: for all $t \in [0,\tau]$

$$\operatorname{int}\Omega_t = \left\{x \in \mathbf{R}^N : \varphi(t,x) < 0\right\} \quad \text{and} \quad \Gamma_t = \left\{x \in \mathbf{R}^N : \varphi(t,x) = 0\right\}. \tag{9.30}$$

Assuming that the function $\varphi_t(x) = \varphi(t,x)$ is at least of class C^1 and that $\nabla\varphi_t \neq 0$ on $\varphi_t^{-1}\{0\}$, the *total derivative* with respect to t of $\varphi(t, T_t(x))$ for $x \in \Gamma$ yields

$$\frac{\partial}{\partial t}\varphi(t,T_t(x)) + \nabla\varphi(t,T_t(x)) \cdot \frac{d}{dt}T_t(x) = 0. \tag{9.31}$$

By substituting the velocity field in (9.18), we get

$$\begin{aligned} &\frac{\partial}{\partial t}\varphi(t,T_t(x)) + \nabla\varphi(t,T_t(x)) \cdot V(t,T_t(x)) = 0, \quad \forall\, T_t(x) \in \Gamma_t \\ &\Rightarrow \boxed{\frac{\partial}{\partial t}\varphi_t + \nabla\varphi_t \cdot V(t) = 0 \text{ on } \Gamma_t,\ t \in [0,\tau].} \end{aligned} \tag{9.32}$$

This last equation, the *level set evolution equation*, is verified *only* on the boundaries or *fronts* Γ_t, $0 \le t \le \tau$. Can we find a representative φ in the equivalence class

$$[\varphi]_{\Omega,V} = \left\{\varphi : \varphi_t^{-1}\{0\} = \Gamma_t \text{ and } \varphi_t^{-1}\{<0\} = \operatorname{int}\Omega_t, \forall t \in [0,\tau]\right\}$$

of functions φ that verify conditions (9.29) and (9.30) in order to extend (9.32) from Γ_t to the whole $\mathbf{R}^N$ or at least almost everywhere in $\mathbf{R}^N$? Another possibility is to consider the larger equivalence class

$$[\varphi]_{\Gamma,V} = \left\{\varphi : \varphi_t^{-1}\{0\} = \Gamma_t, \forall t \in [0,\tau]\right\}.$$

9.4.4 Velocity Carried by the Normal

In D. Adalsteinsson and J. A. Sethian [**1**], J. Gomes and O. Faugeras [**1**], and R. Malladi, J. A. Sethian, and B. C. Vemuri [**1**], the front moves under the effect of a velocity field carried by the normal with a scalar velocity that depends on the curvatures of the level set or the front. So we are led to consider velocity fields V of the form

$$V(t)|_{\Gamma_t} = v(t)\, n_t, \quad x \in \Gamma_t, \tag{9.33}$$

for a *scalar velocity* $(t,x) \mapsto v(t)(x) : [0,\tau] \times \mathbf{R}^N \to \mathbf{R}$. By using the expression $\nabla\varphi_t/|\nabla\varphi_t|$ of the exterior normal n_t as a function of $\nabla\varphi_t$, (9.32) now becomes

$$\boxed{\frac{\partial}{\partial t}\varphi_t + v(t)\,|\nabla\varphi_t| = 0 \text{ on } \Gamma_t.} \tag{9.34}$$

For instance, consider the example of the minimization of the total length of the curve $\mathcal{C}$ of (9.27) for the metric $g(I)\, d\mathcal{C}$ as introduced by V. Caselles, R. Kimmel, and G. Sapiro [**1**]. By using the computation (9.26) of the shape semiderivative with respect to the velocity field V of the objective functional (9.25), a natural direction of descent is given by

$$\boxed{v(t)n_t, \quad v(t) = -\left[H_t\, g(I) + \frac{\partial}{\partial n_t} g(I)\right] \text{ on } \Gamma_t.} \tag{9.35}$$

The normal n_t and the mean curvature H_t can be expressed as a function of $\nabla\varphi_t$:

$$n_t = \frac{\nabla\varphi_t}{|\nabla\varphi_t|} \quad \text{and} \quad H_t = \operatorname{div}_{\Gamma_t} n_t = \operatorname{div}_{\Gamma_t}\left(\frac{\nabla\varphi_t}{|\nabla\varphi_t|}\right) = \operatorname{div}\left(\frac{\nabla\varphi_t}{|\nabla\varphi_t|}\right)\Big|_{\Gamma_t}. \tag{9.36}$$

By substituting in (9.34), we get the following evolution equation:

$$\frac{\partial\varphi_t}{\partial t} - \left(H_t\, g(I) + \frac{\partial}{\partial n_t} g(I)\right) \nabla\varphi_t \cdot \frac{\nabla\varphi_t}{|\nabla\varphi_t|} = 0 \text{ on } \Gamma_t \tag{9.37}$$

and

$$\begin{cases} \dfrac{\partial\varphi_t}{\partial t} - \left(H_t\, g(I) + \dfrac{\partial}{\partial n_t} g(I)\right) |\nabla\varphi_t| = 0 & \text{on } \Gamma_t, \\ \varphi_0 = \varphi^0 & \text{on } \Gamma_0. \end{cases} \tag{9.38}$$

By substituting expressions (9.36) for n_t and H_t in terms of $\nabla\varphi_t$, we finally get

$$\boxed{\begin{cases} \dfrac{\partial \varphi_t}{\partial t} - \left[\operatorname{div}\left(\dfrac{\nabla\varphi_t}{|\nabla\varphi_t|}\right) g(I) + \nabla g(I)\cdot \dfrac{\nabla\varphi_t}{|\nabla\varphi_t|}\right]|\nabla\varphi_t| = 0 & \text{on } \Gamma_t, \\ \varphi_0 = \varphi^0 & \text{on } \Gamma_0. \end{cases}} \tag{9.39}$$

The negative sign arises from the fact that we have chosen the outward rather than the inward normal. The main references to the existence and uniqueness theorems related to (9.39) can be found in Y. G. Chen, Y. Giga, and S. Goto [**1**].

9.4.5 Extension of the Level Set Equations

Equation (9.39) on the fronts Γ_t is not convenient from both the theoretical and the numerical viewpoints. So it would be desirable to be able to extend (9.39) in a small *tubular neighborhood* of thickness h

$$U_h(\Gamma_t) \stackrel{\text{def}}{=} \left\{x \in \mathbf{R}^{\mathrm{N}} \,:\, d_{\Gamma_t}(x) < h\right\} \tag{9.40}$$

of the front Γ_t for a small $h > 0$ (d_{Γ_t}, the distance function to Γ_t). In theory, the velocity field associated with Γ_t is given by

$$V(t) = -\left[\operatorname{div}\left(\frac{\nabla\varphi_t}{|\nabla\varphi_t|}\right) g(I) + \nabla g(I)\cdot \frac{\nabla\varphi_t}{|\nabla\varphi_t|}\right]\frac{\nabla\varphi_t}{|\nabla\varphi_t|} \text{ on } \Gamma_t. \tag{9.41}$$

Given the identities (9.36), the expressions of n_t and H_t extend to a neighborhood of Γ_t and possibly to $\mathbf{R}^{\mathrm{N}}$ if $\nabla\varphi_t$ is sufficiently smooth and $\nabla\varphi_t \neq 0$ on Γ_t.

Equation (9.39) can be extended to $\mathbf{R}^{\mathrm{N}}$ at the price of violating the assumptions (9.29)–(9.30), either by loss of smoothness of φ_t or by allowing its gradient to be zero:

$$\boxed{\begin{cases} \dfrac{\partial \varphi_t}{\partial t} - \left[\operatorname{div}\left(\dfrac{\nabla\varphi_t}{|\nabla\varphi_t|}\right) g(I) + \nabla g(I)\cdot \dfrac{\nabla\varphi_t}{|\nabla\varphi_t|}\right]|\nabla\varphi_t| = 0 & \text{in } \mathbf{R}^{\mathrm{N}}, \\ \varphi_0 = \varphi^0 & \text{in } \mathbf{R}^{\mathrm{N}}. \end{cases}} \tag{9.42}$$

In fact, the starting point of V. Caselles, F. Catté, T. Coll, and F. Dibos [**1**] was the following *geometric partial differential equation*:

$$\begin{cases} \dfrac{\partial \varphi_t}{\partial t} - \left[\operatorname{div}\left(\dfrac{\nabla\varphi_t}{|\nabla\varphi_t|}\right) g(I) + \nu\, g(I)\right]|\nabla\varphi_t| = 0 & \text{in } \mathbf{R}^{\mathrm{N}}, \\ \varphi_0 = \varphi^0 & \text{in } \mathbf{R}^{\mathrm{N}} \end{cases} \tag{9.43}$$

for a constant $\nu > 0$ and the function $g(I) = 1/(1+|\nabla I|^2)$. They prove the existence, in dimension 2, of a *viscosity solution* unique for initial data $\varphi^0 \in C([0,1]\times[0,1]) \cap W^{1,\infty}([0,1]\times[0,1])$ and $g \in W^{1,\infty}(\mathbf{R}^2)$.

9.5 Objective Function Defined on the Whole Image

9.5.1 Tikhonov Regularization/Smoothing

The convolution is only one approach to smoothing an image $I \in L^2(D)$ defined in a frame D. Another way is to use the *Tikhonov regularization*: given $\varepsilon > 0$, find a function $I_\varepsilon \in H^1(D)$ that minimizes the objective functional

$$E_\varepsilon(\varphi) \stackrel{\text{def}}{=} \frac{1}{2} \int_D \varepsilon |\nabla \varphi|^2 + |\varphi - I|^2 \, dx,$$

where the value of ε can be suitably adjusted to get the desired degree of smoothness. The minimizing function $I_\varepsilon \in H^1(D)$ is solution of the following variational equation:

$$\boxed{\forall \varphi \in H^1(D), \quad \int_D \varepsilon \, \nabla I_\varepsilon \cdot \nabla \varphi + (I_\varepsilon - I) \, \varphi \, dx = 0.} \tag{9.44}$$

The price to pay now is that the computation is made on the whole image.

This can be partly compensated by combining in a single operation or level of processing the smoothing of the image with the detection of the edges. To do that D. Mumford and J. Shah [**1**] in 1985 and [**2**] in 1989 introduced a new objective functional that received a lot of attention from the mathematical and engineering communities.

9.5.2 Objective Function of Mumford and Shah

Given a grey level *ideal* image in a fixed bounded open *frame* D we are looking for an open subset Ω of D such that its boundary reveals the edges of the two-dimensional objects contained in the image of Figure 1.9.

Definition 9.2.
Let D be a bounded open subset of $\mathbf{R}^N$ with Lipschitzian boundary.

(i) An *image* in the *frame* D is specified by a function $I \in L^2(D)$.

(ii) We say that $\{\Omega_j\}_{j\in J}$ is an *open partition* of D if $\{\Omega_j\}_{j\in J}$ is a family of disjoint connected open subsets of D such that

$$\mathrm{m}_N(\cup_{j\in J}\Omega_j) = \mathrm{m}_N(D) \quad \text{and} \quad \mathrm{m}_N(\partial \cup_{j\in J} \Omega_j) = 0,$$

where m_N is the N-dimensional Lebesgue measure. Denote by $\mathcal{P}(D)$ the family of all such open partitions of D. □

The idea behind the formulation of D. Mumford and J. Shah [**1**] in 1985 is to do a Tikhonov regularization $I_{\varepsilon,j} \in H^1(\Omega_j)$ of I on each element Ω_j of the partition and then minimize the sum of the local minimizations over Ω_j over all open partitions of the image D: find an open partition $P = \{\Omega_j\}_{j\in J}$ in $\mathcal{P}(D)$ solution of the minimization problem

$$\inf_{P\in\mathcal{P}(D)} \sum_{j\in J} \inf_{\varphi_j\in H^1(\Omega_j)} \int_{\Omega_j} \varepsilon|\nabla\varphi_j|^2 + |\varphi_j - I|^2\, dx \tag{9.45}$$

for some fixed constant $\varepsilon > 0$. Observe that without the condition $\mathrm{m}_N(\cup_{j\in J}\Omega_j) = \mathrm{m}_N(D)$ the empty set would be a solution of the problem.

The question of existence requires a more specific family of open partitions or a penalization term that controls the "length" of the interfaces in some sense:

$$\inf_{P\in\mathcal{P}(D)} \sum_{j\in J} \inf_{\varphi_j\in H^1(\Omega_j)} \int_{\Omega_j} \varepsilon|\nabla\varphi_j|^2 + |\varphi_j - I|^2\, dx + c\, H_{N-1}(\partial \cup_{j\in J} \Omega_j) \tag{9.46}$$

for some $c > 0$, where H_{N-1} is the Hausdorff $(N-1)$-dimensional measure. Equivalently, we are looking for

$$\Omega = \cup_{j\in J}\Omega_j \text{ open in } D \text{ such that } \chi_\Omega = \chi_D \text{ a.e. in } D$$

(χ_Ω and χ_D, characteristic functions) solution of the following minimization problem:

$$\inf_{\substack{\Omega \text{ open } \subset D\\ \chi_\Omega=\chi_D \text{ a.e.}}} \inf_{\varphi\in H^1(\Omega)} \int_\Omega \varepsilon|\nabla\varphi|^2 + |\varphi - I|^2\, dx + c\, H_{N-1}(\partial\Omega). \tag{9.47}$$

In general, the Hausdorff measure is not a lower semicontinuous functional. So either the $(N-1)$-Hausdorff measure H_{N-1} is relaxed to a lower semicontinuous notion of perimeter or the minimization problem is reformulated with respect to a more suitable family of open domains and/or a space of functions larger than $H^1(\Omega)$.

Another way of looking at the problem would be to minimize the number J of connected subsets of the open partition, but this seems more difficult to formalize.

9.5.3 Relaxation of the $(N-1)$-Hausdorff Measure

The choice of a relaxation of the $(N-1)$-Hausdorff measure H_{N-1} is critical. Here the finite perimeter of Caccioppoli (cf. section 6 in Chapter 5) reduces to the perimeter of D since the characteristic function of Ω is almost everywhere equal to the characteristic function of D. However, the relaxation of $H_{N-1}(\partial\Omega)$ to the $(N-1)$-dimensional *upper Minkowski content* by D. Bucur and J.-P. Zolésio [**8**] is much more interesting and in view of its associated compactness theorem (cf. section 12 in Chapter 7) it yields a unique solution to the problem.

9.5.4 Relaxation to BV-, H^s-, and SBV-Functions

The other avenue to explore is to reformulate the minimization problem with respect to a space of functions larger than $H^1(\Omega)$.

One possibility is to use functions with possible jumps, such as functions of bounded variations. In dimension 1, the BV-functions can be decomposed into an absolutely continuous function in $W^{1,1}(0,1)$ plus a jump part at a countable number of points of discontinuity. Such functions have their analogue in dimension N. With this in mind one could consider the penalized objective function

$$J_{\mathrm{BV}}(\varphi) \stackrel{\text{def}}{=} \int_D |\varphi - I|^2\, dx + \varepsilon \|\varphi\|^2_{\mathrm{BV}(D)}, \tag{9.48}$$

which is similar to the Tikhonov regularization when rewritten in the form

$$J_{H^1}(\varphi) \stackrel{\text{def}}{=} \int_D |\varphi - I|^2\, dx + \varepsilon \|\varphi\|^2_{H^1(D)} = \int_D (1+\varepsilon)|\varphi - I|^2\, dx + \varepsilon |\nabla\varphi|^2\, dx,$$

where we square the penalization term to make it differentiable. Unfortunately, the square of the BV-norm is not differentiable.

Going back to Figure 1.9, assume that the triangle T has an intensity of 1, the circle C has an intensity of 2/3, and the remaining part of the square S has an intensity of 1/3. The remainder of the image is set equal to 0. Then the image functional is precisely

$$I(x) = \chi_T(x) + \frac{2}{3}\chi_C + \frac{1}{3}\chi_{S\setminus(T\cup C)}.$$

We know from Theorem 6.9 in section 6.3 of Chapter 5 that the characteristic function of a Lipschitzian set is a BV-function. Therefore, if we minimize the objective functional (9.48) over all $\varphi \in \mathrm{BV}(D)$, we get exactly $\hat{\varphi} = I$. If we insist on formulating the minimization problem over a Hilbert space as for the Tikhonov regularization, the norm of the penalization term can be chosen in the space $H^s(D)$ for some $s \in (0,1]$:

$$J_{H^s}(\varphi) \stackrel{\text{def}}{=} \int_D |\varphi - I|^2\, dx + \varepsilon\, \|\varphi\|^2_{H^s(D)}.$$

For $0 < s < 1/2$, the norm is given by the expression

$$\|\varphi\|_{H^s(D)} \stackrel{\text{def}}{=} \int_D dx \int_D dy\, \frac{|\varphi(y) - \varphi(x)|^2}{|y-x|^{N+2s}}$$

and the relaxation is very close to the one in $\mathrm{BV}(D)$ since $\mathrm{BV}(D) \cap L^\infty(D) \subset H^s(D)$, $0 < s < 1/2$ (cf. Theorem 6.9 (ii) in Chapter 5).

However, as seducing as those relaxations can be, it is not clear that the objectives of the original formulation are preserved. Take the BV-formulation. It is implicitly assumed that there are clean jumps across the interfaces. This is true in dimension 1, but not in dimension strictly greater than 1 as explained in L. Ambrosio [**1**], where he points out that the distributional gradient of a BV-function which is a vector measure has three parts: an absolutely continuous part, a jump part, and a nasty Cantor part. The space $\mathrm{BV}(D)$ contains pathological functions of Cantor–Vitali type that are continuous with an approximate differential 0 almost everywhere, and this class of functions is dense in $L^2(D)$, making the infimum of

J_{BV} over $\mathrm{BV}(D)$ equal to zero without giving information about the segmentation. To get around this difficulty, he introduces the smaller space $\mathrm{SBV}(D)$ of functions whose distributional gradient does not have a Cantor part. He considers the mathematical framework for the minimization of the objective functional

$$J(\varphi, K) \stackrel{\text{def}}{=} \int_{D\setminus K} |\varphi - I|^2 + \varepsilon |\nabla \varphi|^2 \, dx + c\, H_{N-1}(K)$$

with respect to all closed sets $K \subset \Omega$ and $\varphi \in H^1(\Omega \setminus K)$, where $I \in L^\infty(D)$. This problem makes sense and has a solution in $\mathrm{SBV}(D)$ when K is replaced by the set S_φ of points where the function $\varphi \in \mathrm{SBV}(D)$ has a jump, that is,

$$J(\varphi) \stackrel{\text{def}}{=} \int_D |\varphi - I|^2 + \varepsilon |\nabla \varphi|^2 \, dx + c\, H_{N-1}(S_\varphi),$$

that turns out to be well-defined on $\mathrm{SBV}(D)$.

9.5.5 Cracked Sets and Density Perimeter

The alternate avenue to explore is to reformulate the minimization problem with respect to a more suitable family of open domains (cf. section 15 of Chapter 7).

An additional reason to do that would be to remove the term on the length of the interface. When a penalization on the length of the segmentation is included, long slender objects are not "seen" by the numerical algorithms since they have too large a perimeter. In order to retain a segmentation of piecewise H^1-functions without a perimeter term, M. C. Delfour and J.-P. Zolésio [**38**] introduced the family of *cracked sets* (see Figure 1.11) that yields a compactness theorem in the $W^{1,p}$-topology associated with the *oriented distance function* (cf. Chapter 7).

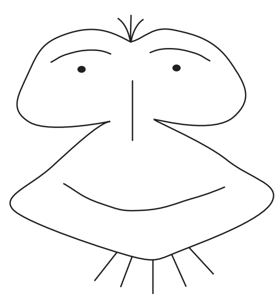

Figure 1.11. *Example of a two-dimensional strongly cracked set.*

The originality of this approach is that it does not require a penalization term on the *length of the segmentation* and that, within the set of solutions, there exists one with minimum *density perimeter* as defined by D. Bucur and J.-P. Zolésio [**8**].

Theorem 9.2. *Let D be a bounded open subset of $\mathbf{R}^N$ and $\alpha > 0$ and $h > 0$ be real numbers.*[3] *Consider the families*

$$\mathcal{F}(D,h,\alpha) \stackrel{\text{def}}{=} \left\{ \Omega \subset \overline{D} : \begin{array}{c} \Gamma \neq \varnothing \text{ and } \forall x \in \Gamma,\, \exists d,\, |d| = 1, \\ \text{such that } \inf\limits_{0<t<h} \dfrac{d_\Gamma(x+td)}{t} \geq \alpha \end{array} \right\}, \tag{9.49}$$

$$C_b^{h,\alpha}(D) \stackrel{\text{def}}{=} \{ b_\Omega : \Omega \in \mathcal{F}(D,h,\alpha) \},$$

$$\mathcal{F}_s(D,h,\alpha) \stackrel{\text{def}}{=} \left\{ \Omega \subset \overline{D} : \begin{array}{c} \Gamma \neq \varnothing \text{ and } \forall x \in \Gamma,\, \exists d,\, |d| = 1, \\ \text{such that } \inf\limits_{0<|t|<h} \dfrac{d_\Gamma(x+td)}{|t|} \geq \alpha \end{array} \right\}, \tag{9.50}$$

$$(C_b^{h,\alpha})_s(D) \stackrel{\text{def}}{=} \{ b_\Omega : \Omega \in \mathcal{F}_s(D,h,\alpha) \}.$$

Then $C_b^{h,\alpha}(D)$ and $(C_b^{h,\alpha})_s(D)$ are compact in $W^{1,p}(D)$, $1 \leq p < \infty$, where d_Γ is the distance function to the boundary of Ω.

In view of the above compactness, it can be shown that there exists a solution to the following minimization problem:

$$\inf_{\substack{\Omega \in \mathcal{F}(D,h,\alpha) \\ \Omega \text{ open } \subset D,\, \mathrm{m}_N(\Omega) = \mathrm{m}_N(D)}} \inf_{\varphi \in H^1(\Omega)} \int_\Omega \varepsilon\, |\nabla \varphi|^2 + |\varphi - f|^2\, dx. \tag{9.51}$$

Cracked sets form a very rich family of sets with a huge potential that is not yet fully exploited in the image segmentation problem. Indeed, they can be used not only to partition the frame of an image but also to detect isolated cracks and points provided an objective functional sharper than the one of Mumford and Shah is used. For instance, in view of the connection between image segmentation and *fracture theory* hinted at in J. Blat and J.-M. Morel [**1**], the theory may have potential applications in problems related to the detection of *fractures or cracks* or *fracture branching and segmentation* in geomaterials (cf. K. B. Broberg [**1**]), but this is way beyond the scope of this book. Some initial considerations about the numerical approximation of cracked sets can be found in M. C. Delfour and J.-P. Zolésio [**41**].

10 Shapes and Geometries: Background and Perspectives

10.1 Parametrize Geometries by Functions or Functions by Geometries?

In the buckling of the column and in the design of the thermal diffuser and the thermal radiator the geometry was a volume of revolution generated by rotation around the z-axis of the hypograph between the axis and the graph of the function

[3] In view of the fact that the distance function d_Γ is Lipschitzian with constant 1, we necessarily have $0 < \alpha \leq 1$.

defined in a variable interval $[0, L]$. They are examples of *sets parametrized by functions* or scalars. A well-documented weakness of this approach is that numerical computations of optimal shapes often yield boundary oscillations. This is typical of the parametrization of the boundary by a function. It provides a good control of the displacement in the direction normal to the graph but little control in the *tangential direction.*

Extending this approach to sets locally defined by several graphs becomes tricky. For instance, it is not possible to represent a sphere or a torus from a bi-Lipschitzian mapping from some domain in the plane unless we replace the plane by a surface with facets as illustrated in Figure 1.12 for a ball (cf. M. C. Delfour [**8**] for the construction of a surface with facets from a $C^{1,1}$-surface). Moreover, the surface with facets of Figure 1.12 is closely associated with the ball and would not be appropriate for a torus.

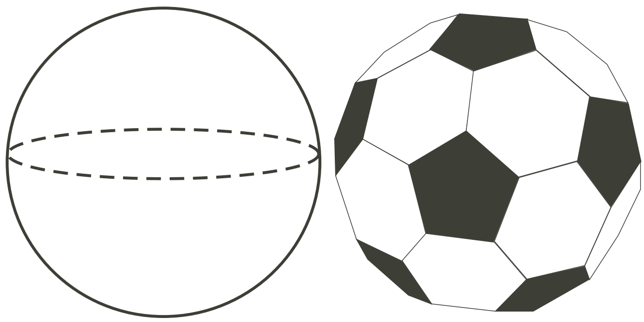

Figure 1.12. *Example of a surface with facets associated with a ball.*

Another family of function-parametrized sets is the family of images of a fixed set by a family of homeomorphisms or diffeomorphisms. This can be advantageous in problems where it is desirable to work with a fixed mesh of the original domain, thus avoiding remeshing the domain at each step of the optimization process. Function-parametrized sets are also used in the *identification of objects* (cf., for instance, the school of R. Azencott, A. Trouvé, and L. Younes). This approach can be traced back to R. Courant and D. Hilbert [**1**] in 1953 and the construction of Courant metrics and complete metric spaces of images of a fixed closed or crack-free open set by A. M. Micheletti [**1**] in 1972. This approach is less interesting in optimization problems where the topology of the set is part of the unknown features of the set. Indeed, the images of a fixed set by an homeomorphism cannot change the topology of the fixed set. For instance, it cannot create holes that are not present in the fixed set.

The example of the distribution of two materials in a fixed domain is generic of problems where the topology of the set is part of the unknowns. The variable chosen to represent the set is the characteristic function χ_A, a *function parametrized by the set A.* As such, arbitrary geometries that are only measurable can be used as independent variables. The topology (for instance, the number of holes) is not fixed a priori and extremely complex design can be obtained. Moreover, the underlying

L^p-spaces where the characteristic functions naturally live induce a metric

$$\rho([A_1],[A_2]) = \|\chi_{A_2} - \chi_{A_2}\|_{L^p(D)}$$

on the family of equivalence classes of Lebesgue measurable subsets of a bounded measurable holdall D. This type of topology has been used in the proof of the Maximum Principle by Pontryagin and later by Ekeland.

Other examples of set-parametrized functions include the uniformly Lipschitzian *distance function* d_A to a set A. The family of such functions for subsets of a holdall $\overline{D}$, D bounded open, can be considered as a family of functions in $C^0(\overline{D})$ or $W^{1,p}(D)$. For instance, the metric

$$\rho([A_1],[A_2]) = \|d_{A_2} - d_{A_2}\|_{C^0(\overline{D})}$$

($[A]$ is the equivalence class of all subsets of $\overline{D}$ with the same closure) coincides with the *écart mutuel* between two sets introduced by D. Pompéiu [**1**] in his thesis presented in Paris in March 1905 that was studied in more detail by F. Hausdorff [**2**, "Quellenangaben," p. 280, and Chap. VIII, sect. 6] in 1914. Unfortunately that metric does not preserve the volume functional, but the norm can be changed to obtain the new metric

$$\rho([A_1],[A_2]) = \|d_{A_2} - d_{A_2}\|_{W^{1,p}(D)}$$

that preserves the volume functional.

Another example of a set-parametrized function is the *support function*

$$\sigma_A(x) \overset{\text{def}}{=} \sup_{a\in A} a\cdot x$$

of convex analysis. Here the equivalence classes are

$$[A] \overset{\text{def}}{=} \left\{B \subset \overline{D} : \overline{\text{co}}\, B = \overline{\text{co}}\, A\right\},$$

that is, sets with the same closed convex hull. The associated metric is

$$\rho([A_1],[A_2]) = \|\sigma_{A_2} - \sigma_{A_2}\|_{C^0(\overline{B})},$$

where B is the unit open ball in $\mathbf{R}^N$ (cf. M. C. Delfour and J.-P. Zolésio [**17**]).

The general pattern that emerges from the last three examples is the following one. Choose a fixed holdall D and let A denote a variable set in D. Select a family of set parametrized functions f_A and let

$$[A] \overset{\text{def}}{=} \left\{B \subset \mathbf{R}^N : f_A = f_B\right\}$$

be the associate equivalence class. Finally, assume that the family of functions $\{f_A\}$ is contained in a Banach space $\mathcal{B}(D)$. Choose as a metric

$$\rho([A_1],[A_2]) = \|f_{A_2} - f_{A_2}\|_{\mathcal{B}(D)}.$$

This approach will be detailed in Chapters 5, 6, and 7 for the *characteristic function* χ_A, the *distance function* d_A, and the *oriented distance function* b_A, but many other set-parametrized functions can be used according to the requirements of the problem at hand.

10.2 Shape Analysis in Mechanics and Mathematics

The terminology *shape analysis* has been introduced independently in at least two different contexts: *continuum mechanics* and *mathematical theory of partial differential equations* (cf., for instance, E. J. HAUG and J. CÉA [**1**]).

In continuum mechanics, shape analysis encompasses contributions to structural mechanics of elastic bodies such as beams, plates, shells, arches, and trusses. In such problems, the objective is to optimize the *compliance*, e.g., the work of the applied loadings, by choosing the design parameters of the structure. Many of the early contributions are in dimension 2 and complete analytical solutions are often provided. Yet, it is not always easy to distinguish between a *shape optimization problem*, such as the shape of a two-dimensional plate, and a *distributed parameter problem*, such as the optimal thickness of that plate. The thickness that is often considered as a *shape parameter* is really a distributed parameter over the two-dimensional domain, which also specifies the plate. When the thickness goes to zero in some parts of the plate, holes are created and induce changes in the topology and the shape of the associated two-dimensional domain. This naturally leads to *topological optimization*, which deals with the connectivity of a domain, the number of holes, the fractal dimension of the boundary, and ultimately the appearance of a microstructure. This was exemplified by K.-T. CHENG and N. OLHOFF [**2**]'s celebrated optimization of the compliance of the circular plate with respect to its thickness under prescribed loading and a constraint on the volume of material. Such questions have received a lot of attention in specific cases and have been analyzed by homogenization methods or Γ-convergence (cf., for instance, F. MURAT and L. TARTAR [**1**], the conference proceedings edited by M. BENDSØE and C. A. MOTA SOARES [**1**], the book by M. BENDSØE [**1**], and the book edited by A. CHERKAEV and R. KOHN [**1**] that contains a selection of translations of key papers written in French or Russian). However, many fundamental questions still remain open. For instance, how does such an analysis affect the validity of the underlying mechanical or physical models?

For convenience we shall refer to this viewpoint as the *compliance analysis* that generally involves the extremum of the minimum of an energy or work functional with respect to some design parameters.

In the mathematical theory of partial differential equations, the analysis dealt with the sensitivity of the solution of boundary value problems with respect to the shape of the geometric domain on which the partial differential equation is defined. This was done for different applications, including free boundary problems, noncylindric problems,[4] and shape identification problems. This *shape sensitivity analysis* was simultaneously developed for the solution of the partial differential equation and for shape functionals depending on their solution.

In that context, the compliance analysis becomes a special case that, quite remarkably, does not require the shape sensitivity analysis of the solution of the associated partial differential equation (cf. section 2 of Chapter 10). This important

[4]A cylindric problem is a partial differential equation where the geometric domain is fixed and independent of the time variable. A noncylindric problem is a partial differential equation where the geometric domain changes with time.

simplification arises from the fact that the compliance is the minimum of an energy or work functional. An historical example that also benefits from this property is the shape derivative of the first eigenvalue of the plate studied at the beginning of the 20th century by J. HADAMARD [**1**]. As for the compliance, the shape sensitivity analysis of the first eigenfunction is not required even when the first eigenvalue is repeated. This follows from the fact that the first eigenvalue can be expressed as a minimum through Rayleigh's quotient or G. AUCHMUTY [**1**]'s dual variational principle.

Shape sensitivity analysis deals with a larger class of shape functionals (e.g., minimal drag, noise reduction) and partial differential equations (e.g., the wave equation, viscous or non-Newtonian fluids), where variational energy functionals are usually not available and for which the compliance analysis is no longer applicable. Yet, the shape sensitivity analysis of the solution of the partial differential equation can again be avoided by incorporating the partial differential equation into a Hamiltonian or Lagrangian formulation. As in *control theory*, this yields a partial differential equation for the *adjoint state* that is coupled with the initial partial differential equation or *state equation*. A precise mathematical justification of this approach can be given when the Lagrangian has saddle points and the shape derivative can be obtained from theorems on the differentiability of saddle points with respect to a parameter even when the saddle point solution is not unique (cf. section 5 in Chapter 10).

The shape sensitivity analysis through a family of diffeomorphisms which preserve the smoothness of the images of a fixed domain is primarily a local analysis. It is used to establish continuity, to define derivatives, or to optimize in a narrow class of domains with fixed regularities and topologies: it cannot create holes or singularities that were not present in the initial domain. Consider the problem of finding the best location and shape of a hole of given volume in a homogeneous elastic plate to optimize the compliance or some other criterion under a given loading. The expectation is that the presence of the hole would improve the compliance over a homogeneous plate without holes. This problem has its analogue in control theory, for instance the optimal placement of sensors and actuators for the control and stabilization of large flexible space structures or flexible arms of robots. Another important example is the localization of sensors and actuators to achieve noise reduction in structures. The optimal placement is usually an integral part of the control synthesis.

In the early 1970s J. CÉA, A. J. GIOAN, and J. MICHEL [**1**] proposed to introduce relaxed problems in which the optimal domain (here an optimal hole, the optimal location of the support of the optimal control) was systematically replaced by a density function ranging between zero and one and that would hopefully be a characteristic function (a bang bang control) in the optimal regime. Furthermore, in order to work on a fixed domain D without holes, they replaced the holes by a very weak elastic material. This changed the original topological optimization problem into the identification of distributed coefficients over the fixed domain D. Equivalently, this identification problem reduces to finding the optimal distribution of two materials for a transmission equation in D under a volume constraint on one of the two materials. Under appropriate conditions, the solution of this problem is a characteristic function, even when the space of distributed parameters is relaxed to the closed convex hull of the set of characteristic functions (cf. Chapter 5). This general technique can also be used to study the continuity of the solution of the homogeneous Dirichlet boundary value problems with respect to the domain Ω, by introducing a transmission problem over a fixed domain D and letting the distributed coefficient go to infinity in the complement of Ω with respect to D. Of course, the coefficient over the complement of Ω could be replaced by a Lagrange multiplier, thus providing a formulation over the fixed domain D. This is one of the many ways to make the domain *fictitious* and avoid dealing directly with the geometry (cf. R. Glowinski, T.-W. Pan, and J. Périaux [**1**]).

10.3 Characteristic Functions: Surface Measure and Geometric Measure Theory

The main advantage of the relaxation to a characteristic function in the formulation of problems involving a domain integral and/or a volume constraint on Lebesgue measurable sets is that the unknown set is completely described by a single function instead of a family of local diffeomorphisms.

Yet, such measurable sets can be quite unstructured. For example, in problems involving a surface tension on the free boundary of a fluid or on the interface between two fluids, the domain must have a locally finite boundary measure. Another example is when the objective functional is a function of the normal derivative of the state variable along the boundary (e.g., a flow or a thermal power flux through the boundary). In both cases the use of characteristic functions is limited by the fact that they are not differentiable on the boundary of the set and cannot be readily used to describe very smooth domains.

Fortunately, there is enough latitude to make sense of a locally finite boundary measure for sets for which the characteristic function is of bounded variation; that is, its gradient is a vector of bounded measures. Such sets are known as *Caccioppoli* or *finite perimeter sets*. They were a key ingredient in the contribution of E. De Giorgi to the *theory of minimal surfaces* in the 1950s, since the norm of that vector measure (which is the total variation of the characteristic function) turns out to be a relaxation of the boundary measure or "perimeter" of the set. To get a compactness result for a family of Caccioppoli sets it is sufficient to put a uniform bound on their perimeters. This field of activities is known as *geometric measure theory*, and its tools have been used very successfully in the theory of free and moving boundary problems. Even though it came rather late to *shape analysis*, this material is both important and fundamental.

10.4 Distance Functions: Smoothness, Normal, and Curvatures

Another set-parametrized function that plays a role similar to the characteristic function is the *distance function* between closed subsets of a fixed holdall[5] D of $\mathbf{R}^N$. The "*écart mutuel*" between two sets was introduced by D. Pompéiu [**1**] in

[5]This set D will play several roles in this book. It is the *universe* in which the variable subsets live. It will often be referred to as the underlying *holdall*. In other circumstances it will have a purely technical role, much like the control volume in fluid mechanics.

his thesis presented in Paris in March 1905. This is the first example of a metric between two sets in the literature. It was studied in more detail by F. HAUSDORFF [**2**, "Quellenangaben," p. 280, and Chap. VIII, sect. 6] in 1914. The Pompéiu–Hausdorff metric between two sets corresponds to the uniform norm of the difference of their respective distance functions. For most applications it is not a very interesting topology since the volume functional is not continuous in that topology. The continuity of the volume functional can be recovered by replacing the uniform norm by the $W^{1,p}$-norm since the characteristic function of the closure of the set can be expressed in terms of the gradient of the distance function.

In 1951, H. FEDERER [**1**] introduced the family of *sets of positive reach* and gave a first access to curvatures from the distance function in much the same spirit as the perimeter from the characteristic function. They are sets for which the projection onto the set is unique in a neighborhood of the set. Since the projection can be expressed in terms of the gradient of the square of the distance function, this is equivalent to requiring that the square of the distance function be $C^{1,1}$ in a neighborhood of the set. Since the gradient of the distance function has a jump discontinuity at the boundary of the set, he managed to recover from the distance function the *curvature measures* of the boundary and made sense of the classical Steiner formula for sets with positive reach.

The jump discontinuity of the characteristic function across the boundary can be bypassed by going to the *oriented distance function*. For smooth domains this function is quite remarkable since it inherits the same degree of smoothness in a neighborhood of the boundary as the boundary itself. Then the gradient, the Hessian matrix, and the higher-order derivatives in the neighborhood of the boundary can be used to characterize and compute normals, curvatures, and their derivatives along the boundary. This correspondence remains true for domains of class $C^{1,1}$. The gradient of the oriented distance function coincides with the outward unit normal on the boundary. This implicit orientation is at the origin of the terminology *oriented distance function* that was introduced by M. C. DELFOUR and J.-P. ZOLÉSIO [**17**] in 1994 to distinguish it from the *algebraic distance function* to a submanifold that can be defined only in terms of some discriminating criterion that distinguishes between what is *above* and what is *below* the submanifold. This is not always possible, while the oriented distance function to a set is always well-defined. The restriction of the Hessian matrix of the oriented distance function to the boundary coincides with the second fundamental form of differential geometry. Its eigenvalues are zero and the principal curvatures of the boundary.

A nice relaxation of the curvatures is obtained by considering sets for which the elements of the Hessian matrix of the oriented distance function are bounded measures. They are called *sets of locally bounded curvature* (cf. M. C. DELFOUR and J.-P. ZOLÉSIO [**17**, **32**]). To get a compactness result, it is sufficient to put a uniform bound on the total variation of the gradient (the elements of the Hessian matrix are bounded measures).

The oriented distance function also provides a framework to compare the smoothness of sets ranging from arbitrary sets with a nonempty boundary to sets of class C^∞ via the smoothness of the oriented distance function in a neighborhood of the boundary of the set. As a result Sobolev spaces can be used to introduce the notion of a *Sobolev domain*, which becomes intertwined with the classical notion of C^k-domains.

10.5 Shape Optimization: Compliance Analysis and Sensitivity Analysis

As in the vector space case, optimization and control problems with respect to geometry are of various degrees of difficulty. When the objective functional does not depend on the solution of a state equation or variational inequality defined on the variable domain, it is sufficient to invoke compactness and continuity arguments. In special cases such as the optimization of the compliance or of the first eigenvalue, the problem can be transformed into the optimization of an objective functional that is itself the minimum of some appropriate functional defined on a fixed function space. There a direct study of the dependence of the solution of the state equation with respect to the underlying domain can be bypassed. In the general case, a state equation constraint has to be handled very carefully from both the mathematical and the application viewpoints. When the analysis can be restricted to families of Lipschitzian or convex domains, it is usually possible to give a meaning and prove the continuity of the solution of the state equation with respect to the underlying varying domain.

When families of arbitrary bounded open domains are considered, new phenomena can occur. As the domains converge in some sense, the corresponding solutions may only converge in a weak sense to the solution of a different type of state equation over the limit domain: boundary conditions may no longer be satisfied, strange terms[6] may occur on the right-hand side of the equation, etc. In such cases, it is often a matter of modeling of the physical or technological phenomenon. Is it natural to accept a generalized solution or a relaxed formulation of the state equation or should the family of domains be sufficiently restricted to preserve the form of the original state equation and maintain the continuity of the solution with respect to the domain? The most suitable relaxation is not necessarily the most general mathematical relaxation. It must always be compatible with the physical or technological problem at hand. Too general a relaxation of the problem or too restrictive conditions on the varying domains can yield completely unsatisfactory solutions however nice the underlying mathematics are. A good balance of mathematical, physical, and engineering intuitions is essential.

The study of the shape continuity of the solution of the partial differential equation is also of independent mathematical interest. In the literature, this issue has been addressed in various ways. Some authors simply introduce a *stability assumption* that essentially says that the limiting domain is such that continuity occurs. Others introduce a set of more technical assumptions that correspond to a minimal set of conditions to make the crucial steps of the proof of the continuity work. On the constructive side the really challenging issue is to characterize the families of domains for which the continuity holds.

[6]Cf., for instance, D. Cioranescu and F. Murat [**1**, **2**].

Many authors have constructed compact families for that purpose. For instance the *Courant metric topology* was used by A. M. MICHELETTI [**1**] in 1972 for C^k-domains, the *uniform cone condition* by D. CHENAIS [**1**] in 1973 for uniformly Lipschitzian domains, and other metric topologies by F. MURAT and J. SIMON [**1**] in 1976 for Lipschitzian domains.

More general capacity conditions were introduced after 1994 by D. BUCUR and J.-P. ZOLÉSIO [**5**] in order to obtain compact subfamilies of domains with respect to the complementary Hausdorff topology and control the curvature of the boundaries of the domains. In dimension 2 they recover the nice result of V. ŠVERÁK [**2**] in 1993, which involves a uniform bound on the number of connected components of the complement of the sets. Intuitively the capacity conditions are such that, locally, the complement of the domains in the chosen family has "enough capacity" to preserve the homogeneous Dirichlet boundary condition in the limit. Yet, the capacity conditions might not be easy to use in a specific example. So D. BUCUR and J.-P. ZOLÉSIO [**1**, **5**] introduced a simpler geometric constraint, called the *flat cone condition*, under which the continuity and compactness results still hold. This generalizes the *uniform cone property* to a much larger class of open domains. All this will be further generalized in section 9 of Chapter 8.

10.6 Shape Derivatives

For functionals defined on a family of domains, it is important to distinguish between a function with values in a topological vector space over a fixed domain and a function such as the solution of a partial differential equation that lives in a Sobolev space defined on the varying domain. In the latter case special techniques have to be used to transport the solution onto a fixed domain or to first embed the varying domains into a fixed *holdall* D, extend the solutions to D, and enlarge the Sobolev space to a large enough space of functions defined over the fixed holdall D. In both cases the function is defined over a family of domains or sets belonging to some *shape space* which is generally nonlinear and nonconvex. Thus, defining derivatives on those spaces is more related to defining derivatives on differentiable manifolds than in vector spaces.

It is perhaps for groups of diffeomorphisms that the most complete theory of shape derivatives is available. This material will be covered in detail in Chapters 2, 3, and 9, but it is useful to briefly introduce the main ideas and definitions here. Given a Banach space Θ of mappings from a fixed open holdall $D \subset \mathbf{R}^N$ into $\mathbf{R}^N$, first consider the group of diffeomorphisms

$$\mathcal{F}(\Theta) \overset{\text{def}}{=} \left\{F : D \to D \,:\, F - I \in \Theta \text{ and } F^{-1} - I \in \Theta\right\}.$$

Then consider the images of a fixed domain Ω_0 in D by $\mathcal{F}(\Theta)$:

$$\mathcal{X}(\Omega_0) \overset{\text{def}}{=} \{F(\Omega_0) \,:\, \forall F \in \mathcal{F}(\Theta)\}.$$

They can be identified with the quotient group

$$\mathcal{F}(\Theta)/\mathcal{G}(\Omega_0), \quad \mathcal{G}(\Omega_0) \overset{\text{def}}{=} \{F \in \mathcal{F}(\Theta) \,:\, F(\Omega_0) = \Omega_0\}$$

of diffeomorphisms of D. For the specific choices of D, Θ, and Ω_0 that are of interest, this quotient group can be endowed with the so-called *Courant metric* to make it a complete metric space. This metric space is neither linear nor convex.

For unconstrained domains ($D = \mathbf{R}^N$) and "sufficiently small" elements $\theta \in \Theta$, transformations of the form $F = I + \theta$ belong to $\mathcal{F}(\Theta)$ and hence *perturbations of the identity* (i.e., of Ω_0) can be chosen in the vector space Θ. This makes it possible to define directional derivatives and speak of Gateaux and Fréchet differentiability with respect to Θ as in the classical case of functions defined on vector spaces. However, this approach does not extend to submanifolds D of $\mathbf{R}^N$ or to domains that are constrained in one way or another (e.g., constant volume, perimeter, etc.).

In the unconstrained case, we get an infinite-dimensional differentiable manifold structure on the quotient group $\mathcal{F}(\Theta)/\mathcal{G}(\Omega_0)$ and the tangent space to $\mathcal{F}(\Theta)$ is the whole linear space Θ. As a result ideas and techniques from *differential geometry* are readily applicable. For instance, the continuity in I (i.e., at Ω_0) can be characterized by the continuity along one-dimensional flows of velocity fields in $\mathcal{F}(\Theta)$ through the point I (that is, $I(\Omega_0) = \Omega_0$). This suggests defining a notion of directional derivative along one-dimensional flows associated with "velocity fields" V which keep the flows inside D. When D is a smooth submanifold of $\mathbf{R}^N$ such velocities are tangent to D and the set of all such velocities has a vector space structure. This key property generalizes to other types of holdalls. So it is possible to define directional derivatives and speak of shape gradient and shape Hessian with respect to the associated vector space of velocities. This second approach has been known in the literature as the *velocity method.* Another very nice property of that method is that perturbations of the identity, as previously defined in the unconstrained case, can be recovered by a special choice of velocity field, thus creating a certain unity in the methodology.

The concept of *topological derivative* was introduced by J. Sokołowski and A. Żochowski [**1**] for problems where the knowledge of the optimal topology of the domain is important. It gives some sensitivity of a shape functional to the presence of a small hole at a point of the domain as the size of that hole goes to zero. For a domain integral, that is, the integral of a locally Lebesgue integrable function, this derivative is the negative of the classical *set-derivative* which is equal to the function at almost every point. This is a direct consequence of the Lebesgue differentiation theorem. So, at least in its simplest form, this approach aims at extending the classical concept of *set-differentiation* as the inverse of integration over sets.

This type of derivative becomes more intricate as we look at the sensitivity of the solution of a boundary value problem or at shape functionals which are functions of that solution. For functions that are the integrals of an integrable function with respect to a Radon measure, the Lebesgue differentiation extends in the form of the Lebesgue–Besicovitch differentiation theorem, which says that the set-derivative of that integral is again equal to the function almost everywhere with respect to the Radon measure. So one could try to determine the class of shape functionals that can be expressed as an integral with respect to some Radon measure. Since this material requires extensive technical results to fully appreciate its impact, we did not include a chapter on this topic in the book, but the approach is an important addition to the global arsenal! Fortunately, the theoretical and numerical work is well documented in the literature and books by J. Sokołowski and his coauthors should be available in the near future.

10.7 Shape Calculus and Tangential Differential Calculus

The velocity method has been used in various applications and contexts, and a very complete *shape calculus* is now available (for instance, the reader is referred to the book of J. Sokołowski and J.-P. Zolésio [**9**]). In the computations of derivatives, and especially of second-order derivatives, an intensive use is made of the *tangential differential calculus*. In order to avoid parametrizations and local bases, tangential derivatives are defined through extensions of functions from the boundary to some small Euclidean neighborhood. Their importance should not be underestimated from both the theoretical and the computational points of view. The use of an intrinsic tangential gradient, divergence, or Laplacian can considerably simplify the computation and the final form of the expressions making more apparent their fine structure. Too many computations using local coordinates, Christoffel symbols, or intricate parametrizations are often difficult to decipher or to use effectively.

This book provides the latest and most important developments of that calculus. They arises from the systematic use of the oriented distance function in the theory of thin and asymptotic shells.[7] In that context, it has been realized that extending functions defined on the boundary Γ of a domain Ω by composition with the projection onto Γ results in sweeping simplifications in the tangential calculus. This is due to the fact that the projection can be expressed in terms of the gradient of the oriented distance function. Computing derivatives on Γ becomes as easy as computing derivatives in the Euclidean space. Curvature terms, when they occur, appear in the right place and in the right form through the Hessian matrix of the oriented distance function that coincides with the second fundamental form of Γ. Chapter 9 will provide a self-contained introduction to these techniques and show how the combined strengths of the shape calculus and the tangential differential calculus considerably simplify computations and expand our capability to tackle highly complex and challenging problems.

10.8 Shape Analysis in This Book

Problems in which the design, control, or optimization variable is no longer a vector of parameters or functions, but is the shape of a geometric domain, a set, or even a "fuzzy entity," cover a much broader range of applications than those for which the compliance or the shape sensitivity analysis have been used. Yet their analysis makes use of common mathematical techniques: partial differential equations, functional analysis, geometry, modern optimization and control theories, finite element analysis, large scale constrained numerical optimization, etc.

In this book the terminology *shape* will be used for domains ranging from unstructured sets to C^∞-domains. Relaxations of their geometric characteristics such as the volume, perimeter, connectivity, curvatures, and their derivatives will

[7] Cf. M. C. Delfour and J.-P. Zolésio [**19**, **20**, **25**] and M. C. Delfour [**3**, **4**, **6**, **7**].

be considered. Shape spaces (often metric and complete) corresponding to different levels of smoothness or degrees of relaxation of the geometry will be systematically constructed. For smooth domains we will emphasize the use of groups of diffeomorphisms endowed with the Courant metric. For more general domains we will use a generic construction based on the use of set-parametrized functions. The characteristic function will be associated with metric spaces of equivalence classes of Lebesgue measurable sets. The distance function will be associated not only with the uniform Hausdorff metric topology in the space of continuous functions but also with $W^{1,p}$-topologies for which the characteristic function and hence the volume function are continuous. The oriented distance function will be used in the same way to generate new metric topologies. In each case, the metric is constructed from the norm of one of the set-parametrized functions in an appropriate function space. The construction is generic and applies to other choices of set-parametrized functions and function spaces. For instance, the support function of convex analysis can be used to generate a complete metric topology on equivalence classes of sets with the same closed convex hull (cf. M. C. Delfour and J.-P. Zolésio [**17**]).

The nice property of the characteristic and distance functions over classical local diffeomorphisms is that the set is globally described in terms of the analytical properties of a single function. For instance, the gradient of the characteristic function yields a relaxed definition of the perimeter, and the Hessian of the distance function yields the boundary measure and curvature terms. But the characteristic function and the gradient of the distance function are both discontinuous at the boundary of the set. This seriously limits their use in the description of smooth domains.

In contrast, the oriented distance function can describe a broad spectrum of sets ranging from arbitrary sets with nonempty boundary to open C^∞-domains according to its degree of smoothness in a neighborhood of the boundary of the set. It readily combines the advantages of local diffeomorphisms and the characteristic and distance functions that are readily obtained from it. This provides, as in functional analysis, a common framework for the classification and comparison of domains according to their relative degree or lack of smoothness. As in geometric measure theory, compact families of sets will be introduced based on the degree of differentiability. One interesting family is made up of the sets with *locally bounded curvature*, which provide a sufficient degree of relaxation for most applications and for which nice compactness theorems are available (cf. sections 5 and 11 in Chapter 7). This family includes Federer's sets of positive reach and hence closed convex and semiconvex sets.

11 Shapes and Geometries: Second Edition

Except for the last two chapters, this second edition is almost a new book. Chapters 2, 3, and 4 give the classical description of sets and domains from the point of view of *differential geometry*. Special attention is paid to domains that verify some segment properties and have a local epigraph representation and to domains that are the image of a fixed set by a family of diffeomorphisms of the Euclidean space.

Chapters 5, 6, and 7 give a *function analytic* description of sets and domains via the set-parametrized characteristic functions, distance functions, and oriented distance functions. It emphasizes the fact that we are now dealing with equivalence classes of sets that may or may not have an invariant open or closed set representative in the class. In particular, they include Lebesgue measurable sets and Federer's sets of positive reach. Many of the classical properties of sets can be recovered from the regularity or function analytic properties of those functions. We concentrate on the basic properties, the construction of spaces of domains, metrics, and topologies, and the characterization of compact families.

Chapter 8 deals with problem formulations and *shape continuity and optimization* for some generic examples.

Chapters 9 and 10 are devoted to a modern version of the *shape calculus*, an introduction to the *tangential differential calculus*, and the shape derivatives under a state equation constraint.

11.1 Geometries Parametrized by Functions

Chapter 2 gives several classical characterizations and properties of sets or domains from the point of view of *differential geometry*: sets *locally described* in a neighborhood of each point of their boundary by an homeomorphism or a diffeomorphism (e.g., C^k or Hölderian diffeomorphisms), by the epigraph of a C^0 function (e.g., Lipschitzian or Hölderian domains), or by a geometric property (e.g., segment, cone, cusp), or sets *globally described* by the level sets of a C^1-function.

The sections on sets that are locally the epigraph of a C^0 function and sets having one of the segment properties have been completely reorganized, rewritten, and enriched with new results and older ones that are difficult to find in the literature other than in the form of *folk theorems*[8] or theorems without satisfying proofs. Special attention is given to the *uniform fat segment property* that was introduced in the first edition of the book under the name *uniform cusp property.* Several equivalent properties are given for domains that satisfy a *uniform segment property* and the stronger *uniform fat segment property* that can be expressed in terms of a *dominating function* and its modulus of continuity. All this is specialized to domains verifying a uniform cone or cusp property.

Chapter 3 adopts another point of view by considering families of sets that are the images of a fixed subset of $\mathbf{R}^N$ by some family of transformations of $\mathbf{R}^N$. The structure and the topology of the images can then be specified via the natural algebraic and topological structures of transformations or equivalence classes of transformations for which the full power of function analytic methods is available. In 1972, A. M. Micheletti [**1**] introduced the so-called *Courant metric* and gave what seems to be the first construction of a complete metric topology on the images

[8]As the term is understood by mathematicians, folk mathematics or mathematical folklore means theorems, definitions, proofs, or mathematical facts or techniques that are found by investigation and may circulate among mathematicians by word-of-mouth but have not appeared in print, either in books or in scholarly journals. Knowledge of folklore is the coin of the realm of academic mathematics, showing the relative insight of investigators (http://en.wikipedia.org/wiki/Mathematical_folklore).

of a fixed set. It is the nonlinear and nonconvex character of such *shape spaces* that will make the differential calculus and the analysis of shape optimization problems more challenging than their counterparts in topological vector spaces. The construction of A. M. MICHELETTI [**1**] is generic and readily extends to many families of transformations of $\mathbf{R}^{\mathrm{N}}$.

The new version of Chapter 3 considerably expands the material and ideas of the first edition by extracting the fundamental assumptions behind the generic framework of A. M. Micheletti that leads to the *Courant metrics* on the quotient space of families of transformations by subgroup of isometries such as identities, rotations, translations, or flips. Constructions are given for a large spectrum of transformations of the Euclidean space and for arbitrary closed subgroups. New complete metrics on the whole spaces of homeomorphisms and C^k-diffeomorphisms are also introduced. They extend classical results for transformations of compact manifolds to general unbounded closed sets and open sets that are crack-free. This material is central in classical mechanics and physics and in modern applications such as imaging and detection.

The former Chapter 7 on *transformations versus flows of velocities* has been moved right after the Courant metrics as Chapter 4 and considerably expanded. It now specializes the results of Chapter 3 to spaces of transformations that are generated by the flow of a *velocity field* over a generic time interval. The main motivation is to introduce a notion of semiderivatives as well as a tractable criterion for continuity with respect to Courant metrics. The velocity point of view was also adopted by R. Azencott and A. Trouvé starting in 1994 to construct complete metrics and *geodesic paths* in spaces of diffeomorphisms generated by a velocity field with applications to imaging.

Chapter 4 also gives general equivalences between transformations and flows of velocity fields for unconstrained and constrained families of domains and further sharpens the results to the specific families of transformations associated with the Courant metrics studied in Chapter 2. Several examples of transformations and velocities associated with widely used families of domains are given: C^∞-domains, C^k-domains, Cartesian graphs, polar coordinates and star-shaped domains, and level sets. The chapter clarifies the long-standing issue of the equivalence of the continuity of shape functionals with respect to the Courant metrics and along the flow of velocity fields.

This chapter also prepares the ground for and motivates the definition of shape semiderivatives that will be given in Chapter 9. As in the case of the continuity the equivalent characterizations via transformations and flows of velocities are very much in the background and at the origin of the many seemingly different definitions which can be found in the literature. Preliminary considerations are first given to the definition of a shape functional and to two candidates for the definition of a directional shape semiderivative. They respectively correspond to *perturbations of the identity* associated with any one of the metric spaces constructed in Chapter 2 and to the *velocity method* associated with the flow of a generally nonautonomous vector field. The first one seems to be limited to domains in $\mathbf{R}^{\mathrm{N}}$, while the second one naturally extends to domains living in a fixed smooth submanifold of $\mathbf{R}^{\mathrm{N}}$. Moreover, the shape directional derivative obtained by perturbations of the identity can be recovered by a special choice of velocity field. Most definitions of shape derivatives which can be found in the literature can be brought down to one of the two approaches.

11.2 Functions Parametrized by Geometries

Almost all compactness theorems for families of sets are specified by function analytic conditions on special set-parametrized families of functions. In this book we single out the *characteristic function*, the *distance function*, and the *oriented distance function*. In the last case, a complete and convivial tangential differential calculus on the boundary of smooth sets will be obtained without local bases or Christoffel symbols and will be exploited in the computation of shape derivatives later in Chapters 9 and 10.

Chapter 5 relaxes the family of classical domains to the equivalence classes of Lebesgue measurable sets. Using the *characteristic function* associated with a set, complete *metric groups* of equivalence classes of characteristic functions are constructed. As for Courant metrics on groups, they are nonlinear and nonconvex. On one hand, this type of relaxation is desirable in optimization problems where the topology of the optimal set is not a priori specified; on the other hand, it necessitates the relaxation of the theory of partial differential equations on a smooth open domain to measurable sets that have no smoothness and may not even be open. Furthermore some optimization problems yield optimal solutions where the characteristic function is naturally relaxed to a function between zero and one. Such solutions can be interpreted as *microstructures*, *fuzzy sets*, probability measures, etc. As a first illustration of the use of characteristic functions in optimization, the solution of the original problem of J. Céa and K. Malanowski [**1**] for the optimization of the compliance with respect to the distribution of two materials is given with complete details. A second example deals with the buckling of columns, which is one of the very early optimal design problems formulated by Lagrange in 1770. This is followed by the construction of the *nice representative* of an equivalence class of measurable functions: the *measure theoretic representative*. The last section is devoted to the *Caccioppoli or finite perimeter sets*, which have been introduced to solve the Plateau problem of minimal surfaces (J. A. F. Plateau [**1**]). Even if, by nature, a characteristic function is discontinuous at the boundary of the associated set, the characteristic function of Caccioppoli sets has some smoothness: it belongs to $W^{\epsilon,p}(D)$, $1 \leq \epsilon < 1/p$, $p \geq 1$. This is sufficient to obtain compact families of sets by putting a uniform bound on the perimeter. One such family is the set of (locally) Lipschitzian (epigraph) domains contained in a fixed bounded *holdall* and satisfying a uniform cone property. This property puts a uniform bound on the perimeter of the sets. The use of the theory of finite perimeter sets is illustrated by an application to a free boundary problem in fluid mechanics: the modeling of the Bernoulli wave where the surface tension of the water enters via the perimeter of the free boundary. The chapter concludes with an approximation of the Dirichlet problem by transmission problems over a fixed larger space in order to study the continuity of its solution with respect to the underlying moving domains.

Chapter 6 moves on to the classical Pompéiu–Hausdorff metric topology which is associated with the space of equivalence classes of distance functions of sets with the same closure. As in Chapter 2, the construction of the metric on equivalence classes of sets with the same closure is generic. By going to the distance function of the complement of the set in the uniform topology of the continuous functions, we get the *complementary Hausdorff topology.* These uniform topologies are often too coarse for applications to physical or technological systems. But, since the distance function is uniformly Lipschitzian, it can also be embedded into $W^{1,p}$-Sobolev spaces, and finer metric topologies can be generated. They offer definite advantages over the uniform Hausdorff topology in the sense that they preserve the continuity of the volume of sets since the characteristic function is continuous with respect to $W^{1,p}$-topologies. Yet, we lose the compactness of the family of subsets of a fixed bounded holdall of $\mathbf{R}^N$. Compact families are recovered by imposing some smoothness on the Hessian matrix as in the case of the characteristic functions of Chapter 5. Sets for which the gradient of the distance function is a vector of functions of bounded variation are said to be of *bounded curvature* since their Hessian matrix is intimately connected with the curvatures of the boundary. This class of sets is sufficiently large for applications and at the same time sufficiently structured to obtain interesting theoretical results. For instance, such sets turn out to be Caccioppoli sets. Furthermore, the squared distance function is directly related to the projection of a point onto the set and is used to characterize *Federer's sets of positive reach* and construct compact families. Closed convex sets that are completely characterized by the convexity of their distance function are also of locally bounded curvature. To complete the list of families of sets that are associated with the distance function we introduce Federer's *sets of positive reach* and a first compactness theorem. The chapter is complemented with a general compactness theorem for families of sets of global or local bounded curvature in a tubular neighborhood of their boundary.

The use of the distance function of Chapter 6 to characterize the smoothness of sets is limited by the fact that its gradient presents a jump discontinuity at the boundary. This is similar to the jump discontinuity of the characteristic function. To get around this difficulty we introduce the *oriented distance function* that is obtained by subtracting from the distance function to a set the distance function to its complement, providing a level set description of the set. One remarkable property of this function is the fact that a set is of class $C^{1,1}$ (resp., C^k, $k \geq 2$) if and only if its oriented distance function is locally $C^{1,1}$ (resp., C^k) in a neighborhood of its boundary. It also provides an orientation of the boundary since its gradient coincides with the unit outward normal. This is why we use the terminology *oriented distance function.* As in Chapter 6, Hausdorff and $W^{1,p}$-topologies and sets of global or locally bounded curvature can be introduced. We now have a continuous classification of sets ranging from sets with a nonempty boundary to C^∞-sets, much as in the theory of functions. Closed convex sets are characterized by the convexity of their oriented distance function. Convex sets and semiconvex sets are of locally bounded curvature. This property extends to Federer's sets of positive reach.

Several compactness theorems in the $W^{1,p}$-topology of the oriented distance function are presented in this chapter for families of subsets of a bounded open holdall with either bounded curvature in tubular neighborhoods of their boundary, or with a bound on the *density perimeter* of D. Bucur and J.-P. Zolésio [**8**], or with a *uniform fat segment property* or equivalently the uniform boundedness and equicontinuity of all the local graphs of the sets in the family. The theorem on the compactness under the uniform fat segment property is specialized to the family of subsets of a bounded holdall satisfying a uniform cone or cusp property. It is more general than the one we obtained for Lipschitzian domains in Chapter 5, where the perimeter of each set was finite and uniformly bounded for all subsets of the holdall. Hölderian domains do not necessarily have a locally finite boundary measure. Yet, the fact that the boundary of a Hölderian domain may have cusp does not mean that all Hölderian domains do not have a locally finite boundary measure. In order to include applications where the perimeter is bounded, the general compactness theorem is specialized to families of subsets of a holdall that verify the uniform fat segment property with a uniform bound on either the De Giorgi perimeter or the density perimeter. A last section 15 deals with the family of *cracked sets*. They have been used in M. C. Delfour and J.-P. Zolésio [**38**] in the context of the image segmentation problem of D. Mumford and J. Shah [**2**] that will be detailed in this section. Cracked sets are more general than sets which are locally the epigraph of a continuous function in the sense that they include domains with cracks, and sets that can be made up of components of different codimensions. The Hausdorff $(N-1)$ measure of their boundary is not necessarily finite. Yet, compact families (in the $W^{1,p}$-topology) of such sets can be constructed.

11.3 Shape Continuity and Optimization

Chapter 8 deals with problem formulations, continuity, or semicontinuity of shape functionals or of the solutions of boundary value problems with respect to their underlying domain of definition for selected generic examples. Combined with compact families of sets studied in the previous chapters, they are all essential to getting the existence of optimal shapes. This is illustrated to some extent in Chapter 5 for the modeling of the transmission problem with the help of the characteristic function that occurs both in the model and in the specification of the metric on the equivalence classes of measurable sets.

This chapter first reviews the continuity of the transmission problem using characteristic functions. It characterizes the upper semicontinuity of the first eigenvalue of the generalized Laplacian with respect to the domain using the complementary Hausdorff topology. Then it studies the continuity of the solution of the homogeneous Dirichlet and Neumann boundary value problems with respect to their underlying domain of definition since they require different constructions and topologies that are generic of the two types of boundary conditions even for more complex nonlinear partial differential equations. In problems where the objective functional depends explicitly on the domain and the solution of an elliptic equation defined on the same domain, the strong continuity of the solution with respect to the underlying domain is the key element in the proof of the existence of optimal domains. To get that continuity, some extra conditions have to be imposed on the family of open domains, such as the uniform fat segment property.

The second part extends some of the results to a larger family of domains satisfying *capacity conditions* which turn out to be important to obtain the continuity of solutions of partial differential equations with homogeneous Dirichlet boundary conditions with respect to their underlying domain of definition. One special case of a capacity condition is the *flat cone condition* that generalizes the condition of V. Šverák [**2**] involving a bound on the number of connected components of the complement of the sets.

11.4 Derivatives, Shape and Tangential Differential Calculuses, and Derivatives under State Constraints

Chapter 9 is devoted to the essential *shape calculus* and the no less essential *tangential or boundary calculuses* since traces, normals, or tangential gradients will naturally occur in the final expressions.

After a self-contained review of differentiation in topological vector spaces that emphasizes Gateaux and Hadamard differentials, it introduces basic definitions of first- and second-order Eulerian shape semiderivatives and derivatives by the *velocity method*. General structure theorems are given for Eulerian semiderivatives of a shape functional. They arise from the fact that shape functionals are usually defined over equivalence classes of sets, and hence only the normal part of the velocity along the boundary really affects the shape functional. Bridges are provided with the *method of perturbations of the identity*. A section is devoted to a modern version of the shape calculus. It gives the general formulae for the shape derivative of domain and boundary integrals. From these formulae several examples are worked out, including the semiderivative of the boundary integral of the square of the normal derivative. For the computation of a broader range of shape derivatives the reader is referred to the book by J. Sokołowski and J.-P. Zolésio [**9**]. In most cases, both domain and boundary expressions are available for derivatives. The boundary expression usually contains more information on the structure of the derivative than its domain counterpart. Finally, to effectively deal with the differential calculus in boundary integrals, we provide the latest version of the *tangential calculus* on C^2-submanifolds of codimension 1 which has been developed in the context of the theory of shells (cf. M. C. Delfour and J.-P. Zolésio [**28**, **33**] and M. C. Delfour [**3**, **7**]). This calculus has been significantly simplified by using the projection associated with the oriented distance function studied in Chapter 7. This powerful tool combined with the shape calculus makes it possible to obtain clean explicit expressions of second-order shape derivatives of domain integrals along with a better understanding of their fine structure.

Chapter 10, the final chapter, completes the shape calculus by introducing the basic theoretical results and computational tools for the shape derivative of functionals that depend on a state variable that is usually the solution of a partial differential equation or inequality defined over the varying underlying domain. The first section concentrates on shape functionals that are of *compliance type*; that is, they are the minimum of an energy functional associated with the state equation or inequality. Such functionals are very nice in the sense that they do not generate an adjoint state equation, and their derivative can be obtained by theorems on the differentiability of a minimum with respect to a parameter even when the minimizers are not unique. A detailed generic example is provided to illustrate how to use the *function space parametrization* to transport the functions in Sobolev spaces over variable domains to a Sobolev space over the fixed larger holdall. These techniques extend to more complex situations. For instance, Sobolev spaces of vector functions with zero divergence can be transported by the so-called Piola transformation. Domain and boundary expressions are provided. The main theorem is applied to the example of the buckling of columns. An explicit expression of the semiderivative of *Euler's buckling load* with respect to the cross-sectional area is obtained from the main theorem, and a necessary and sufficient analytical condition is given to characterize the maximum Euler's buckling load with respect to a family of cross-sectional areas. The theory is further illustrated by providing the semiderivative of the first eigenvalue of several boundary value problems over a bounded open domain: Laplace equation, bi-Laplace equation, linear elasticity. In general, the first eigenvalue is not simple over an arbitrary bounded open domain and the eigenvalue is not differentiable; yet the main theorem provides explicit domain and boundary expressions of the semiderivatives.

For general shape functionals, a Lagrangian formulation is used to incorporate the state equation and to avoid the study of the derivative of the state equation with respect to the domain. The computation of the shape derivative of a state-constrained functional reduces to the computation of the derivative of a saddle point with respect to a parameter even when the saddle point solution is not unique. It yields an expression that depends on the associated *adjoint state equation*, much like in control theory, but here the domains play the role of the controls. The technique is illustrated on both the homogeneous Dirichlet and the Neumann boundary value problems by function space parametrization. An alternative to this method is the *function space embedding* combined with the use of Lagrange multipliers. It consists in extending solutions of the boundary value problems over the variable domains to a larger fixed holdall, rather than transporting them. This approach offers many technical advantages over the other one. The computations are easier and they apply to larger classes of problems. This is illustrated on the nonhomogeneous Dirichlet boundary value problem. Again domain and boundary expressions for the shape gradient are obtained. Yet the relative advantages of one method over the other are very much problem and objective dependent. Finally, it is important to acknowledge that the above techniques seem quite robust and are systematically used for nonlinear state equations and in contexts where optimization or saddle point formulations are not available.