Shape optimization: geometrical and topological aspects

Master of Mathematics and Applications, 2^{nd} year
Specialty Mathematics of Modeling
Ecole polytechnique and Sorbonne university
Winter 2023

Samuel Amstutz
CMAP and Department of Applied Mathematics
Ecole Polytechnique
samuel.amstutz@polytechnique.edu

Contents

Introduction

These notes deal with the wide topic of shape optimization, also called optimal design. In this field of research and engineering, one is interested in finding a geometrical domain, usually of $\mathbb{R}^2$ or $\mathbb{R}^3$, that is optimal for some objective (or cost) function(s), under possible constraints. Actually, in many situations there exists no optimal domain, and / or some local minima occur. In such cases the attention is mainly focused on improving a given domain (the initial guess), or sometimes even finding a domain that satisfies the constraints.

We often distinguish between three types of shape optimization problems.

1. In **parametric shape optimization** (or sizing optimization), the shape is described by a vector of a normed vector space (finite dimensional or not). This allows the use of standard optimization strategies.

2. In **classical shape optimization** (or geometry optimization), the shape admits no natural parametric representation, but it is assumed to have the same topology as a given reference shape. In particular, all shapes obtained in this way are homeomorphic. In 2D they have the same number of holes.

3. In **topology optimization** both the geometry and the topology are subject to optimization.

This course aims at providing some important tools and concepts in order to address and analyze such problems.

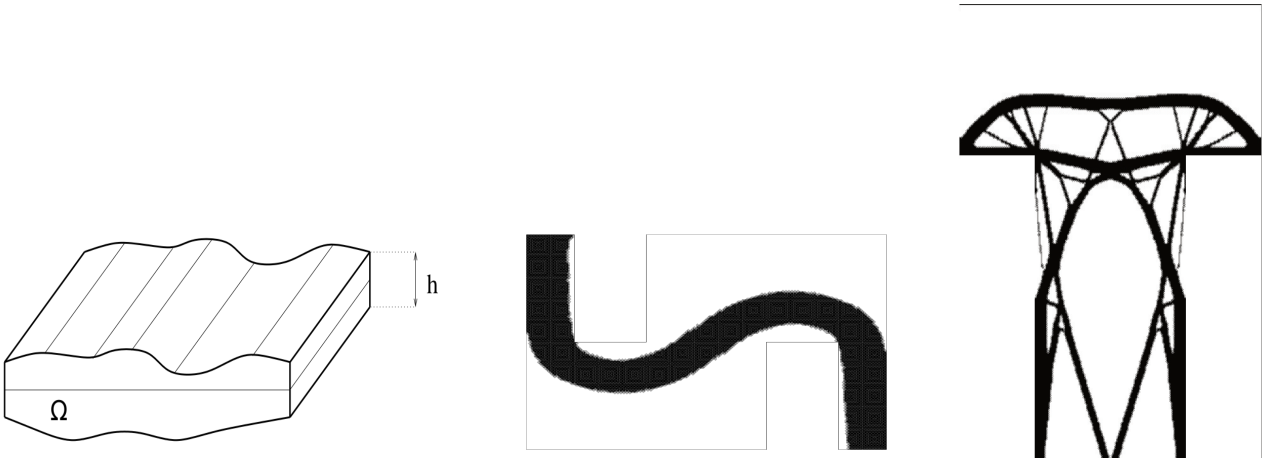


Figure 1: Sizing vs geometry vs topology.

Chapter 1

Prerequisites

1.1 Differential calculus

We recall here some classical elements of (Fréchet) differential calculus. All vector spaces will be on the real field. We refer to [7] for proofs and further results.

1.1.1 Definitions

Definition 1.1 *Let X, Y be two normed vector spaces, U be an open subset of X, $f : U \to Y$ and $x_0 \in U$. We say that f is Fréchet differentiable at x_0 if there exists $L \in \mathcal{L}(X, Y)$, the set of continuous linear maps from X to Y, such that*

$$\lim_{h \to 0} \frac{\|f(x_0 + h) - f(x_0) - L(h)\|}{\|h\|} = 0.$$

In this case the map L is unique, it is called the Fréchet derivative of f at x_0, denoted by $df(x_0)$.

If the map $x \in U \mapsto df(x) \in \mathcal{L}(X, Y)$ is continuous we say that f is continuously Fréchet differentiable.

It is clear that if $f \in \mathcal{L}(X, Y)$ then f is (continuously) Fréchet differentiable on X with $df(x)h = f(h)$ for all $x, h \in X$. It is also straightforward that if a function f is Fréchet differentiable at a point x_0 then it is continuous at x_0.

It is sometimes useful to consider the weaker notion of directional differentiability.

Definition 1.2 *Let X, Y be two normed vector spaces, U be an open subset of X, $f : U \to Y$ and $x_0 \in U$. The derivative of f at x_0 in the direction $h \in X$, if it exists, is defined by*

$$f'(x; h) = \lim_{t \searrow 0} \frac{f(x_0 + th) - f(x_0)}{t}.$$

It is easy to see that if f is Fréchet differentiable at x_0 then f admits directional derivatives in all directions and we have $f'(x_0; h) = df(x_0)h$.

1.1.2 Finite dimensional case

In finite dimension the Fréchet derivative is represented by the Jacobian matrix.

Proposition 1.3 *Let $f : x = (x_1, ..., x_n) \in U \subset \mathbb{R}^n \mapsto f(x) = (f_1(x), ..., f_m(x)) \in \mathbb{R}^m$. If f is Fréchet differentiable at $x \in U$ then $df(x)$ is represented in the canonical bases by the Jacobian matrix*

$$Df(x) = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(x) & \dots & \frac{\partial f_1}{\partial x_n}(x) \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1}(x) & \dots & \frac{\partial f_m}{\partial x_n}(x) \end{pmatrix}.$$

1.1.3 Rules of calculus

We have seen that continuous linear maps admit straightforward Fréchet derivatives. This extends to continuous multilinear maps as follows.

Proposition 1.4 *Let $X_1, \cdots, X_n, Y$ be normed vector spaces, and $f \in \mathcal{L}_n(X_1 \times \cdots \times X_n, Y)$. Then f is Fréchet differentiable at all $(x_1, \cdots, x_n) \in X_1 \times \cdots \times X_n$ with*

$$df(x_1, \cdots, x_n)(h_1, \cdots, h_n) = \sum_{i=1}^{n} f(x_1, \cdots, x_{i-1}, h_i, x_{i+1}, \cdots, x_n).$$

We now state the chain rule.

Theorem 1.5 *Let X, Y, Z be normed vector spaces, U be an open subset of X, V be an open subset of Y, $f : U \to Y$, $g : V \to Z$ and $x_0 \in U$ be such that $f(x_0) \in V$. If f is Fréchet differentiable at x_0 and g is Fréchet differentiable at $f(x_0)$ then $g \circ f$ is Fréchet differentiable at x_0 with*

$$d(g \circ f)(x_0)h = dg(f(x_0))(df(x_0)h).$$

In finite dimension, the Jacobian matrix of a composite function is the product of the Jacobian matrices.

For example, consider an inner product space X and the composite map

$$\phi : x \in X \mapsto (x, x) \in X \times X \mapsto \langle x, x \rangle = \|x\|^2.$$

Using the chain rule, the linearity of the first map and the bilinearity of the inner product we immediately get

$$d\phi(x)h = \langle x, h \rangle + \langle h, x \rangle = 2\langle x, h \rangle.$$

1.1.4 Mean value theorem

The mean value theorem extends the mean value inequality for functions of one variable.

Theorem 1.6 *Let X, Y be two normed vector spaces, U be an open subset of X, $f : U \to Y$ be Fréchet differentiable on U and $x, y \in U$ be such that the segment $[x, y] := \{\theta x + (1 - \theta y), 0 \leq \theta \leq 1\}$ is in U. We have*

$$\|f(x) - f(y)\| \leq \|x - y\| \sup_{0 \leq \theta \leq 1} \|df(\theta x + (1 - \theta y))\|_{\mathcal{L}(X,Y)}.$$

One of its consequences is the following.

Theorem 1.7 *Let X, Y be two normed vector spaces, U be an open and connected subset of X, $f : U \to Y$ be Fréchet differentiable on U. If $df(x) = 0$ for all $x \in U$ then f is constant.*

1.1.5 Implicit functions

The implicit function theorem is usually proven using the Banach fixed point theorem. Therefore Banach spaces are required.

Theorem 1.8 *Let X, Y, Z be three Banach spaces, $\mathcal{O}$ be an open subset of $X \times Y$, $f : \mathcal{O} \to Z$ be continuously Fréchet differentiable, $(x_0, y_0) \in \mathcal{O}$ be such that $f(x_0, y_0) = 0$. If the map $h \in Y \mapsto df(x_0, y_0)(0, h) \in Z$ is an isomorphism then there exists open neighborhoods U and V of x_0 and y_0, respectively, and a Fréchet differentiable function $\varphi : U \to V$ such that*

$$\forall (x, y) \in U \times V, \qquad f(x, y) = 0 \Leftrightarrow y = \varphi(x).$$

1.1.6 Two useful examples

Proposition 1.9 *Let X, Y be a Banach spaces. The set* $\operatorname{isom}(X, Y)$ *of isomorphisms from X into Y is an open subset of $\mathcal{L}(X, Y)$. The map $f : u \in \operatorname{isom}(X, Y) \mapsto u^{-1}$ is Fréchet differentiable with*

$$df(u)h = -u^{-1} \circ h \circ u^{-1} \qquad \forall h \in \mathcal{L}(X, Y).$$

Proof. Given $u \in \operatorname{isom}(X, Y)$ and $h \in \mathcal{L}(X, Y)$ with $\|h\|_{\mathcal{L}(X,Y)} < \|u^{-1}\|^{-1}_{\mathcal{L}(Y,X)}$, we have $\|u^{-1} \circ h\|_{\mathcal{L}(X)} < 1$, hence the Neumann series $\sum(-u^{-1} \circ h)^k$ is normally converging. We have by cancellation of terms

$$(\mathrm{Id}_X + u^{-1} \circ h) \circ \sum_{k=0}^{+\infty} (-u^{-1} \circ h)^k = \mathrm{Id}_X,$$

leading to

$$(u + h) \circ \sum_{k=0}^{+\infty} (-u^{-1} \circ h)^k \circ u^{-1} = \mathrm{Id}_X .$$

Likewise,

$$\left(\sum_{k=0}^{+\infty} (-u^{-1} \circ h)^k \circ u^{-1}\right) \circ (u + h) = \left(\sum_{k=0}^{+\infty} (-u^{-1} \circ h)^k\right) \circ (\mathrm{Id}_X + u^{-1} \circ h) = \mathrm{Id}_X .$$

It follows that $u + h \in \operatorname{isom}(X, Y)$ with

$$\begin{aligned}
(u + h)^{-1} &= \sum_{k=0}^{+\infty} (-u^{-1} \circ h)^k \circ u^{-1} = \left(\mathrm{Id}_X - u^{-1} \circ h + (u^{-1} \circ h)^2 \circ \sum_{k=0}^{+\infty} (-u^{-1} \circ h)^k\right) \circ u^{-1} \\
&= u^{-1} - u^{-1} \circ h \circ u^{-1} + (u^{-1} \circ h)^2 \circ \sum_{k=0}^{+\infty} (-u^{-1} \circ h)^k \circ u^{-1} \\
&= u^{-1} - u^{-1} \circ h \circ u^{-1} + o(\|h\|_{\mathcal{L}(X,Y)}),
\end{aligned}$$

from which we infer the claim. □

In particular the map $f : A \in GL_n(\mathbb{R}) \mapsto A^{-1}$ is Fréchet differentiable with

$$df(A)H = -A^{-1}HA^{-1}. \tag{1.1}$$

Proposition 1.10 *The map $f : A \in \mathcal{M}_n(\mathbb{R}) \mapsto \det A$ is Fréchet differentiable with*

$$df(A)H = \operatorname{cof}(A) : H \qquad \forall H \in \mathcal{M}_n(\mathbb{R}),$$

with $\operatorname{cof}(A)$ *the cofactor matrix of A and $:$ the Frobenius inner product of matrices.*

Proof. Consider the map as a function of the column vectors. For clarity we set

$$\begin{array}{rrcl} F : & (\mathbb{R}^n)^n & \to & \mathbb{R} \\ & (u_1, ..., u_n) & \mapsto & \det(u_1, ..., u_n). \end{array}$$

By $n-$ linearity, F is Fréchet differentiable with

$$dF(u_1, ..., u_n)(h_1, ..., h_n) = \sum_{j=1}^{n} \det(u_1, ..., u_{j-1}, h_j, u_{j+1}, ..., u_n).$$

Expanding the determinant with respect to column j yields

$$dF(u_1, ..., u_n)(h_1, ..., h_n) = \sum_{j=1}^{n} \sum_{i=1}^{n} h_{i,j} \operatorname{cof}_{ij}(A) = \operatorname{cof}(A) : H.$$

□

We will often apply this result at the identity matrix, where we have $df(I)H = \operatorname{tr} H$.

1.2 Sobolev spaces

Let Ω be an open subset of $\mathbb{R}^N$. We equip $\mathbb{R}^N$ with the Lebesgue measure dx. We refer to [6, 8, 11] for proofs and further results.

1.2.1 Lebesgue spaces

Definition 1.11 *The Lebesgue spaces are defined by*

$$L^p(\Omega) = \left\{ u : \Omega \to \mathbb{R} : u \textit{ measurable}, \int_\Omega |u(x)|^p dx < +\infty \right\} \qquad \textit{for } 1 \le p < +\infty,$$
$$L^\infty(\Omega) = \{ u : \Omega \to \mathbb{R} : u \textit{ measurable}, \exists M > 0 \textit{ s.t. } |u(x)| \le M \textit{ a.e. } x \in \Omega \}.$$

To be completely rigorous, the Lebesgue spaces are spaces of classes of functions up to the equality almost everywhere. This aspect is made implicit for the sake of readability.

Theorem 1.12 *Equipped with the norms*

$$\|u\|_{L^p(\Omega)} = \left(\int_\Omega |u(x)|^p dx \right)^{1/p} \qquad \textit{for } 1 \le p < +\infty,$$
$$\|u\|_{L^\infty(\Omega)} = \inf \ \{ M > 0 \textit{ s.t. } |u(x)| \le M \textit{ a.e. } x \in \Omega \},$$

the Lebesgue spaces are Banach spaces. Moreover, $L^2(\Omega)$ is a Hilbert space for the inner product

$$\langle u, v \rangle_{L^2(\Omega)} = \int_\Omega u(x)v(x)dx.$$

We recall Hölder's inequality:

Theorem 1.13 *If $u \in L^p(\Omega)$, $v \in L^q(\Omega)$, $1 \le p \le +\infty$, $1/p + 1/q = 1$, then $uv \in L^1(\Omega)$ and we have*

$$\|uv\|_{L^1(\Omega)} \le \|u\|_{L^p(\Omega)} \|v\|_{L^q(\Omega)}.$$

This provides a continuous embedding $L^p(\Omega) \hookrightarrow L^q(\Omega)'$, the continuous dual of $L^q(\Omega)$, by

$$\Phi : u \in L^p(\Omega) \mapsto \left[v \in L^q(\Omega) \mapsto \int_\Omega uv dx \right] \in L^q(\Omega)'.$$

Theorem 1.14 *When $1 < p \le +\infty$ the map Φ is bijective and isometric. We systematically identify $L^p(\Omega)$ and $L^q(\Omega)'$. For $1 < p < +\infty$, we have through this identification $L^p(\Omega)'' = L^p(\Omega)$: we say that $L^p(\Omega)$ is reflexive.*

We denote by $\mathcal{C}^k_c(\Omega)$ the set of k times continuously differentiable functions on Ω with compact support, i.e.

$$\mathcal{C}^k_c(\Omega) = \left\{ u \in \mathcal{C}^k(\Omega), u(x) = 0 \ \forall x \in \Omega \setminus K \text{ for some } K \text{ compact} \subset \Omega \right\}.$$

Theorem 1.15 *For $1 \le p < +\infty$, the set $\mathcal{C}^\infty_c(\Omega)$ is dense in $L^p(\Omega)$ and $L^p(\Omega)$ is separable (i.e. it admits a countable dense subset).*

Theorem 1.16 *Let $u_n, u \in L^p(\Omega)$, $1 \le p \le +\infty$ be such that $\lim_{n\to+\infty} \|u_n - u\|_{L^p(\Omega)} = 0$. There exists a subsequence $(u_{\iota(n)})$ such that $u_{\iota(n)} \to u$ a.e. in Ω.*

We will also use the set of locally integrable functions

$$L^1_{\text{loc}}(\Omega) = \left\{ u : \Omega \to \mathbb{R} : u \text{ measurable}, \int_K |u(x)| dx < +\infty \ \forall K \subset \Omega, K \text{ compact} \right\}.$$

Hölder's inequality yields $L^p(\Omega) \subset L^1_{\text{loc}}(\Omega)$ for all $p \in [1, +\infty]$.

1.2.2 Weak derivatives

We denote by $\cdot$ the canonical inner product of $\mathbb{R}^N$.

Definition 1.17 *A function $u \in L^1_{\text{loc}}(\Omega)$ is said to be weakly differentiable if there exists $v \in L^1_{\text{loc}}(\Omega)^N$ such that*

$$-\int_\Omega u(x)\operatorname{div}\varphi(x)dx = \int_\Omega v(x)\cdot\varphi(x)dx \qquad \forall\varphi \in \mathcal{C}^1_c(\Omega)^N.$$

Such a function v is unique (in the almost everywhere sense) and it is called weak derivative (or gradient) of u. It is denoted by $v = \nabla u$.

The notion of weak derivative is a particular case of the notion of derivative in the sense of distributions. Of course, by integration by parts, ∇u coincides with the usual gradient if $u \in \mathcal{C}^1(\Omega)$.

By a density argument, Theorem 1.7 extends as folows.

Theorem 1.18 *Let Ω be an open connected subset of $\mathbb{R}^N$ and let $u \in L^1_{\text{loc}}(\Omega)$ be weakly differentiable with $\nabla u = 0$. Then u is constant.*

1.2.3 First order Sobolev spaces

Definition 1.19 *For any $p \in [1, +\infty]$ the Sobolev space $W^{1,p}(\Omega)$ is defined by*

$$W^{1,p}(\Omega) = \left\{u \in L^p(\Omega) : \nabla u \in L^p(\Omega)^N\right\},$$

where ∇u is intended in the weak sense. The space $W^{1,2}(\Omega)$ is most often denoted by $H^1(\Omega)$.

We donote by $\left(\frac{\partial u}{\partial x_1}, \cdots, \frac{\partial u}{\partial x_N}\right)$ the components of ∇u.

Theorem 1.20 *Equipped with the norms*

$$\|u\|_{W^{1,p}(\Omega)} = \left(\|u\|^p_{L^p(\Omega)} + \sum_{i=1}^N \left\|\frac{\partial u}{\partial x_i}\right\|^p_{L^p(\Omega)}\right)^{1/p} \qquad \text{for } 1 \le p < +\infty,$$

$$\|u\|_{W^{1,\infty}(\Omega)} = \sup\left\{\|u\|_{L^\infty(\Omega)}, \left\|\frac{\partial u}{\partial x_1}\right\|_{L^\infty(\Omega)}, \cdots, \left\|\frac{\partial u}{\partial x_N}\right\|_{L^\infty(\Omega)}\right\}$$

the Sobolev space $W^{1,p}(\Omega)$ is a Banach space. Moreover, $H^1(\Omega)$ is a Hilbert space for the inner product

$$\langle u, v\rangle_{H^1(\Omega)} = \langle u, v\rangle_{L^2(\Omega)} + \sum_{i=1}^N \left\langle \frac{\partial u}{\partial x_i}, \frac{\partial v}{\partial x_i}\right\rangle_{L^2(\Omega)}.$$

Note that we could have used the equivalent norm on $W^{1,p}(\Omega)$, $1 \le p \le +\infty$, defined by

$$\|u\|_{L^p(\Omega)} + \sum_{i=1}^N \left\|\frac{\partial u}{\partial x_i}\right\|_{L^p(\Omega)},$$

but the previous ones have the advantage of being consistent with the inner product in the case $p = 2$.

The space $\mathcal{C}^\infty_c(\Omega)$ is generally not dense in $W^{1,p}(\Omega)$. This motivates the following definition.

Definition 1.21 *We define*

$$W^{1,p}_0(\Omega) = \text{ the closure of } \mathcal{C}^\infty_c(\Omega) \text{ in } W^{1,p}(\Omega).$$

The set $W^{1,2}_0(\Omega)$ is usually denoted by $H^1_0(\Omega)$.

It is nevertheless true that, for $p \in [1, +\infty[$, $\mathcal{C}^\infty_c(\mathbb{R}^N)$ is dense in $W^{1,p}(\mathbb{R}^N)$, hence $W^{1,p}_0(\mathbb{R}^N) = W^{1,p}(\mathbb{R}^N)$.

1.2.4 Traces

Definition 1.22 *We say that Ω is of class $\mathcal{C}^k$, $k \geq 1$, if for all $x \in \partial\Omega$ (the topological boundary of Ω) there exist an open neighborhood $\mathcal{O}$ of x, a vector $d \in \mathbb{R}^N$ and a $\mathcal{C}^k$ diffeomorphism φ from $B(0,1)$ to $\mathcal{O}$ such that*

$$\Omega \cap \mathcal{O} = \varphi(\{x \in B(0,1), x \cdot d > 0\}),$$

$$\partial\Omega \cap \mathcal{O} = \varphi(\{x \in B(0,1), x \cdot d = 0\}).$$

In words, this means that $\partial\Omega$ is a submanifold of $\mathbb{R}^N$ of dimension $N-1$ and of class $\mathcal{C}^k$, and that Ω is locally on one side of $\partial\Omega$.

We denote

$$\mathcal{C}^k(\bar{\Omega}) = \left\{ u_{|\Omega}, u \in \mathcal{C}^k(\mathbb{R}^N) \right\}.$$

Theorem 1.23 *If Ω is bounded and of class $\mathcal{C}^1$ then $\mathcal{C}^\infty(\bar{\Omega})$ is dense in $W^{1,p}(\Omega)$ for $1 \leq p < +\infty$.*

If Ω is of class $\mathcal{C}^1$, then we can define integrals over $\partial\Omega$. This permits to define the space $L^p(\partial\Omega)$. In what follows we assume that Ω is bounded and of class $\mathcal{C}^1$, and, unless other specified, that $p \in [1, +\infty[$.

Theorem 1.24 *The restriction mapping*

$$u \in \mathcal{C}^\infty(\bar{\Omega}) \mapsto u_{|\partial\Omega} \in L^p(\partial\Omega)$$

extends by continuity into a linear and continuous mapping

$$\gamma_0 : W^{1,p}(\Omega) \to L^p(\partial\Omega)$$

called trace operator of order 0.

Definition 1.25 *For $p \in]1, +\infty[$ we define the trace space $W^{1-1/p,p}(\partial\Omega)$ as the image of γ_0, namely*

$$W^{1-1/p,p}(\partial\Omega) = \gamma_0(W^{1,p}(\Omega)).$$

For $p = 2$ the space $W^{\frac{1}{2},2}(\partial\Omega)$ is usually denoted by $H^{1/2}(\partial\Omega)$.

Theorem 1.26 *Equipped with the norm*

$$\|p\|_{W^{1-1/p,p}(\partial\Omega)} = \inf\{\|u\|_{W^{1,p}(\Omega)} : \gamma_0 u = p\}$$

the trace space $W^{1-1/p,p}(\partial\Omega)$ is a Banach space. Moreover, for any $p \in [1, +\infty[$, the map $\gamma_0 : W^{1,p}(\Omega \to W^{1-1/p,p}(\partial\Omega)$ is linear, continuous and surjective with kernel

$$\ker \gamma_0 = W_0^{1,p}(\Omega).$$

Note that the space $W^{1-1/p,p}(\partial\Omega)$ admits equivalent intrinsic characterizations that justify the notation. All these characterizations are fairly involved, we will not use them.

1.2.5 Sobolev embeddings

There are a number of embedding results involving Sobolev spaces. We limit ourselves to a particular case of Rellich's theorem.

Theorem 1.27 *If Ω is bounded then the embedding $W_0^{1,p}(\Omega) \hookrightarrow L^p(\Omega)$ is compact for all $p \in [1, +\infty[$. If Ω is bounded and of class $\mathcal{C}^1$ then the embedding $W^{1,p}(\Omega) \hookrightarrow L^p(\Omega)$ is compact for all $p \in [1, +\infty]$.*

1.2.6 Higher order Sobolev spaces

By induction, we define for $p \in [1, +\infty]$

$$W^{k,p}(\Omega) = \left\{ u \in W^{k-1,p}(\Omega) : \frac{\partial u}{\partial x_i} \in W^{k-1,p}(\Omega), i = 1, \cdots, N \right\}.$$

Here again, the derivatives are meant in the weak sense. The space $W^{k,2}(\Omega)$ is denoted by $H^k(\Omega)$. The space $W^{k,p}(\Omega)$ is a Banach space for the norm

$$\|u\|_{W^{k,p}(\Omega)} = \left(\|u\|^p_{L^p(\Omega)} + \sum_{i=1}^{N} \left\| \frac{\partial u}{\partial x_i} \right\|^p_{W^{k-1,p}(\Omega)} \right)^{1/p} \qquad \text{for } 1 \leq p < +\infty,$$

$$\|u\|_{W^{k,\infty}(\Omega)} = \sup \left\{ \|u\|_{L^\infty(\Omega)}, \left\| \frac{\partial u}{\partial x_1} \right\|_{W^{k-1,\infty}(\Omega)}, \cdots, \left\| \frac{\partial u}{\partial x_N} \right\|_{W^{k-1,\infty}(\Omega)} \right\}.$$

The space $H^k(\Omega)$ is a Hilbert space.

We also define

$$W_0^{k,p}(\Omega) = \text{ the closure of } \mathcal{C}_c^\infty(\Omega) \text{ in } W^{k,p}(\Omega).$$

Higher order Sobolev spaces allow to define higher order trace operators. For simplicity we restrict ourselves to the trace of order 1. We assume that Ω is bounded and of class $\mathcal{C}^2$. For $p \in]1, +\infty[$ we define the trace space $W^{2-1/p,p}(\partial\Omega)$ as

$$W^{2-1/p,p}(\partial\Omega) = \gamma_0(W^{2,p}(\Omega)).$$

By construction we have $W^{2-1/p,p}(\partial\Omega) \subset W^{1-1/p,p}(\partial\Omega)$. For $p = 2$ the space $W^{\frac{3}{2},2}(\partial\Omega)$ is denoted by $H^{3/2}(\partial\Omega)$. The space $W^{2-1/p,p}(\partial\Omega)$ is a Banach space for the norm

$$\|p\|_{W^{2-1/p,p}(\partial\Omega)} = \inf\{\|u\|_{W^{2,p}(\Omega)} : \gamma_0 u = p\}.$$

We denote by n the outward unit normal to $\partial\Omega$.

Theorem 1.28 *The restriction mapping*

$$u \in \mathcal{C}^\infty(\bar{\Omega}) \mapsto \frac{\partial u}{\partial n} = \nabla u \cdot n \in L^p(\partial\Omega)$$

can be extended by continuity and density into a linear and continuous mapping

$$\gamma_1 : W^{2,p}(\Omega) \to W^{1-1/p,p}(\partial\Omega)$$

called trace operator of order 1*. Moreover, the map*

$$(\gamma_0, \gamma_1) : u \in W^{2,p}(\Omega) \to (\gamma_0 u, \gamma_1 u) \in W^{2-1/p,p}(\partial\Omega) \times W^{1-1/p,p}(\partial\Omega)$$

is linear, continuous and surjective with kernel

$$\ker(\gamma_0, \gamma_1) = W_0^{2,p}(\Omega).$$

1.2.7 Dual Sobolev spaces

We limit ourselves to $p = 2$. For $k \in \mathbb{N}^*$ we define

$$H^{-k}(\Omega) = \text{ the continuous dual of } H_0^k(\Omega).$$

We have the canonical embedding $L^2(\Omega) \hookrightarrow H^{-k}(\Omega)$ by

$$\langle u, \varphi \rangle = \int_\Omega u\varphi dx \qquad \forall \varphi \in H_0^k(\Omega).$$

Given now $L \in H^{-k}(\Omega)$, if there exists $u \in L^2(\Omega)$ such that

$$\langle L, \varphi \rangle = \int_\Omega u\varphi dx \qquad \forall \varphi \in H^k(\Omega),$$

then this u is unique and it represents canonically L. In this case we identify L with u.

As to trace spaces we define for Ω bounded and of class $\mathcal{C}^1$

$$\begin{aligned} H^{-1/2}(\partial\Omega) &= \text{ the continuous dual of } H^{1/2}(\partial\Omega), \\ H^{-3/2}(\partial\Omega) &= \text{ the continuous dual of } H^{3/2}(\partial\Omega). \end{aligned}$$

Similarly we have the embeddings

$$L^2(\partial\Omega) \hookrightarrow H^{-1/2}(\partial\Omega) \hookrightarrow H^{-3/2}(\partial\Omega).$$

1.3 Elliptic boundary value problems

We refer to [6, 8, 10].

1.3.1 Green's formula

The Green (or integration by parts) formula for smooth functions extends by continuity to H^1 functions as follows.

Theorem 1.29 *Suppose that Ω is an open and bounded subset of $\mathbb{R}^N$ of class $\mathcal{C}^1$, with outward unit normal n. For all $u, v \in H^1(\Omega)$ we have*

$$\int_\Omega \frac{\partial u}{\partial x_i} v dx = -\int_\Omega u \frac{\partial v}{\partial x_i} dx + \int_{\partial\Omega} \gamma_0(u)\gamma_0(v) n_i ds, \qquad i = 1, \cdots, N.$$

Applying this formula to each component U_i of $U \in H^1(\Omega)^N$ and summing over i yields the perhaps more classical formula

$$\int_\Omega \operatorname{div} U v dx = -\int_\Omega U \cdot \nabla v dx + \int_{\partial\Omega} \gamma_0(U) \cdot n \gamma_0(v) ds. \tag{1.2}$$

In particular, taking $v = 1$ and extending by continuity yields for all $U \in W^{1,1}(\Omega)^N$

$$\int_\Omega \operatorname{div} U dx = \int_{\partial\Omega} \gamma_0(U) \cdot n ds. \tag{1.3}$$

Also, if $U = \nabla u$, $u \in H^2(\Omega)$, then

$$\int_\Omega \Delta u v dx = -\int_\Omega \nabla u \cdot \nabla v dx + \int_{\partial\Omega} \gamma_0(\nabla u) \cdot n \gamma_0(v) ds. \tag{1.4}$$

If Ω is of class $\mathcal{C}^2$ then by construction $\gamma_0(\nabla u) \cdot n = \gamma_1(u)$.

1.3.2 Weak and strong formulations of the Poisson problem

Let Ω be an open, bounded and connected subset of $\mathbb{R}^N$.

We begin with the Dirichlet problem. Our aim is to give a precise meaning to the boundary value problem

$$\begin{cases} -\Delta u = f & \text{in } \Omega \\ u = h & \text{on } \partial\Omega. \end{cases} \tag{1.5}$$

There are several possibilities. We will focus on two of them.

Definition 1.30 *Suppose that $f \in L^2(\Omega)$ and $h \in H^{3/2}(\partial\Omega)$, with Ω of class $\mathcal{C}^2$. A strong solution of* (1.5) *is a function $u \in H^2(\Omega)$ such that $-\Delta u = f$ in Ω and $\gamma_0 u = h$.*

This definition is natural but it is not well suited to existence theories. We usually prefer the concept of weak solution.

Definition 1.31 *Suppose that $f \in H^{-1}(\Omega)$ and $h \in H^{1/2}(\partial\Omega)$, with Ω of class $\mathcal{C}^1$. A weak solution of* (1.5) *is a function $u \in H^1(\Omega)$ such that $\gamma_0 u = h$ and*

$$\int_\Omega \nabla u \cdot \nabla \varphi dx = \int_\Omega f \varphi dx \qquad \forall \varphi \in H_0^1(\Omega). \tag{1.6}$$

By abuse of notation but for the sake of readability we have denoted the duality pairing on $H^{-1}(\Omega)$ by an integral. This is only a notation, which we will constantly use throughout these notes.

By the Green formula (1.4), a strong solutions is a weak solution. Conversely, a weak solution which is in $H^2(\Omega)$ is a strong solution.

Note that the concept of weak solution requires less regularity assumptions on the data and the domain. If $h = 0$, then the solution can be directly sought in $H_0^1(\Omega)$, hence it is even not needed to make any regularity assumption on Ω.

We now turn to the mixed Dirichlet - Neumann problem:

$$\begin{cases} -\Delta u = f & \text{in } \Omega \\ u = h & \text{on } \Gamma_D \\ \dfrac{\partial u}{\partial n} = g & \text{on } \Gamma_N, \end{cases} \tag{1.7}$$

where $\partial\Omega = \Gamma_D \cup \Gamma_N$, $\Gamma_D \cap \Gamma_N = \emptyset$.

Definition 1.32 *Suppose that $f \in L^2(\Omega)$, $g \in H^{1/2}(\partial\Omega)$ and $h \in H^{3/2}(\partial\Omega)$, with Ω of class $\mathcal{C}^2$. A strong solution of* (1.7) *is a function $u \in H^2(\Omega)$ such that $-\Delta u = f$ in Ω, $\gamma_1 u = g$ on Γ_N and $\gamma_0(u) = h$ on Γ_D.*

Definition 1.33 *Suppose that $f \in L^2(\Omega)$, $g \in H^{-1/2}(\partial\Omega)$ and $h \in H^{1/2}(\partial\Omega)$, with Ω of class $\mathcal{C}^1$. A weak solution of* (1.7) *is a function $u \in H^1(\Omega)$ such that $\gamma_0 u = h$ on Γ_D and*

$$\int_\Omega \nabla u \cdot \nabla \varphi dx = \int_\Omega f \varphi dx + \int_{\partial\Omega} g \gamma_0(\varphi) ds \qquad \forall \varphi \in H^1(\Omega) \text{ s.t. } \gamma_0 \varphi = 0 \text{ on } \Gamma_D. \tag{1.8}$$

By the Green formula, a strong solution is a weak solution, and a weak solution which is in $H^2(\Omega)$ is a strong solution.

1.3.3 Existence and uniqueness of solutions

The classical approach to prove the existence and uniqueness of a solution of an elliptic boundary value problem relies on the Lax-Milgram theorem.

Theorem 1.34 (Lax-Milgram) *Let H be a Hilbert space, $a(\cdot,\cdot)$ be a bilinear and continuous form on H and $b \in H'$ (the continuous dual space of H). We assume that a is coercive, namely there exists $\alpha > 0$ such that*

$$a(u,u) \geq \alpha \|u\|^2 \qquad \forall u \in H.$$

Then there exists a unique $u \in H$ such that

$$a(u,v) = \langle b, v \rangle \qquad \forall v \in H.$$

As first example consider the problem

$$\begin{cases} -\Delta u + u = f & \text{in } \Omega \\ u = h & \text{on } \partial\Omega, \end{cases}$$

with $f \in H^{-1}(\Omega)$, $h \in H^{1/2}(\partial\Omega)$, whose weak formulation is: find $u \in H^1(\Omega)$ such that $\gamma_0 u = h$ and

$$\int_\Omega (\nabla u \cdot \nabla \varphi + u\varphi) dx = \langle f, \varphi \rangle \qquad \forall \varphi \in H_0^1(\Omega).$$

Observe that it is sufficient here to suppose $f \in H^{-1}(\Omega)$ since test functions belong to $H_0^1(\Omega)$. To show the existence and uniqueness of a weak solution we proceed in two steps. First we "lift" the Dirichlet boundary condition, by surjectivity of the trace operator: let $u_1 \in H^1(\Omega)$ be such that $\gamma_0 u_1 = h$. Then $u = u_1 + u_2$ will be a solution if and only if $u_2 \in H_0^1(\Omega)$ and

$$\int_\Omega (\nabla u_2 \cdot \nabla \varphi + u_2\varphi) dx = \langle f, \varphi \rangle - \int_\Omega (\nabla u_1 \cdot \nabla \varphi - u_1\varphi) dx \qquad \forall \varphi \in H_0^1(\Omega).$$

Lax-Milgram's theorem ensures the existence of such u_2. For the uniqueness we suppose that u and u' are two weak solutions and we set $\hat{u} = u - u'$. We have $\hat{u} \in H_0^1(\Omega)$ and

$$\int_\Omega (\nabla \hat{u} \cdot \nabla \varphi + \hat{u}\varphi) dx = 0 \qquad \forall \varphi \in H_0^1(\Omega).$$

Choosing $\varphi = \hat{u}$ yields $\hat{u} = 0$.

Now consider the map

$$\begin{array}{rcl} \Lambda : H^1(\Omega) & \to & H^{-1}(\Omega) \times H^{1/2}(\partial\Omega) \\ u & \mapsto & \left([\varphi \mapsto \int_\Omega (\nabla u \cdot \nabla \varphi + u\varphi) dx], \gamma_0 u \right). \end{array}$$

Obviously it is a linear and continuous map and we have just shown that it is bijective. By the open mapping theorem, Λ^{-1} is continuous. This means that the weak solution satisfies

$$\|u\|_{H^1(\Omega)} \leq c \left(\|f\|_{H^{-1}(\Omega)} + \|h\|_{H^{1/2}(\partial\Omega)} \right),$$

for some constant $c > 0$. In particular, we have shown the following (choose $f = 0$):

Corollary 1.35 *The trace operator $\gamma_0 : H^1(\Omega) \to H^{1/2}(\partial\Omega)$ admits a linear and continuous right inverse.*

In order to apply Theorem 1.34 to the weak formulations (1.6) and (1.8), an important ingredient is missing in order to prove the coercivity. It is the Poincaré inequality, which may take several forms. The most classical one states that if Ω is bounded then there exists a constant $C_P > 0$ such that

$$\|u\|_{L^2(\Omega)} \leq C_P \|\nabla u\|_{L^2(\Omega)} \qquad \forall u \in H_0^1(\Omega).$$

This permits to deal with the Dirichlet problem, but not with the mixed problem. Here is a more general version.

Theorem 1.36 *Let Ω be a bounded open subset of $\mathbb{R}^N$ of class $\mathcal{C}^1$, and H be a closed subspace of $H^1(\Omega)$. Let $\|\cdot\|$ be a norm on H such that, for some constants $c_1, c_2 > 0$,*

$$c_1 \|\nabla u\|_{L^2(\Omega)} \leq \|u\| \leq c_2 \|u\|_{H^1(\Omega)} \qquad \forall u \in H.$$

Then the norm $\|\cdot\|$ is equivalent to the norm $\|\cdot\|_{H^1(\Omega)}$ on H.

PROOF. We argue by contradiction, assuming that there exists no $c > 0$ such that $\|u\| \geq c\|u\|_{L^2(\Omega)}$ for all $u \in H$. Therefore, we can construct a sequence (u_k) of nonzero elements of H such that

$$\lim_{k\to+\infty} \frac{\|u_k\|_{L^2(\Omega)}}{\|u_k\|} = +\infty.$$

We set $v_k = u_k/\|u_k\|_{L^2(\Omega)}$, so that

$$\|v_k\|_{L^2(\Omega)} = 1, \qquad \lim_{k\to+\infty} \|v_k\| = 0.$$

We infer from the assumptions that $\lim_{k\to+\infty} \|\nabla v_k\|_{L^2(\Omega)} = 0$. In particular $(\|v_k\|_{H^1(\Omega)})$ is bounded. By Theorem 1.27 we can extract a subsequence $(v_{\iota(k)})$ such that

$$\lim_{k\to+\infty} \|v_{\iota(k)} - v\|_{L^2(\Omega)} = 0 \qquad \text{for some } v \in L^2(\Omega).$$

In particular

$$\|v\|_{L^2(\Omega)} = \lim_{k\to+\infty} \|v_{\iota(k)}\|_{L^2(\Omega)} = 1. \tag{1.9}$$

Moreover we have for all $\varphi \in H_0^1(\Omega)^N$

$$-\int_\Omega v \operatorname{div} \varphi dx = \lim_{k\to+\infty} -\int_\Omega v_{\iota(k)} \operatorname{div} \varphi dx = \lim_{k\to+\infty} \int_\Omega \nabla v_{\iota(k)} \cdot \varphi dx = 0.$$

We recognize that v is weakly differentiable with $\nabla v = 0$. In particular $v \in H^1(\Omega)$ and

$$\lim_{k\to+\infty} \|\nabla v_{\iota(k)} - \nabla v\|_{L^2(\Omega)} = \lim_{k\to+\infty} \|\nabla v_{\iota(k)}\|_{L^2(\Omega)} = 0.$$

We arrive at

$$\lim_{k\to+\infty} \|v_{\iota(k)} - v\|_{H^1(\Omega)} = 0.$$

Since H is closed this yields $v \in H$, and from the assumptions

$$\lim_{k\to+\infty} \|v_{\iota(k)} - v\| = 0.$$

Hence

$$\|v\| = \lim_{k\to+\infty} \|v_{\iota(k)}\| = 0.$$

This implies that $v = 0$, which contradicts (1.9). □

Let us now show the existence and uniqueness of a solution for the mixed problem (of which the Dirichlet problem is a particular case.

Proposition 1.37 *Suppose that $f \in L^2(\Omega)$, $g \in H^{-1/2}(\partial\Omega)$ and $h \in H^{1/2}(\partial\Omega)$, with Ω an open, bounded, connected subset of $\mathbb{R}^N$ of class $\mathcal{C}^1$ and Γ_D of nonzero measure. There exists a unique weak solution u to* (1.7). *Moreover there exists a constant $c > 0$, depending only on the geometric data, such that*

$$\|u\|_{H^1(\Omega)} \leq c\left(\|f\|_{L^2(\Omega)} + \|g\|_{H^{-1/2}(\partial\Omega)} + \|h\|_{H^{1/2}(\partial\Omega)}\right).$$

PROOF. We first prove existence. The first step is to "lift" the Dirichlet condition: let $u_1 \in H^1(\Omega)$ be such that $\gamma_0 u_1 = h$. Next, in order to obtain a weak solution decomposed as $u = u_1 + u_2$, we need to find

$$u_2 \in H := \left\{v \in H^1(\Omega) : \gamma_0(v) = 0 \text{ on } \Gamma_D\right\} \tag{1.10}$$

such that

$$\int_\Omega \nabla u_2 \cdot \nabla\varphi dx = \int_\Omega f\varphi dx + \int_{\partial\Omega} g\gamma_0(\varphi)ds - \int_\Omega \nabla u_1 \cdot \nabla\varphi dx \qquad \forall\varphi \in H. \tag{1.11}$$

By continuity of γ_0, H is a closed subspace of $H^1(\Omega)$. Moreover, since Γ_D has nonzero measure and Ω is connected, it is immediately seen that the map $v \mapsto \|\nabla v\|_{L^2(\Omega)}$ is a norm on H. By Theorem 1.36 it is equivalent to the norm $\|\cdot\|_{H^1(\Omega)}$. This shows the coercivity of the bilinear form

$$(v, w) \in H \times H \mapsto \int_\Omega \nabla v \cdot \nabla w dx.$$

Lax-Milgram's theorem yields the existence of u_2.

We turn to uniqueness. Suppose that u, u' are two solutions and set $\hat{u} = u - u'$. We have $\hat{u} \in H$ and

$$\int_\Omega \nabla \hat{u} \cdot \nabla \varphi dx = 0 \qquad \forall \varphi \in H.$$

Choosing $\varphi = \hat{u}$ yields $\hat{u} = 0$ by coercivity.

Finally we prove the Lipschitz-continuous dependence on the data. We consider the decomposition $u = u_1 + u_2$ described above. By Corollary 1.35, there exists $c_1 > 0$ such that $\|u_1\|_{H^1(\Omega)} \leq c_1 \|h\|_{H^{1/2}(\partial\Omega)}$. Choosing $\varphi = u_2$ in (1.11) yields by coercivity

$$\|u_2\|_{H^1(\Omega)} \leq c_2 \left(\|f\|_{L^2(\Omega)} + \|g\|_{H^{-1/2}(\partial\Omega)} + \|u_1\|_{H^1(\Omega)} \right).$$

Then it suffices to combine the two inequalities above. □

Remark 1.38 *For the Neumann problem ($\partial\Omega = \Gamma_N$), in order to obtain coercivity, we set the weak formulation (both for the unknown and the test function) in the space*

$$H = \left\{ v \in H^1(\Omega) : \int_\Omega v dx = 0 \right\}.$$

To retrieve the correspondence between strong and weak solutions, one has to assume the equilibrium condition

$$\int_\Omega f dx + \int_{\partial\Omega} g ds = 0.$$

Alternatively, we can work in the quotient space $H^1(\Omega)/\mathbb{R}$, under the same equilibrium condition needed to have the linear form well-defined.

1.3.4 The linear elasticity system

A typical linear elasticity problem is to find a displacemend field $u : \Omega \to \mathbb{R}^N$ (where Ω is an open and bounded subset of $\mathbb{R}^N$, $N = 2, 3$) solution of

$$\begin{cases} -\operatorname{div} \sigma(u) = f & \text{in } \Omega \\ u = h & \text{on } \Gamma_D \\ \sigma(u) n = g & \text{on } \Gamma_N, \end{cases} \tag{1.12}$$

where $\partial\Omega = \Gamma_D \cup \Gamma_N$, $\Gamma_D \cap \Gamma_N = \emptyset$. The stress field $\sigma(u)$ is a symmetric $N \times N$ matrix field. The divergence is computed row-wise. In the linear setting the stress depend linearly on the linearized strain

$$e(u) = \frac{1}{2}(\nabla u + \nabla^\top u).$$

The gradient $\nabla u = Du$ is the Jacobian matrix of u (the rows are the gradients of the components of u). We denote by A the elasticity tensor (or Hooke's tensor) connecting the stress to the strain by

$$\sigma(u) = Ae(u) \qquad \text{i.e. } (\sigma(u))_{ij} = \sum_{kl} A_{ijkl}(e(u))_{kl}.$$

Here we assume that A is constant over Ω. Moreover we assume an isotropic constitutive law, namely

$$\sigma(u) = Ae(u) = \lambda \operatorname{tr} e(u) I + 2\mu e(u) \tag{1.13}$$

where $\lambda \geq 0, \mu > 0$ are the Lamé coefficients. It then immediately found that the eigenvalues of A are 2μ and $\kappa := \lambda N + 2\mu$ (μ is also called shear modulus and κ is the bulk modulus). It follows that A is symmetric ($Ae : e' = e : Ae'$) positive definite ($Ae : e > 0 \ \forall e \neq 0$), with more specifically

$$Ae : e \geq 2\mu |e|^2 \qquad \forall e \in \mathcal{S}_N(\mathbb{R})$$

($|e| = \sqrt{e : e}$ is the Frobenius norm). From the modeling standpoint, the Lamé coefficients are related to the Young modulus E and the Poisson ratio ν of the material by the formulas

$$\mu = \frac{E}{2(1+\nu)}, \quad \begin{array}{ll} \lambda = \dfrac{E\nu}{(1+\nu)(1-2\nu)} & \text{in 3D and plane strain} \\ \lambda = \dfrac{E\nu}{(1+\nu)(1-\nu)} & \text{in plane stress.} \end{array} \tag{1.14}$$

Definition 1.39 *Suppose that $f \in L^2(\Omega)^N$, $g \in H^{1/2}(\partial\Omega)^N$ and $h \in H^{3/2}(\partial\Omega)^N$, with Ω of class $\mathcal{C}^2$. A strong solution of* (1.12) *is a function $u \in H^2(\Omega)^N$ such that $-\operatorname{div}\sigma(u) = f$ in Ω, $\gamma_0(\sigma(u))n = g$ on Γ_N and $\gamma_0(u) = h$ on Γ_D.*

Definition 1.40 *Suppose that $f \in L^2(\Omega)^N$, $g \in H^{-1/2}(\partial\Omega)^N$ and $h \in H^{1/2}(\partial\Omega)^N$, with Ω of class $\mathcal{C}^1$. A weak solution of* (1.12) *is a function $u \in H^1(\Omega)^N$ such that $\gamma_0(u) = h$ on Γ_D and*

$$\int_\Omega \sigma(u) : e(v) dx = \int_\Omega f \cdot v dx + \int_{\partial\Omega} g \cdot \gamma_0(v) ds \qquad \forall v \in H^1(\Omega)^N \ \text{ s.t. } \gamma_0(v) = 0 \ \text{ on } \Gamma_D. \tag{1.15}$$

Due to the symmetry of $\sigma(u)$, we have $\sigma(u) : e(u) = \sigma(u) : \nabla u$. This remark leads to the following variant of the Green formula: if Ω is of class $\mathcal{C}^1$ then for all $(u, v) \in H^2(\Omega) \times H^1(\Omega)$ we have

$$\int_\Omega \operatorname{div}\sigma(u) \cdot v dx = -\int_\Omega \sigma(u) : e(v) + \int_{\partial\Omega} (\gamma_0(\sigma(u))n) \cdot \gamma_0(v) ds.$$

This shows that a strong solution is a weak solution, and that a weak solution which is in $H^2(\Omega)^N$ is a strong solution.

In order to prove the existence and uniqueness of a weak solution we will make use of a variant of Korn's inequality which will provide a counterpart of Poincaré's inequality. We refer to [10] for the following version of Korn's inequality.

Theorem 1.41 (Korn's inequality in $H^1(\Omega)$) *Let Ω be an open, bounded, connected subset of $\mathbb{R}^N$ with Lipschitz boundary. If $u \in L^2(\Omega)^N$ and $e(u) \in L^2(\Omega)^{N\times N}$ then $u \in H^1(\Omega)^N$ and there exists a geometric constant $c > 0$ such that*

$$\|u\|_{H^1} \leq c(\|u\|_{L^2} + \|e(u)\|_{L^2}).$$

Lemma 1.42 *Let Ω be a connected, open subset of $\mathbb{R}^N$. If $u \in H^1(\Omega)^N$ satisfies $e(u) = 0$ then there exists a skew-symmetric matrix R and a vector $b \in \mathbb{R}^N$ such that*

$$u(x) = Rx + b \qquad \forall x \in \Omega.$$

We say that u is an infinnitesimal rigid body displacement field.

Proof. By definition of the weak derivative, an immediate calculation shows that for all $\varphi \in \mathcal{C}^2_c(\Omega)$

$$0 = \int_\Omega \left(e_{ij}(u)\frac{\partial\varphi}{\partial x_k} + e_{ik}(u)\frac{\partial\varphi}{\partial x_j} - e_{jk}(u)\frac{\partial\varphi}{\partial x_i} \right) dx = -\int_\Omega u_i \frac{\partial^2\varphi}{\partial x_j \partial x_k} dx = \int_\Omega \frac{\partial u_i}{\partial x_j}\frac{\partial\varphi}{\partial x_k} dx.$$

Hence by Theorem 1.18 there exist constants R_{ij} such that

$$\frac{\partial u_i}{\partial x_j} = R_{ij}.$$

The condition $e(u) = 0$ yields $R_{ij} = -R_{ij}$, i.e. the matrix R is skew-symmetric. Set $w(x) = Rx$. We have

$$\frac{\partial w_i}{\partial x_j} = R_{ij} = \frac{\partial u_i}{\partial x_j},$$

hence for all $\varphi \in \mathcal{C}_c^1(\Omega)$

$$\int_\Omega u_i \frac{\partial \varphi}{\partial x_j} dx = -\int_\Omega \frac{\partial u_i}{\partial x_j} \varphi dx = -\int_\Omega \frac{\partial w_i}{\partial x_j} \varphi dx = \int_\Omega w_i \frac{\partial \varphi}{\partial x_j} dx.$$

This shows that $u_i - w_i = b_i$ for some constant b_i. □

We denote by $\mathcal{R}$ the set of infinitesimal rigid body displacements, namely

$$\mathcal{R} = \{x \mapsto Rx + b,\ b \in \mathbb{R}^N, R \text{ skew-symmetric}\}.$$

Equivalently, in dimension $N = 3$,

$$\mathcal{R} = \{x \mapsto \omega \wedge x + b,\ \omega, b \in \mathbb{R}^3\}.$$

The same holds in dimension 2 with ω restricted to the antiplane direction.

Theorem 1.43 *Let Ω be a bounded, connected, open subset of $\mathbb{R}^N$ of class $\mathcal{C}^1$ and let Γ_D be a subset of $\partial\Omega$ of nonzero measure. Let*

$$H = \left\{u \in H^1(\Omega) : \gamma_0(u) = 0 \ \textit{on}\ \Gamma_D\right\}.$$

There exists $C_K > 0$ such that

$$\|u\|_{H^1(\Omega)^N} \le C_K \|e(u)\|_{L^2(\Omega)^{N\times N}} \qquad \forall u \in H.$$

Proof. It follows from a slight modification of the proof of Theorem 1.36, using Theorem 1.41 and Lemma 1.42 and setting $\|u\| = \|e(u)\|_{L^2(\Omega)^{N\times N}}$. The details are left to the reader, but the key point is to show that $\|\cdot\|$ is a norm. Thus let us assume that $e(u) = 0$. Then $u(x) = Rx + b$ for some skew-symmetric matrix R and some vector b. Let $x_0 \in \Gamma_D$. Since $\partial\Omega$ is a $\mathcal{C}^1$ manifold of dimension $N-1$ we can find $x_1, \cdots, x_{N-1} \in \Gamma_D$ such that the vectors $x_1 - x_0, \cdots, x_{N-1} - x_0$ are linearly independent. We have

$$u(x_i) = Rx_i + b = 0 \qquad \forall i = 0, \cdots, N-1$$

hence

$$R(x_i - x_0) = 0 \qquad \forall i = 1, \cdots, N-1.$$

This shows that $\dim\ker R \ge N-1$. If $\dim\ker R = N-1$ then $\operatorname{rank}(R) = 1$, which is not possible because R is skew-symmetric. Therefore $R = 0$ and subsequently $b = 0$, hence $u = 0$. □

We arrive at the following existence and uniqueness result. The proof is an adaptation of Proposition 1.37, using Theorem 1.43.

Proposition 1.44 *Suppose that $f \in L^2(\Omega)^N$, $g \in H^{-1/2}(\partial\Omega)^N$ and $h \in H^{1/2}(\partial\Omega)^N$, with Ω an open, bounded, connected subset of $\mathbb{R}^N$ of class $\mathcal{C}^1$ and Γ_D of nonzero measure. There exists a unique weak solution u to* (1.12). *Moreover there exists a constant $c > 0$, depending only on the geometric data, such that*

$$\|u\|_{H^1(\Omega)^N} \le c\left(\|f\|_{L^2(\Omega)^N} + \|g\|_{H^{-1/2}(\partial\Omega)^N} + \|h\|_{H^{1/2}(\partial\Omega)^N}\right).$$

Remark 1.45 *For the full Neumann case ($\partial\Omega = \Gamma_N$), through working with the quotient space $H^1(\Omega)/\mathcal{R}$, one obtains existence and uniqueness up to infinitesimal rigid body displacements, under the equilibrium condition*

$$\int_\Omega f \cdot w dx + \int_{\partial\Omega} g \cdot w ds = 0 \qquad \forall w \in \mathcal{R},$$

itself equivalent to

$$\int_\Omega f dx + \int_{\partial\Omega} g ds = 0 \ \textit{and} \ \int_\Omega f \wedge x dx + \int_{\partial\Omega} g \wedge x ds = 0$$

(equilibrium of forces and moments).

1.3.5 Variational principles

Variational principles aim at characterizing weak solutions of boundary value problems as solutions of optimization problems. This approach has many advantages but it is not always possible. We will present primal and dual variational principles.

Proposition 1.46 *Let $\mathcal{H}$ be a Hilbert space, H be a closed linear subspace of $\mathcal{H}$, $w \in \mathcal{H}$, $a(\cdot,\cdot)$ be a continuous and symmetric bilinear form on $\mathcal{H}$ coercive on H and $b \in \mathcal{H}'$. The "primal energy" functional*

$$\mathcal{E} : v \in \mathcal{H} \mapsto \frac{1}{2}a(v,v) - \langle b, v\rangle$$

admits a unique minimizer u over the affine space $\{w\} + H$. It satisfies

$$a(u,\varphi) = \langle b, \varphi\rangle \qquad \forall v \in H. \tag{1.16}$$

Proof. Let $u \in \{w\} + H$ be the unique solution of (1.16) (by Lax-Milgram's theorem for $u - w$). Consider an arbitrary $v \in \{w\} + H$. We have

$$\mathcal{E}(v) - \mathcal{E}(u) = \frac{1}{2}a(v-u,v-u) + a(u,v-u) - \langle b, v-u\rangle = \frac{1}{2}a(v-u,v-u) \geq 0.$$

It remains to show that the minimizer is unique. Let $\tilde{u}$ be an arbitrary minimizer. From the previous calculation and $\mathcal{E}(\tilde{u}) = \mathcal{E}(u)$ we infer $a(\tilde{u}-u,\tilde{u}-u) = 0$, hence $\tilde{u} = u$ by coercivity. □

The equality (1.16) is the first order optimality condition (vanishing of the derivative, called Euler-Lagrange equation) for the primal energy. This is why the weak formulation (1.16) is also called **variational formulation**. The shift w is useful to handle non-homogeneous Dirichlet conditions. Otherwise we simply take $H = \mathcal{H}$ and $w = 0$.

We now turn to the dual principle. We begin with a classical lemma known as weak duality inequality.

Lemma 1.47 *Let U, V be two sets and $L : U \times V \to \mathbb{R}$ be a function. We always have*

$$\sup_{v\in V} \inf_{u\in U} L(u,v) \leq \inf_{u\in U} \sup_{v\in V} L(u,v).$$

Proof. We have

$$\inf_{u\in U} L(u,\hat{v}) \leq L(\hat{u},\hat{v}) \leq \sup_{v\in V} L(\hat{u},v) \qquad \forall(\hat{u},\hat{v}) \in U \times V$$

hence

$$\inf_{u\in U} L(u,\hat{v}) \leq \inf_{u\in U} \sup_{v\in V} L(u,v) \qquad \forall\hat{v} \in V.$$

The conclusion follows. □

Proposition 1.48 *Under the assumptions of Proposition 1.46, we further assume that*

$$a(u,v) = \langle Pu, Pv\rangle_Y$$

where Y is a Hilbert space, $\langle\cdot,\cdot\rangle_Y$ is its inner product and $P \in \mathcal{L}(\mathcal{H}, Y)$. The "dual energy" functional (or complementary energy)

$$\mathcal{E}^* : \tau \in Y \mapsto -\frac{1}{2}\|\tau\|_Y^2 + \langle \tau, Pw\rangle_Y - \langle b, w\rangle$$

admits a unique maximizer τ^* in the set

$$\mathcal{T} = \{\tau \in Y : \langle \tau, Pv\rangle_Y = \langle b, v\rangle \ \forall v \in H\}.$$

It satisfies $\tau^* = Pu$ where $u \in \mathcal{H}$ is the minimizer of the primal energy. In addition we have

$$\min_{v\in\{w\}+H} \mathcal{E}(v) = \max_{\tau\in\mathcal{T}} \mathcal{E}^*(\tau).$$

Proof. We define the Lagrangian

$$L(\tau, v) = \mathcal{E}^*(\tau) + \langle \tau, Pv\rangle_Y - \langle b, v\rangle \qquad \forall (\tau, v) \in Y \times \mathcal{H}.$$

By construction we have

$$\sup_{\tau\in\mathcal{T}} \mathcal{E}^*(\tau) = \sup_{\tau\in Y} \inf_{v\in H} L(\tau, v).$$

Now we write that

$$L(\tau, v) = -\frac{1}{2}\|\tau\|_Y^2 + \langle \tau, Pw\rangle_Y - \langle b, w\rangle + \langle \tau, Pv\rangle_Y - \langle b, v\rangle = \hat{L}(\tau, v+w)$$

with

$$\hat{L}(\tau, \hat{v}) = -\frac{1}{2}\|\tau\|_Y^2 + \langle \tau, P\hat{v}\rangle_Y - \langle b, \hat{v}\rangle.$$

Therefore

$$\sup_{\tau\in\mathcal{T}} \mathcal{E}^*(\tau) = \sup_{\tau\in Y} \inf_{\hat{v}\in\{w\}+H} \hat{L}(\tau, \hat{v}).$$

Now we recognize that

$$\hat{L}(\tau, \hat{v}) = -\frac{1}{2}\|\tau - P\hat{v}\|_Y^2 + \frac{1}{2}\|P\hat{v}\|_Y^2 - \langle b, \hat{v}\rangle = -\frac{1}{2}\|\tau - P\hat{v}\|_Y^2 + \mathcal{E}(\hat{v}).$$

Clearly,

$$\inf_{\hat{v}\in\{w\}+H} \sup_{\tau\in Y} \hat{L}(\tau, \hat{v}) = \inf_{\hat{v}\in\{w\}+H} \mathcal{E}(\hat{v}).$$

The weak duality inequality implies

$$\sup_{\tau\in\mathcal{T}} \mathcal{E}^*(\tau) = \sup_{\tau\in Y} \inf_{\hat{v}\in\{w\}+H} \hat{L}(\tau, \hat{v}) \le \inf_{\hat{v}\in\{w\}+H} \sup_{\tau\in Y} \hat{L}(\tau, \hat{v}) = \inf_{\hat{v}\in\{w\}+H} \mathcal{E}(\hat{v}).$$

Let $u \in \{w\} + H$ be the minimizer of the primal energy and set $\tau^* = Pu$. We have

$$\begin{aligned}\mathcal{E}^*(\tau^*) = -\frac{1}{2}\|Pu\|_Y^2 + \langle Pu, Pw\rangle_Y - \langle b, w\rangle &= -\frac{1}{2}a(u,u) + a(u,w) - \langle b, w\rangle \\ &= \frac{1}{2}a(u,u) - \langle b, u\rangle - a(u, u-w) + \langle b, u-w\rangle = \mathcal{E}(u),\end{aligned}$$

by (1.16), and obviously $\tau^* \in \mathcal{T}$. It follows that

$$\mathcal{E}^*(\tau^*) = \sup_{\tau\in\mathcal{T}} \mathcal{E}^*(\tau) = \inf_{v\in\{w\}+H} \mathcal{E}(v) = \mathcal{E}(u).$$

It remains to show that the maximizer is unique. Let $\tau \in \mathcal{T}$ be an arbitrary maximizer. We have

$$\mathcal{E}^*(\tau) = L(\tau, u-w) = \hat{L}(\tau, u) = -\frac{1}{2}\|\tau - Pu\|_Y^2 + \mathcal{E}(u) = -\frac{1}{2}\|\tau - \tau^*\|_Y^2 + \mathcal{E}^*(\tau^*),$$

and since $\mathcal{E}^*(\tau) = \mathcal{E}^*(\tau^*)$, we infer that $\tau = \tau^*$. □

Let us now give two applications of the preceding results. We first consider the Poisson problem (1.7) with homogeneous Dirichlet condition ($h = 0$). The space H is defined by (1.10), and the bilinear and linear forms by

$$a(u, v) = \int_\Omega \nabla u \cdot \nabla v dx, \qquad \langle b, v\rangle = \int_\Omega f v dx + \int_{\partial\Omega} g\gamma_0(v) ds.$$

Since $h = 0$ we choose $\mathcal{H} = H$ and $w = 0$. We set $Y = L^2(\Omega)^N$ endowed with its canonical inner product and $P = \nabla$. We obtain

$$\mathcal{T} = \left\{\tau \in L^2(\Omega)^N : \int_\Omega \tau \cdot \nabla v dx = \int_\Omega f v dx + \int_{\partial\Omega} g\gamma_0(v) ds \ \forall v \in H\right\}.$$

This is the weak formulation of

$$\mathcal{T}=\left\{\tau\in L^2(\Omega)^N : -\operatorname{div}\tau=f \text{ in } \Omega, \tau\cdot n=g \text{ on } \Gamma_N\right\}.$$

The dual energy is

$$\mathcal{E}^*(\tau)=-\frac{1}{2}\|\tau\|^2_{L^2(\Omega)^N}$$

Remark 1.49 *For the variant where*

$$a(u,v)=\int_\Omega \alpha\nabla u\cdot\nabla v dx, \qquad \alpha\in L^\infty(\Omega), \alpha\geq\underline{\alpha}>0$$

we define $Pv=\alpha\nabla v$ and equip $Y=L^2(\Omega)^N$ with the inner product

$$\langle\tau,\hat{\tau}\rangle_Y=\int_\Omega \alpha^{-1}\tau\cdot\hat{\tau}dx.$$

By simplification this leads to the same set $\mathcal{T}$, but the dual energy becomes

$$\mathcal{E}^*(\tau)=-\frac{1}{2}\int_\Omega \alpha^{-1}|\tau|^2dx.$$

Let us now consider the elasticity problem (1.12), again with $h=0$. Here

$$H=\left\{v\in L^2(\Omega)^N : \gamma_0(v)=0 \text{ on } \Gamma_D\right\},$$

$$a(u,v)=\int_\Omega \sigma(u):e(v)dx=\int_\Omega Ae(u):e(v)dx, \qquad \langle b,v\rangle=\int_\Omega f\cdot vdx+\int_{\partial\Omega} g\cdot\gamma_0(v)ds.$$

We define $Y=L^2(\Omega,\mathcal{S}_N(\mathbb{R}))$ endowed with the inner product

$$\langle\tau,\hat{\tau}\rangle_Y=\int_\Omega A^{-1}\tau:\hat{\tau}dx$$

and $Pv=\sigma(v)=Ae(v)$. We obtain

$$\mathcal{T}=\left\{\tau\in L^2(\Omega,\mathcal{S}_N(\mathbb{R})) : \int_\Omega \tau:e(v)dx=\int_\Omega f\cdot vdx+\int_{\partial\Omega} g\cdot\gamma_0(v)ds\ \forall v\in H\right\},$$

which can be rewritten in strong form as

$$\mathcal{T}=\left\{\tau\in L^2(\Omega,\mathcal{S}_N(\mathbb{R})) : -\operatorname{div}\tau=f \text{ in } \Omega, \tau n=g \text{ on } \Gamma_N\right\}.$$

The dual energy is

$$\mathcal{E}^*(\tau)=-\frac{1}{2}\int_\Omega A^{-1}\tau:\tau dx.$$

1.4 The direct method of the calculus of variations

1.4.1 Problem statement

We recall here in the general setting the method of converging minimizing sequences very often used to prove the existence of solutions to optimization problems. Note that there are also non-sequential approaches, usually based on general compactness. As maximization problems are transformed into minimization ones by change of sign, we focus on minimization. We consider the problem

$$\underset{x\in X}{\operatorname{minimize}}\ f(x) \tag{1.17}$$

where X is a topological space and $f : X \to \mathbb{R} \cup \{+\infty\}$ is the cost function (or objective function, or criterion). This problem may incorporate constraints throught setting $f(x) = j(x) + I_{\mathcal{U}}(x)$, where j is the original cost and $I_{\mathcal{U}}$ is the indicator function of the admissible set defined by

$$I_{\mathcal{U}}(x) = \begin{cases} 0 & \text{if } x \in \mathcal{U} \\ +\infty & \text{otherwise.} \end{cases}$$

Solving (1.17) not only means finding the value of the minimum (or infimum), but also finding minimizers, if there are some, i.e. points where the minimum is attained. Let us recall a few basic definitions.

Definition 1.50 *We say that f is proper if $f(x)$ is not everywhere equal to $+\infty$.*

Definition 1.51 *We say that $z \in X$ is a global minimizer of f if*

$$f(z) \leq f(x) \qquad \forall x \in X.$$

Definition 1.52 *We say that $z \in X$ is a local minimizer of f if $f(z) \in \mathbb{R}$ and there exists a neighborhood $\mathcal{N}$ of z in X such that*

$$f(z) \leq f(x) \qquad \forall x \in \mathcal{N}.$$

By definition the problem stated in (1.17) deals with global minima. However, it is important to keep in mind that optimization algorithms often find local minimizers, and get stuck there.

1.4.2 Lower semicontinuity and inf-compactness

Definition 1.53 *Let $\gamma \in \mathbb{R} \cup \{+\infty\}$. The γ-level set of f is the set*

$$\operatorname{lev}_\gamma f = \{x \in X \ \textit{s.t.}\ f(x) \leq \gamma\}.$$

Definition 1.54 *We say that $f : X \to \mathbb{R} \cup \{+\infty\}$ is (sequentially) lower semicontinuous if for all sequence (x_n) of elements of X such that*

$$\lim_{n\to+\infty} x_n = x \in X \qquad \textit{and} \qquad \lim_{n\to+\infty} f(x_n) = y \in \mathbb{R} \cup \{-\infty, +\infty\}$$

it holds $f(x) \leq y$. This is equivalent to

$$\lim_{n\to+\infty} x_n = x \Rightarrow f(x) \leq \liminf_{n\to+\infty} f(x_n).$$

Of course, if $f : X \to \mathbb{R}$ is continuous then it is lower semicontinuous.

Proposition 1.55 *$f : X \to \mathbb{R} \cup \{+\infty\}$ is lower semicontinuous if and only if $\operatorname{lev}_\gamma f$ is sequentially closed for all $\gamma \in \mathbb{R}$.*

Proof. Suppose first that f is lower semicontinuous and consider a sequence (x_n) of $\operatorname{lev}_\gamma f$, for some $\gamma \in \mathbb{R}$, such that $\lim_{n\to+\infty} x_n = x \in X$. Set $y_n = f(x_n) \leq \gamma$. We have for a non-relabeled subsequence $\lim_{n\to+\infty} y_n = y \in [-\infty, \gamma]$, hence $f(x) \leq y \leq \gamma$. This means that $x \in \operatorname{lev}_\gamma f$. Assume now that $\operatorname{lev}_\gamma f$ is sequentially closed for all $\gamma \in \mathbb{R}$. Suppose that there exists a sequence (x_n) of elements of X such that

$$\lim_{n\to+\infty} x_n = x \in X \qquad \text{and} \qquad \lim_{n\to+\infty} f(x_n) = y < f(x).$$

Let $\gamma \in]y, f(x)[$. For n large enough we have $f(x_n) \leq \gamma$, i.e. $x_n \in \operatorname{lev}_\gamma f$. Since $\operatorname{lev}_\gamma f$ is sequentially closed we infer that $x \in \operatorname{lev}_\gamma f$. This contradicts $\gamma < f(x)$. □

Definition 1.56 *The function $f : X \to \mathbb{R} \cup \{+\infty\}$ is said to be inf-compact if for all $\gamma \in \mathbb{R}$ the level set $\operatorname{lev}_\gamma f$ either is empty or has sequentially compact closure in X.*

1.4.3 Existence of minimizers

Definition 1.57 *We call minimizing sequence of f a sequence (x_n) of elements of X such that $\lim_{n\to+\infty} f(x_n) = \inf_X f$.*

By definition of the infimum, minimizing sequences always exist.

Theorem 1.58 *Let $f : X \to \mathbb{R}\cup\{+\infty\}$ be a proper, lower semicontinuous, inf-compact function. Then f admits (at least) a global minimizer.*

Proof. Let (x_n) be a minimizing sequence. By definition we have

$$\lim_{n\to+\infty} f(x_n) = \inf_{x\in X} f(x) \in \mathbb{R}\cup\{-\infty\}.$$

The sequence $(f(x_n))$ is bounded from above for $n \geq n_0$, hence there exists $\gamma \in \mathbb{R}$ such that

$$x_n \in \operatorname{lev}_\gamma f \qquad \forall n \geq n_0.$$

By sequential compactness of $\overline{\operatorname{lev}_\gamma f}$, there exists $\hat{x} \in \overline{\operatorname{lev}_\gamma f}$ such that

$$\lim_{n\to+\infty} x_n = \hat{x},$$

for a non-relabelled subsequence. By lower-semicontinuity we have $f(\hat{x}) \leq \inf_{x\in X} f(x)$. We conclude that $f(\hat{x}) = \inf_{x\in X} f(x) = \min_{x\in X} f(x)$. □

Inf-compactness is often related to the property

$$\lim_{\|x\|\to+\infty} f(x) = +\infty$$

called "f is infinite at infinity". If X is a normed vector space of finite dimension and f goes to infinity at infinity, then the level sets of f are bounded, which implies that f is inf-compact.

In infinite dimension, inf-compactness property turn out to be easier to achieve using weak topologies. However, weak lower semicontinuity is a stronger property than strong lower semicontinuity. Although counter-intuitive, this is a straightforward consequence of the definition. Weak lower semicontinuity is enhanced by convexity.

Definition 1.59 *Let X be a vector space. A function $f : X \to \mathbb{R}\cup\{+\infty\}$ is said to be convex if*

$$f(\theta x + (1-\theta)y) \leq \theta f(x) + (1-\theta)f(y) \qquad \forall x, y \in X,\ \forall \theta \in]0,1[. \tag{1.18}$$

Corollary 1.60 *Let X be a separable reflexive Banach space and $f : X \to \mathbb{R}\cup\{+\infty\}$ be a proper, convex, lower semicontinuous function which is infinite at infinity. Then f admits a global minimizer.*

Proof. Let $\gamma \in \mathbb{R}$. The growth condition yields that $\operatorname{lev}_\gamma f$ is bounded, the lower semicontinuity yields that $\operatorname{lev}_\gamma f$ is closed, and the convexity yields that $\operatorname{lev}_\gamma f$ is convex. We recall (see [8]) that closed convex sets are weakly closed, that in a reflexive Banach space closed balls are weakly compact, and that under the additional separability assumption the weak topology is metrizable in closed balls. We infer that $\operatorname{lev}_\gamma f$ is weakly compact. Lastly, closedness/compactness is equivalent to sequential closedness/compactness in metric spaces. We have shown that f is inf-compact and lower-semi-continuous for the weak topology. The conclusion follows from Theorem 1.58. □

Remark 1.61 *The separability assumption can actually be dropped: we show that the level-sets are sequentially weakly compact through working in the closure of the vector space spanned by the sequence, which is always separable by construction (see e.g. [8] th. III.27).*

Remark 1.62 *When working with non-reflexive Banach spaces one has to distinguish between the weak and weak-$*$ topologies and develop carefully all the arguments in the specific cases.*

Let us now have a closer look to the decomposition $f = j + I_{\mathcal{U}}$.

Definition 1.63 *Let X be a vector space. A set $\mathcal{U} \subset X$ is said to be convex if*

$$\forall x, y \in \mathcal{U}, \forall \theta \in [0,1], \qquad \theta x + (1-\theta) y \in \mathcal{U}.$$

Proposition 1.64 *Let X be a normed vector space, $\mathcal{U} \subset X$, $j : X \to \mathbb{R} \cup \{+\infty\}$, $f = j + I_{\mathcal{U}}$.*

1. *If $\mathcal{U}$ is convex then $I_{\mathcal{U}}$ is convex.*
2. *If j is convex and $\mathcal{U}$ is convex then f is convex.*
3. *If $\mathcal{U}$ is closed then $I_{\mathcal{U}}$ is lower semicontinuous.*
4. *If $\mathcal{U}$ is closed and j is is lower semicontinuous then f is lower semicontinuous.*
5. *If $\mathcal{U}$ is bounded then f is infinite at infinity.*

Proof. 1. It stems from

$$I_{\mathcal{U}}(\theta x + (1-\theta) y) = +\infty \Rightarrow \theta x + (1-\theta) y \notin \mathcal{U} \Rightarrow x \notin \mathcal{U} \text{ or } y \notin \mathcal{U} \Rightarrow I_{\mathcal{U}}(x) = +\infty \text{ or } I_{\mathcal{U}}(y) = +\infty.$$

2. It is immediately seen that the sum of two convex functions is convex.
3. This is due to $\operatorname{lev}_\gamma I_{\mathcal{U}} = \mathcal{U}$ if $\gamma \geq 0$ and $\operatorname{lev}_\gamma I_{\mathcal{U}} = \emptyset$ if $\gamma < 0$.
4. It results from

$$\operatorname{lev}_\gamma f = \operatorname{lev}_\gamma j \cap \mathcal{U}.$$

5. This is simply because $f(x) = +\infty$ outside a ball. □

Finally, a very nice property of convex functions is the following.

Proposition 1.65 *Let X be a normed vector space and $f : X \to \mathbb{R} \cup \{+\infty\}$ be convex function. Every local minimizer of f is a global minimizer.*

Proof. Let x be a local minimizer of f and $y \in X$ be arbitrary. There exists $\theta \in]0,1[$ such that

$$f(x) \leq f((1-\theta)x + \theta y) \leq (1-\theta) f(x) + \theta f(y).$$

This yields $f(x) \leq f(y)$. □

1.4.4 Uniqueness

Definition 1.66 *Let X be a vector space. A proper function $f : X \to \mathbb{R} \cup \{+\infty\}$ is said to be strictly convex if the inequality* (1.18) *is strict whenever $x \neq y$ and $f(x), f(y) < +\infty$.*

Proposition 1.67 *If a proper and strictly convex function admits a global minimizer then it is unique.*

Proof. If $x \neq y$ are two minimizers then

$$f(\frac{x+y}{2}) < \frac{1}{2} f(x) + \frac{1}{2} f(y) = f(x),$$

which is in contradiction with x being a minimizer. □

Chapter 2

Examples of shape optimization problems

We refer to the general introduction for the distinction between parametric, geometry and topology optimization.

2.1 Examples of parametric optimization problems

2.1.1 Thickness optimization for a membrane model

Let Ω be a bounded open subset of $\mathbb{R}^2$. We assume that Ω is occupied by an elastic plate of unitary shear modulus, clamped on $\partial\Omega$, and submitted to a surface force $f \in L^2(\Omega)$. We suppose that shear strains (e_{xz} and e_{yz}) and stresses are dominant, which can be the case for a sufficiently thick plate. If $h(x)$ is the thickness of the plate at point x then the vertical displacement field u is solution of

$$\left\{ \begin{aligned} -\operatorname{div}(h\nabla u) &= f \quad \text{in } \Omega \\ u &= 0 \quad \text{on } \partial\Omega. \end{aligned} \right. \tag{2.1}$$

For h constant this also models an elastic membrane, this is why we call this the membrane model.

Actually, for the mathematical analysis, we are interested in weak solutions, i.e., in some $u \in H_0^1(\Omega)$ such that

$$\int_\Omega h\nabla u \cdot \nabla\varphi dx = \int_\Omega f\varphi dx \qquad \forall\varphi \in H_0^1(\Omega). \tag{2.2}$$

We know (by a slight modification of Proposition 1.37) that this problem has a unique solution provided that $h \in L^\infty(\Omega)$ satisfies $h(x) \geq h_{\min} > 0$ for a.e. $x \in \Omega$. We will search for thickness profiles within the set

$$\mathcal{U} = \left\{ h \in L^\infty(\Omega) : h_{\min} \leq h(x) \leq h_{\max} \text{ a.e. } x \in \Omega \right\}.$$

Note that (2.1) can be interpreted in other physical contexts. For instance it is the stationary heat equation, where u is the temperature and h is the heat conductivity. In this case it makes sense to also consider the 3D setting. We sometimes call (2.1) the conductivity problem.

A typical cost function is the compliance, i.e. the work done by the load,

$$J_{\text{comp}}(u) = \int_\Omega fudx.$$

Note that $J_{\text{comp}}(u) = -2\mathcal{E}(u)$, with the primal energy

$$\mathcal{E}(v) = \frac{1}{2}\int_\Omega h|\nabla v|^2 dx - \int_\Omega fvdx.$$

This energy, together with its dual counterpart, provide variational formulations for (2.2), see section 1.3.5. This makes the compliance rather convenient to deal with.

Another classical example is the least square cost

$$J_{\text{l.s.}}(u) = \int_\Omega (u - \bar{u})^2 dx,$$

where $\bar{u} \in L^2(\Omega)$ is a target displacement field.

2.1.2 Thickness optimization of a Kirchhoff plate

If now the behavior of the plate is dominated by internal bending moments, then it is classical to consider the Kirchhoff model (written here with clamped boundary conditions):

$$\left\{\begin{aligned} \operatorname{div}\operatorname{div}(\frac{h^3}{12}A\nabla\nabla u) = f \quad & \text{in } \Omega \\ u = \frac{\partial u}{\partial n} = 0 \quad & \text{on } \partial\Omega, \end{aligned}\right. \tag{2.3}$$

where A is Hooke's tensor in plane stress, see (1.13), (1.14). It is a fourth order boundary value problem with weak formulation: find $u \in H_0^2(\Omega)$ such that

$$\int_\Omega \frac{h^3}{12} A\nabla\nabla u : \nabla\nabla\varphi dx = \int_\Omega f\varphi dx \qquad \forall\varphi \in H_0^2(\Omega).$$

The existence and uniqueness of a solution follows from the Lax-Milgram theorem together with the Poincaré inequality

$$\|u\|_{H^2(\Omega)} \leq c\|\nabla\nabla u\|_{L^2(\Omega)} \qquad \forall u \in H_0^2(\Omega),$$

itself proven along the same lines as its counterpart in $H_0^1(\Omega)$. Here also typical cost functions are the compliance or a least square cost. Optimization of eigenfrequencies are also of interest.

2.1.3 Optimization of CAD parameters

In the two above situations the unknown h is sought within an infinite-dimensional vector space. In some industrial applications the "design space" is discrete. It is typically represented by CAD variables such as lengths, control points of splines, NURBS...

2.2 Examples of geometry optimization problems

2.2.1 Membrane / conductivity problems

We consider again (2.1), where now h is fixed (but not necessarily constant). The unknown is the open set Ω. It is classical to assume as constraint $\Omega \ni \omega$, where ω is a prescribed set containing the support of f. Of course, the load can also apply on a part of the boundary as a Neumann or a Dirichlet condition, and it is standard to assume that this part is fixed.

2.2.2 Linear elasticity problems

A very classical problem is to optimize $\Omega \subset \mathbb{R}^N$, $N = 2, 3$, in the linear elasticity problem (see section 1.3.4)

$$\left\{\begin{array}{ll} -\operatorname{div}\sigma(u) = f & \text{in } \Omega \\ u = 0 & \text{on } \Gamma_D \\ \sigma(u)n = g & \text{on } \Gamma_N. \end{array}\right. \tag{2.4}$$

As cost function we often consider the compliance

$$J_{\text{comp}}(u) = \int_\Omega f \cdot u dx + \int_{\Gamma_N} g \cdot u ds.$$

Other classical criteria involve local or averaged values of the stress $\sigma(u)$.

2.2.3 Some other problems involving PDEs

Geometry optimization is used in many other applicative contexts. Let us cite

- optimal design of structures with complex behaviors (nonlinear elasticity, plasticity...),
- flow optimization (pipes, profiles) based on fluid dynamics models,
- optimization of electromagnetic devices (antennas, motors, wave guides...).

2.2.4 Perimetric problems

In all the aforementioned problems it is standard to incorporate a volume constraint, for instance to account for weight. It is sometimes also useful to consider the perimeter for its regularizing properties, as we will see later. The perimeter is sometimes also involved for its intrinsic meaning. Let us mention the isoperimetric problem: what is the shape of maximal volume with prescribed perimeter? In the absence of constraint the answer is well known to be the ball. A less academic problem is to find minimal surfaces: given a closed curve Γ in $\mathbb{R}^3$, what is the surface of minimal area which admits Γ as boundary? This kind of problem is used to model surface tensions.

2.3 Examples of topology optimization problems

In the examples of section 2.2 it was implicitly assumed that the topology of Ω was prescribed as that of a reference shape Ω_0, typically the initial guess of an optimization procedure. Intuitively, we can say that two shapes have the same topology if there exists a continuous deformation from one to the other. This leaves unchanged the number of holes in dimension 2, and further properties in dimension 3. Precise mathematical definitions are rather complicated and there are several concepts which are not truly equivalent. One of them, that will be useful for later purposes, is to say that the two sets are homeomorphic. At the moment, the intuitive understanding is sufficient.

When the topology is unknown we speak of topology optimization. Topology optimization is particularly usefull in solid mechanics.

2.4 Outline

There is a large amount of notions related to the treatment of shape optimization problems. Some are more theoretical, like the existence and regularity of optimal shapes, others have more direct practical implications, like the various notions of derivatives used to build descent directions. In this course we will address a selection of concepts dedicated to both aspects. We will mainly focus on geometry and topology optimization problems, since parametric optimization is more closely related to standard nonlinear optimization.

Chapter 3

Existence and non-existence of optimal shapes

This chapter is dedicated to the difficult question of the existence of optimal shapes. We will restrict ourselves to the main ideas and results.

The main argument to prove existence is the direct method of the calculus of variations described in Theorem 1.58. The difficulty is to find a topology (or at least a notion of convergence) on the set of domains that guarrantees at the same time the lower semi-continuity and the inf-compactness of the cost function.

3.1 Examples

3.1.1 Example of existence

Let D be an open, bounded and connected subset of $\mathbb{R}^N$ with boundary $\partial D = \Gamma_D \cup \Gamma_N$, $\Gamma_D \cap \Gamma_N = \emptyset$, $|\Gamma_D| > 0$. For $\Omega \subset D$ we define the piecewice constant thermal conductivity

$$\sigma_\Omega = \chi_\Omega \alpha + (1 - \chi_\Omega)\beta, \qquad \alpha, \beta > 0.$$

Given $h \in H^{1/2}(\partial D)$ we consider the problem

$$\begin{cases} -\operatorname{div}(\sigma_\Omega \nabla u_\Omega) = 0 & \text{in } D \\ u_\Omega = h & \text{on } \Gamma_D \\ \sigma_\Omega \frac{\partial u_\Omega}{\partial n} = 0 & \text{on } \Gamma_N. \end{cases} \tag{3.1}$$

For a Dirichlet load, it is of interest to maximize the thermal compliance. Therefore we consider as cost function the half negative compliance

$$J(\Omega) = -\frac{1}{2}\int_D \sigma_\Omega |\nabla u_\Omega|^2.$$

Let $w \in H^1(\Omega)$ be such that $\gamma_0 w = h$ on Γ_D. The weak formulation reads

$$u_\Omega \in \{w\} + H, \qquad H := \left\{ v \in H^1(D) : \gamma_0 v = 0 \text{ on } \Gamma_D \right\},$$

$$\int_D \sigma_\Omega \nabla u_\Omega \cdot \nabla \varphi dx = 0 \qquad \forall \varphi \in H.$$

The primal and dual variational principles yield (see sections 1.3.5 and 2.1.1)

$$J(\Omega) = \max_{v \in \{w\}+H} -\frac{1}{2}\int_D \sigma_\Omega |\nabla v|^2 dx,$$

$$J(\Omega) = \min_{\tau \in \mathcal{T}} \frac{1}{2} \int_D \sigma_\Omega^{-1} |\tau|^2 dx - \int_D \tau \cdot \nabla w dx,$$

with

$$\begin{aligned} \mathcal{T} &= \left\{ \tau \in L^2(D)^N : \int_D \tau \cdot \nabla v dx = 0 \ \forall v \in H \right\} \\ &= \left\{ \tau \in L^2(D)^N : -\operatorname{div} \tau = 0 \text{ in } D, \ \tau \cdot n = 0 \text{ on } \Gamma_N \right\}. \end{aligned}$$

We address the problem

$$\operatorname*{minimize}_{\Omega \in \mathcal{U}} J(\Omega), \tag{3.2}$$

$$\text{with} \qquad \mathcal{U} = \{\Omega \subset D \text{ measurable} : |\Omega| = V\},$$

given a target volume $0 \leq V \leq |D|$.

Proposition 3.1 *Consider the dimension $N = 2$, the square $D =]-\frac{1}{2}, \frac{1}{2}[\times]-\frac{1}{2}, \frac{1}{2}[$, the left and right borders $\Gamma_D = \{-\frac{1}{2}\}\times]-\frac{1}{2}, \frac{1}{2}[\cup \{\frac{1}{2}\}\times]-\frac{1}{2}, \frac{1}{2}[$, the bottom and top borders $\Gamma_N =]-\frac{1}{2}, \frac{1}{2}[\times\{-\frac{1}{2}\} \cup]-\frac{1}{2}, \frac{1}{2}[\times\{\frac{1}{2}\}$, the volume constraint $V = \frac{1}{2}$, the Dirichlet data $h = -\frac{1}{2}$ on $\{-\frac{1}{2}\}\times]-\frac{1}{2}, \frac{1}{2}[$, $h = \frac{1}{2}$ on $\{\frac{1}{2}\}\times]-\frac{1}{2}, \frac{1}{2}[$ (see Fig. 3.1). Then problem (3.2) admits solutions.*

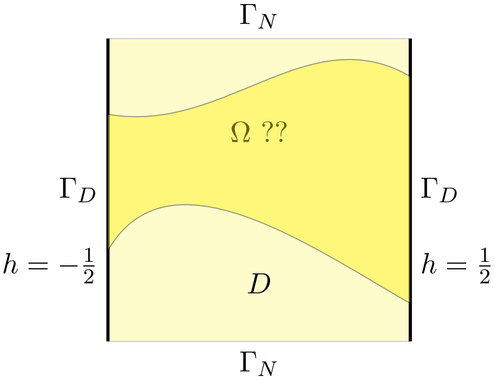


Figure 3.1: Domain D and boundary conditions.

Proof. Step 0: finding some w. We simply take

$$w(x_1, x_2) = x_1, \qquad \nabla w(x_1, x_2) = \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

Note that due to the existence of such a lifting, the boundary value problem is well-posed although D is not of class $\mathcal{C}^1$.

Step 1: lower bound. We use the primal variational principle. Let $v = w \in \{w\} + H$. We obtain

$$J(\Omega) \geq -\frac{1}{2} \int_D \sigma_\Omega |\nabla w|^2 dx = -\frac{1}{2} \int_D \sigma_\Omega = -\frac{1}{2} \frac{\alpha + \beta}{2} = -\frac{\alpha + \beta}{4}.$$

Step 2: upper bound. We use the dual variational principle. Let

$$\Omega_0 = \left]-\frac{1}{2}, \frac{1}{2}\right[\times \left]-\frac{1}{4}, \frac{1}{4}\right[, \qquad \tau_0 = \sigma_{\Omega_0} \begin{pmatrix} 1 \\ 0 \end{pmatrix}.$$

This Ω_0 satisfies the volume constraint. Moreover, although τ_0 is discontinuous, there is no jump of the normal component to the discontinuity lines. This results in $\operatorname{div} \tau_0 = 0$. Since clearly we have $\tau_0 \cdot n = 0$ on Γ_N, we infer that $\tau_0 \in \mathcal{T}$. We have

$$J(\Omega_0) \leq \frac{1}{2} \int_D \sigma_{\Omega_0}^{-1} |\tau_0|^2 dx - \int_D \tau_0 \cdot \nabla w dx = \frac{1}{2} \int_D \sigma_{\Omega_0} dx - \int_D \sigma_{\Omega_0} dx = -\frac{\alpha + \beta}{4}.$$

We conclude that

$$\inf_{\Omega\in\mathcal{U}} J(\Omega) = -\frac{\alpha+\beta}{4}.$$

This bound is attained by the set Ω_0. □

Note that the construction from the proof shows that there are infinitely many solutions.

3.1.2 Example of non-existence

We modify the previous example as follows. Given $g \in H^{-1/2}(\partial D)$ with $\int_{\partial D} g ds = 0$ we consider the Neumann problem

$$\begin{cases} -\operatorname{div}(\sigma_\Omega \nabla u_\Omega) = 0 & \text{in } D \\ \sigma_\Omega \dfrac{\partial u_\Omega}{\partial n} = g & \text{on } \partial D. \end{cases} \tag{3.3}$$

The weak formulation reads

$$u_\Omega \in H := \left\{ v \in H^1(D) : \int_D v dx = 0 \right\}.$$

$$\int_D \sigma_\Omega \nabla u_\Omega \cdot \nabla \varphi dx = \int_{\partial D} g \gamma_0 \varphi ds \qquad \forall \varphi \in H.$$

Here we minimize the thermal compliance. Therefore we consider as cost function the half compliance

$$J(\Omega) = \frac{1}{2}\int_D \sigma_\Omega |\nabla u_\Omega|^2 = \frac{1}{2}\int_{\partial D} g\gamma_0 u_\Omega ds.$$

The primal and dual variational principles yield

$$J(\Omega) = \max_{v\in H} -\frac{1}{2}\int_D \sigma_\Omega |\nabla v|^2 dx + \int_{\partial D} g\gamma_0 v ds,$$

$$J(\Omega) = \min_{\tau\in\mathcal{T}} \frac{1}{2}\int_D \sigma_\Omega^{-1}|\tau|^2 dx,$$

with

$$\mathcal{T} = \left\{ \tau \in L^2(D)^N : -\operatorname{div}\tau = 0 \text{ in } D \text{ and } \tau\cdot n = g \text{ on } \partial D \right\}.$$

We again address the problem

$$\underset{\Omega\in\mathcal{U}}{\text{minimize}}\ J(\Omega), \tag{3.4}$$

$$\text{with} \qquad \mathcal{U} = \{\Omega \subset D \text{ measurable} : |\Omega| = V\}.$$

We will use the following algebraic lemma, which in particular shows the joint convexity of the function $(\sigma,\tau) \mapsto \sigma^{-1}|\tau|^2$.

Lemma 3.2 *For all $\sigma, \sigma_0 > 0$, $\tau, \tau_0 \in \mathbb{R}^N$ we have the equality*

$$\sigma^{-1}|\tau|^2 - \sigma_0^{-1}|\tau_0|^2 = -\frac{\sigma-\sigma_0}{\sigma_0^2}|\tau_0|^2 + \frac{2}{\sigma_0}\tau_0\cdot(\tau-\tau_0) + \sigma^{-1}\left|\tau - \frac{\sigma}{\sigma_0}\tau_0\right|^2.$$

PROOF. Just expand and simplify the right hand side. □

Proposition 3.3 *Take $N = 2$, $D =]-\frac{1}{2},\frac{1}{2}[\times]-\frac{1}{2},\frac{1}{2}[$, $\Gamma_D = \{-\frac{1}{2}\}\times]-\frac{1}{2},\frac{1}{2}[\cup \{\frac{1}{2}\}\times]-\frac{1}{2},\frac{1}{2}[$, $\Gamma_N =]-\frac{1}{2},\frac{1}{2}[\times\{-\frac{1}{2}\} \cup]-\frac{1}{2},\frac{1}{2}[\times\{\frac{1}{2}\}$, $V = \frac{1}{2}$, $g = -\frac{1}{2}$ on $\{-\frac{1}{2}\}\times]-\frac{1}{2},\frac{1}{2}[$, $g = \frac{1}{2}$ on $\{\frac{1}{2}\}\times]-\frac{1}{2},\frac{1}{2}[$, $g = 0$ elsewhere (see Fig. 3.2). Then problem* (3.4) *admits no solution.*

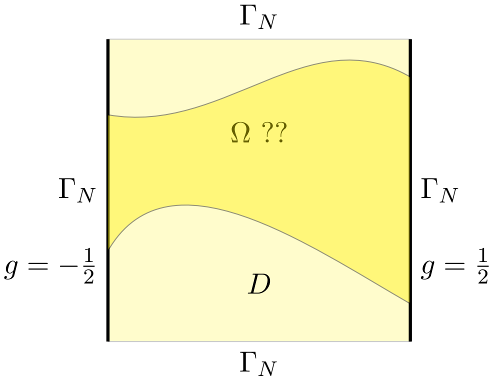


Figure 3.2: Domain D with Neumann boundary conditions.

Proof. Step 1: lower bound. We use the primal variational principle. Let

$$v(x_1, x_2) = kx_1, \qquad \nabla v(x_1, x_2) = \begin{pmatrix} k \\ 0 \end{pmatrix},$$

for some constant k to be fixed. We obtain

$$J(\Omega) \geq -\frac{1}{2}\int_D \sigma_\Omega |\nabla v|^2 dx + \int_{\Gamma_N} g\gamma_0 v ds = -\frac{k^2}{2}\frac{\alpha+\beta}{2} + \left(-\frac{1}{2}\right)\left(-\frac{k}{2}\right) + \frac{1}{2}\frac{k}{2} = -k^2\frac{\alpha+\beta}{4} + \frac{k}{2}.$$

This quantity is maximized for $k = 1/(\alpha+\beta)$, which results in

$$J(\Omega) \geq \frac{1}{4(\alpha+\beta)}.$$

Step 2: upper bound. We use the dual variational principle. Let $n \in \mathbb{N}^*$ and

$$\Omega_n = \bigcup_{0\leq i<n} \left]-\frac{1}{2}, \frac{1}{2}\right[\times \left]-\frac{1}{2} + \frac{i}{n}, -\frac{1}{2} + \frac{2i+1}{2n}\right[.$$

We also define

$$D_n = \left]-\frac{1}{2} + \frac{1}{n}, \frac{1}{2} - \frac{1}{n}\right[\times \left]-\frac{1}{2}, \frac{1}{2}\right[.$$

We set

$$\tau_n = \frac{\sigma_{\Omega_n}}{\alpha+\beta}\begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad \text{in } D_n.$$

It is important to notice that

$$\int_{\Delta_n^1} \tau_n \cdot e_1 ds = \frac{1}{2} = -\int_{\Gamma_N^1} g ds$$

with

$$\Gamma_N^1 = \left\{-\frac{1}{2}\right\} \times \left]-\frac{1}{2}, \frac{1}{2}\right[, \qquad \Delta_n^1 = \left\{-\frac{1}{2} + \frac{1}{n}\right\} \times \left]-\frac{1}{2}, \frac{1}{2}\right[.$$

Of course the same holds on the other side. Hence we can extend τ_n over D in order to fulfill $\tau \in \mathcal{T}$. In addition, this extension can be constructed by gluing elementary solutions on squares of form

$$Q_{i,n}^1 = \left]-\frac{1}{2}, -\frac{1}{2} + \frac{1}{n}\right[\times \left]-\frac{1}{2} + \frac{i}{n}, -\frac{1}{2} + \frac{i+1}{n}\right[.$$

Such solutions can be first found on the unit square $Q =]0,1[\times]0,1[$ then transported by an affine change of variable. This construction ensures that

$$\int_{D\setminus D_n} \sigma_{\Omega_n}^{-1}|\tau_n|^2 = O(2n \times \frac{1}{n^2}) = O(\frac{1}{n}).$$

We arrive at

$$J(\Omega_n) \leq \frac{1}{2}\int_{D_n} \sigma_{\Omega_n}^{-1}|\tau_n|^2 + O(\frac{1}{n}) = \frac{1}{2}\frac{1}{(\alpha+\beta)^2}\int_{D_n} \sigma_{\Omega_n} + O(\frac{1}{n}) = \frac{1}{4(\alpha+\beta)} + O(\frac{1}{n}).$$

Together with step 1 we conclude that

$$\inf_{\Omega\in\mathcal{U}} J(\Omega) = \frac{1}{4(\alpha+\beta)}.$$

Step 3: non-existence. We use Lemma 3.2 with $\sigma = \sigma_\Omega$, $\Omega \in \mathcal{U}$, $\sigma_0 = \frac{\alpha+\beta}{2}$, $\tau_0 = \frac{1}{2}e_1$, and $\tau \in \mathcal{T}$. Observing that

$$\int_D \tau\cdot\tau_0 dx = \frac{1}{2}\int_D \tau\cdot\nabla(x\cdot e_1)dx = \frac{1}{2}\int_{\partial D} \tau\cdot n(x\cdot e_1)ds = \frac{1}{4} = \int_D |\tau_0|^2 dx,$$

we obtain that

$$\frac{1}{2}\int_D \sigma_\Omega^{-1}|\tau|^2 dx - \frac{1}{4(\alpha+\beta)} = \frac{1}{2}\int_D \sigma_\Omega^{-1}|\tau - \frac{\sigma_\Omega}{\sigma_0}\tau_0|^2 dx.$$

Suppose that Ω is optimal. Then there exists $\tau \in \mathcal{T}$ such that the left hand side vanishes. It follows that

$$\tau = \frac{\sigma_\Omega}{\sigma_0}\tau_0 = \frac{\sigma_\Omega}{\alpha+\beta}e_1.$$

This τ does not satisfy the boundary condition from $\mathcal{T}$. □

Remark 3.4 *An alternative proof for step 3 is to work with the primal energy: if Ω is optimal then*

$$J(\Omega) = \max_{v\in H} -\frac{1}{2}\int_D \sigma_\Omega|\nabla v|^2 dx + \int_{\partial D} g\gamma_0 v ds = \frac{1}{4(\alpha+\beta)} = -\frac{1}{2}\int_D \sigma_\Omega|\nabla v_k|^2 dx + \int_{\partial D} g\gamma_0 v_k ds,$$

where v_k is the function constructed in step 1. It follows by uniqueness that $v_k = u_\Omega$, but it does not satisfy the boundary condition $\sigma_\Omega\frac{\partial v_k}{\partial n} = g$.

Remark 3.5 *We have considered a two-phase shape optimization problem. When one of the conductivities, say α, tends to 0, the solution of (3.1) converges in some sense to the solution of the one-phase problem*

$$\begin{cases} -\alpha\Delta u_\Omega = f & \textit{in } \Omega \\ u_\Omega = 0 & \textit{on } \partial\Omega\cap\Gamma_D \\ \alpha\frac{\partial u_\Omega}{\partial n} = g & \textit{on } \partial\Omega\cap\Gamma_N \\ \frac{\partial u_\Omega}{\partial n} = 0 & \textit{on } \partial\Omega\setminus(\Gamma_N\cup\Gamma_D). \end{cases} \tag{3.5}$$

A counter-example to the existence of optimal shape for the one-phase problem can be found in [16].

The example described above illustrates well a typical phenomenon that can be visualized through numerical simulations: minimizing sequences tend to mix the phases at a smaller and smaller scale, without ever converging. We say that **homogenization** occurs. A way to "enforce" the existence of optimal configurations is to include homogenized phases in the admissible set. This procedure is called relaxation. Details can be found in [1, 2]. Another way is to prevent homogenization, as we will see.

3.2 Notions of convergence for sequences of domains

3.2.1 Convergence of characteristic functions

Let D be an open subset of $\mathbb{R}^N$. We recall the characteristic function of a set $\Omega \subset D$:

$$\chi_\Omega : x \in D \mapsto \begin{cases} 1 & \text{if } x \in \Omega \\ 0 & \text{otherwise.} \end{cases}$$

Definition 3.6 *Let Ω_n, Ω be measurable subsets of D. We say that the sequence (Ω_n) converges to Ω in the sense of characteristic functions if $\chi_{\Omega_n} \to \chi_\Omega$ a.e. in D.*

If $|D| < +\infty$, this convergence implies $\|\chi_{\Omega_n} - \chi_\Omega\|_{L^1(D)} \to 0$, itself equivalent to $|\Omega \backslash \Omega_n| + |\Omega_n \backslash \Omega| \to 0$.

The convergence in the sense of characteristic functions has nice lower semicontinuity properties. An elementary example is shown below.

Proposition 3.7 *Let Ω_n, Ω be measurable subsets of D such that Ω_n converges to Ω in the sense of characteristic functions. Then $|\Omega| \leq \liminf_{n\to+\infty} |\Omega_n|$. If $|D| < +\infty$ then $|\Omega| = \lim_{n\to+\infty} |\Omega_n|$.*

Proof. We have by Fatou's lemma

$$|\Omega| = \int_D \chi_\Omega dx = \int_D \lim_{n\to+\infty} \chi_{\Omega_n} dx \leq \liminf_{n\to+\infty} \int_D \chi_{\Omega_n} dx.$$

If $|D| < +\infty$ then we apply Lebesgue's dominated convergence theorem. □

Unfortunately, it is well-known that a major drawback of strong topologies is that they hardly yield compactness properties. For extracting converging subsequences one usually prefers weak topologies.

Theorem 3.8 *Let $(\Omega_n)_{n\in\mathbb{N}}$ be a sequence of measurable subsets of D. There exists $w \in L^\infty(D, [0,1])$ such that, up to a subsequence,*

$$\lim_{n\to+\infty} \int_D (\chi_{\Omega_n} - w)\varphi dx = 0 \qquad \forall \varphi \in L^1(D). \tag{3.6}$$

Proof. We simply observe that each $u_n := 2\chi_{\Omega_n} - 1$ belongs to the unit ball of $L^\infty(D)$. Since $L^\infty(D)$ identifies with the continuous dual of $L^1(D)$, its unit ball is weakly-* compact. In addition, $L^1(D)$ is separable, which yields that the unit ball of $L^\infty(D)$ is metrisable for the weak-* topology. Therefore compactness is equivalent to sequential compactness. Consequently, there exists $u \in L^\infty(D, [-1,1])$ such that $u_n \rightharpoonup u$ weakly-* in $L^\infty(D)$. It follows that $\chi_{\Omega_n} \rightharpoonup w := (u+1)/2$ weakly-* in $L^\infty(D)$. □

The main drawback of theorem 3.8 is that it does not guarrantee that the limit χ is a characteristic function. In fact, if it is the case, then the convergence becomes strong.

Proposition 3.9 *Under the assumptions of Theorem 3.8, if $w = \chi_\Omega$ for some measurable set Ω and $|D| < +\infty$, then χ_{Ω_n} converges to w in $L^1(D)$. Therefore, up to a possible further subsequence, Ω_n converges to Ω in the sense of characteristic functions.*

Proof. Since $w, \chi_{\Omega_n} \in L^\infty(D, \{0,1\})$ we have

$$\|\chi_{\Omega_n} - w\|_{L^1(D)} = \int_{\{w=0\}} \chi_{\Omega_n} dx + \int_{\{w=1\}} (1 - \chi_{\Omega_n}) dx.$$

Choosing $\varphi = \chi_{\{w=0\}}$ in (3.6) yields that the first integral goes to 0. Choosing now $\varphi = \chi_{\{w=1\}}$ yields that the second integral goes to 0. □

3.2.2 Hausdorff distances

We denote by $|\cdot|$ the Euclidean norm on $\mathbb{R}^N$.

Definition 3.10 *Let K_1, K_2 be two nonempty compact subsets of $\mathbb{R}^N$. We define the distance between x and K_1 by*

$$\forall x \in \mathbb{R}^N, \qquad d_{K_1}(x) = d(x, K_1) = \min_{y \in K_1} |x - y|,$$

and the Hausdorff distance between K_1 and K_2 by

$$h(K_1, K_2) = \max\{\max_{x \in K_1} d(x, K_2), \max_{x \in K_2} d(x, K_1)\}.$$

Let Ω_1, Ω_2 be two (relatively) open strict (i.e. $\Omega_1 \neq C$ and $\Omega_2 \neq C$) subsets of a compact set C of $\mathbb{R}^N$. We define the complementary Hausdorff distance between Ω_1 and Ω_2 relatively to C by

$$h_C(\Omega_1, \Omega_2) = h(C \setminus \Omega_1, C \setminus \Omega_2).$$

It is immediate to show that the distance function d_{K_i} is 1-Lipschitz continuous. It can be shown that the complementary Hausdorff distance actually does not depend on the compact C when $\Omega_1, \Omega_2 \subset\subset C$, see [16].

Proposition 3.11 *Let K_1, K_2 be two nonempty compact subsets of $\mathbb{R}^N$ and $\|\cdot\|_\infty$ be the norm of uniform convergence over a set $C \supset K_1 \cup K_2$. We have*

$$h(K_1, K_2) = \|d_{K_1} - d_{K_2}\|_\infty.$$

Proof. Denote $\sigma(K_1, K_2) = \|d_{K_1} - d_{K_2}\|_\infty$. If $x \in K_1$ then

$$d(x, K_2) = |d(x, K_1) - d(x, K_2)| \leq \sigma(K_1, K_2).$$

Of course a similar inequality holds for $d(x, K_1)$ when $x \in K_2$, hence

$$h(K_1, K_2) \leq \sigma(K_1, K_2).$$

For the converse inequality we proceed as follows. Let $x \in C$. By continuity of the distance and compactness of K_1, K_2, there exists $y_1 \in K_1$, $y_2 \in K_2$ such that $d(x, K_1) = |x - y_1|$ and $d(x, K_2) = |x - y_2|$. We have by the triangle inequality

$$d(x, K_1) \leq |x - y_2| + d(y_2, K_1) = d(x, K_2) + d(y_2, K_1),$$

whereby

$$d(x, K_1) - d(x, K_2) \leq d(y_2, K_1) \leq h(K_1, K_2).$$

The claim follows by exchanging the roles of K_1 and K_2. □

One of the consequences of Proposition 3.11 is that it immediately entails the triangle inequality of the Hausdorff distance. We arrive at the following.

Corollary 3.12 *The set of the nonempty compact subsets of $\mathbb{R}^N$ endowed with the Hausdorff distance is a metric space.*

The set of the open strict subsets of a fixed compact C of $\mathbb{R}^N$, endowed with the complementary Hausdorff distance is a metric space.

Each of the above metrics leads to its own notion of convergence, neither weaker nor stronger than the convergence in the sense of characteristic functions.

The Hausdorff distances turn out to have very good compactness properties.

Theorem 3.13 *Let (K_n) be a sequence of nonempty compact subsets of $\mathbb{R}^N$ contained in a fixed compact C. There exists a nonempty compact $K \subset C$ such that K_n converges to K for the Hausdorff distance, up to a subsequence.*

Proof. Consider the sequence (d_{K_n}) of $\mathcal{C}(C)$ endowed with the norm $\|\cdot\|_\infty$. It is bounded, since C is bounded. Moreover, since d_{K_n} is 1-Lipschitz continuous, the sequence (d_{K_n}) is uniformly equicontinuous. By Ascoli's theorem, (d_{K_n}) is relatively compact: it admits a converging subsequence. Up to relabeling, suppose that $d_{K_n} \to f \in \mathcal{C}(C)$. Let

$$K = \{x \in C : f(x) = 0\}.$$

It is a closed and bounded subset of $\mathbb{R}^N$, thus a compact. We will show that $f = d_K$. Then by Proposition 3.11 we will infer that $K_n \to K$ for the Hausdorff distance.

We use again the Lipschitz property

$$|d_{K_n}(x) - d_{K_n}(y)| \le |x-y| \qquad \forall x, y \in C,$$

which yields at the limit $|f(x) - f(y)| \le |x-y|$. Choosing any $y \in K$ provides $f(x) \le |x-y|$, thus $f(x) \le d_K(x)$. Let now $x \in C$ and $x_n \in K_n$ such that $d(x, K_n) = |x - x_n|$. Since C is compact, up to a subsequence, $x_n \to y \in C$. We obtain at the limit $f(x) = |x-y|$, in particular $f(y) = 0$. Therefore $y \in K$. It follows that $f(x) \ge d(x, K)$ and the proof is complete. □

Corollary 3.14 *Let (Ω_n) be a sequence of open strict subsets of a fixed compact set C of $\mathbb{R}^N$. There exists an open strict subset Ω of C such that Ω_n converges to Ω for the complementary Hausdorff distance relatively to C, up to a subsequence.*

Unfortunately, the most standard cost functions fail to be lower semicontinuous for these distances. The reason is that the Hausdorff distance essentially controls only "half" of the L^1 distance of characteristic functions.

Proposition 3.15 *Let Ω_n, Ω be open strict subsets of a fixed compact set C. If Ω_n converges to Ω for the complementary Hausdorff distance relatively to C then*

(i) $|\Omega \setminus \Omega_n| \to 0$;

(ii) $\chi_\Omega \le \liminf_{n\to+\infty} \chi_{\Omega_n}$;

(iii) $|\Omega| \le \liminf_{n\to+\infty} |\Omega_n|$.

Proof. Let $A_n = C \setminus \Omega_n$, $A = C \setminus \Omega$. We have by definition $h(A_n, A) \to 0$. In particular $d_n := \max_{x\in A_n} d(x, A) \to 0$. Set $\delta_n = \max_{k\ge n} d_k$. Then also $\delta_n \to 0$, (δ_n) is nonincreasing, and obviously $d_n \le \delta_n$. Define

$$B_n = \{x \in C : 0 < d(x, A) \le \delta_n\}.$$

The sequence (χ_{B_n}) is nonincreasing and it converges pointwise to 0. Now we have

$$x \in A_n \setminus A \Rightarrow 0 < d(x, A) \le d_n \Rightarrow x \in B_n.$$

Hence

$$\chi_{\Omega\setminus\Omega_n} = \chi_{A_n \setminus A} \le \chi_{B_n}.$$

We infer by monotone convergence that

$$|\Omega \setminus \Omega_n| = \int_C \chi_{\Omega\setminus\Omega_n} dx \to 0.$$

For the second assertion we write

$$\chi_\Omega = \chi_{\Omega\setminus\Omega_n} + \chi_{\Omega\cap\Omega_n} \le \chi_{\Omega\setminus\Omega_n} + \chi_{\Omega_n}$$

and use that $\chi_{\Omega\setminus\Omega_n} \to 0$. The third assertion results from Fatou's lemma, or directly from $|\Omega| \le |\Omega \setminus \Omega_n| + |\Omega_n|$. □

The volume is lower semicontinuous for the complementary Hausdorff distance, but it can happen that the volume of the limit is strictly less than the limit of the volumes. This is achieved by considering oscillating boundaries. It can even happen that the limit of a sequence of sets of same volumes is empty: think of an homogenization phenomenon as in section 3.1.2. This prevents functionals that decrease with the volume, like the compliance, from being lower semicontinuous. The perimeter is not lower semicontinuous either for the complementary Hausdorff distance (see [16]). So, let us go back to the convergence in the sense of characteristic functions and search for compactness.

3.3 Existence under perimeter control

Let D be an open subset of $\mathbb{R}^N$.

3.3.1 Total variation and generalized perimeter

For any $\varphi \in \mathcal{C}(D)^N$ we set $\|\varphi\|_\infty = \sup_{x\in D} |\varphi(x)|_2$.

Definition 3.16 *Let $u \in L^1_{\text{loc}}(D)$. The total variation of u relatively to D is*

$$TV_D(u) = \sup\left\{\int_D u \operatorname{div}\varphi dx, \varphi \in \mathcal{C}^1_c(D)^N, \|\varphi\|_\infty \le 1\right\} \in [0, +\infty].$$

The space of functions with bounded variation in D is

$$BV(D) = \left\{u \in L^1(D) : TV_D(u) < +\infty\right\}.$$

Definition 3.17 *Let Ω be an arbitrary measurable subset of D. The perimeter of Ω relatively to D is $P_D(\Omega) := TV_D(\chi_\Omega)$.*

In the sequel we will use the classical notation $\omega \subset\subset \Omega$ to say that ω is open, $\bar\omega$ is compact and $\bar\omega \subset \Omega$.

Proposition 3.18 *If Ω is a bounded, open subset of D of class $\mathcal{C}^1$ then*

$$P_D(\Omega) = \int_{\partial\Omega\cap D} ds.$$

Proof. Let $\varphi \in \mathcal{C}^1_c(D)^N$ with $\|\varphi\|_\infty \le 1$. We have by the divergence formula

$$\int_D \chi_\Omega \operatorname{div}\varphi dx = \int_\Omega \operatorname{div}\varphi dx = \int_{\partial\Omega} \varphi\cdot n ds.$$

Since φ is compactly supported in D we have

$$\int_D \chi_\Omega \operatorname{div}\varphi dx \le \int_{\partial\Omega\cap D} \varphi\cdot n ds \le \int_{\partial\Omega\cap D} ds.$$

This gives a first inequality. For the converse one we need to construct an appropriate φ. As Ω is of class $\mathcal{C}^1$, its normal can be extended (by projection) over a tubular neighborhood $\mathcal{N}$ of $\partial\Omega$ into a continuous function $\tilde n$. Let $D_\varepsilon \subset\subset D$ and η_ε be a smooth function with values in $[0,1]$, equal to 1 on $\partial\Omega\cap D_\varepsilon$ and compactly supported in $\mathcal{N}\cap D$. Then $\eta_\varepsilon\tilde n \in \mathcal{C}_c(D)^N$ and $\eta_\varepsilon\tilde n = n$ on $\partial\Omega\cap D_\varepsilon$. By density, we construct $\varphi_k \in \mathcal{C}^1_c(D)^N$ converging uniformly to $\eta_\varepsilon\tilde n$. We have

$$P_D(\Omega) \ge \int_{\partial\Omega} \frac{\varphi_k}{\|\varphi_k\|_\infty}\cdot n ds \to \int_{\partial\Omega} \eta_\varepsilon\tilde n\cdot n ds = \int_{\partial\Omega\cap D} \eta_\varepsilon ds \ge \int_{\partial\Omega\cap D_\varepsilon} ds.$$

We can now vary D_ε :

$$P_D(\Omega) \ge \sup_{D_\varepsilon\subset\subset D} \int_{\partial\Omega\cap D_\varepsilon} ds = \int_{\partial\Omega\cap D} ds.$$

□

3.3.2 A compactness result

Theorem 3.19 *We assume that $|D| < +\infty$. Let (Ω_n) be a sequence of measurable subsets of D such that*

$$P_D(\Omega_n) \le C \qquad \forall n.$$

There exists a measurable subset Ω of D such that Ω_n converges to Ω in the sense of characteristic functions (see Definition 3.6), up to a subsequence.

For the proof we will use a classical compactness criterion in Lebesgue spaces, see e.g. [8].

Theorem 3.20 *Let Ω, ω be open subsets of $\mathbb{R}^N$ with $\bar{\omega} \subset\subset \Omega$. Let $\mathcal{F}$ be a bounded subset of $L^p(\Omega)$, $1 \le p < +\infty$. We assume that*

$$\forall \varepsilon > 0\ \exists \delta > 0, \delta < d(\omega, \Omega^c) \ \text{ s.t. } \int_\omega |f(x+h) - f(x)|^p dx < \varepsilon\ \forall h \in \mathbb{R}^N, |h| < \delta,\ \forall f \in \mathcal{F}.$$

Then $\mathcal{F}_{|\omega}$ is relatively compact in $L^p(\omega)$.

Lemma 3.21 *Let $u \in BV(D)$, $\delta > 0$ and ω be an open subset of D such that $\omega \subset\subset D_\delta = \{x \in D : d(x, \partial D) > \delta)\}$. We have*

$$\int_\omega |u(x+h) - u(x)| dx \le TV_D(u)|h| \qquad \forall h \in \mathbb{R}^N, |h| < \delta.$$

Proof. Let $\varphi \in \mathcal{C}^1_c(\omega)$ with $\|\varphi\|_\infty \le 1$. We obtain by change of variable and after extending φ by 0

$$\int_\omega (u(x+h) - u(x))\varphi(x) dx = \int_D u(x)(\varphi(x-h) - \varphi(x)) dx.$$

Then we write that

$$\varphi(x-h) - \varphi(x) = -\int_0^1 \nabla\varphi(x - th) \cdot h dt = |h| \int_0^1 \operatorname{div} \psi_t(x) dt$$

with $\psi_t(x) = -\varphi(x - th)\frac{h}{|h|}$. We arrive at

$$\int_\omega (u(x+h) - u(x))\varphi(x) dx = |h| \int_0^1 \left(\int_D u(x) \operatorname{div} \psi_t(x) dx \right) dt.$$

Since $\|\psi_t\|_\infty \le 1$ we infer that

$$\int_\omega u(x) \operatorname{div} \psi_t(x) dx \le TV_D(u).$$

This leads to

$$\int_\omega (u(x+h) - u(x))\varphi(x) dx \le |h| TV_D(u).$$

This extends by linearity to

$$\int_\omega (u(x+h) - u(x))\varphi(x) dx \le |h| TV_D(u) \|\varphi\|_\infty \qquad \forall \varphi \in \mathcal{C}^1_c(\omega).$$

By density this even holds true for any $\varphi \in \mathcal{C}_c(\omega)$. We derive the claim using a sequence $\varphi_n \in \mathcal{C}_c(\omega)$ such that $\varphi_n \to \varphi := \operatorname{sign}(u(x+h) - u(x))$ in $L^1(\omega)$, and setting $\bar{\varphi}_n(x) = \max(-1, \min(1, \varphi_n(x)))$. We have $\bar{\varphi}_n \in \mathcal{C}_c(\omega)$, $\|\bar{\varphi}_n\|_\infty \le 1$, and $\bar{\varphi}_n \to \varphi$ in $L^1(\omega)$ as the projection onto $[-1, 1]$ is 1-Lipschitz. □

Proof of Theorem 3.19.
By Theorem 3.8 there exists $w \in L^\infty(D, [0,1])$ such that $\chi_{\Omega_{\iota(n)}} \rightharpoonup w$ weakly-$*$ in $L^\infty(D)$, for a subsequence of indices $\iota(n)$. In particular, $\chi_{\Omega_{\iota(n)}} \to w$ in the sense of distributions. Let

$$\mathcal{F} = \left\{ u \in L^1(D) : TV_D(u) \le C \right\}$$

and fix some $\omega \subset\subset D$. By Theorem 3.20 and Lemma 3.21, $\mathcal{F}_{|\omega}$ is relatively compact in $L^1(\omega)$. Hence there exists $v_\omega \in L^1(\omega)$ and a further subsequence such that $\chi_{\Omega_{\iota\circ\lambda(n)}} \to v_\omega$ in $L^1(\omega)$. This implies that $\chi_{\Omega_{\iota(n)}} \to v_\omega$ in the sense of distributions in ω. It follows that $v_\omega = w$ a.e. in ω. Therefore, by uniqueness of the accumulation point in a compact space, $\chi_{\Omega_{\iota(n)}} \to w$ in $L^1(\omega)$. This yields that $w(x) \in \{0,1\}$ for a.e. $x \in \omega$, but as ω is arbitrary, $w(x) \in \{0,1\}$ for a.e. $x \in D$. We infer from Proposition 3.9 that $\chi_{\Omega_{\iota(n)}} \to w$ in $L^1(D)$, and a.e. in D up to a further subsequence. □

3.3.3 Lower semicontinuity

Theorem 3.22 *Let $u_n, u \in L^1_{\text{loc}}(D)$ such that*

$$\lim_{n\to+\infty} \|u_n - u\|_{L^1(K)} = 0 \qquad \forall K \, compact \subset D.$$

Then $TV_D(u) \leq \liminf_{n\to+\infty} TV_D(u_n)$.

PROOF. Let $\varphi \in \mathcal{C}^1_c(D)^N$ with $\|\varphi\|_\infty \leq 1$. We have

$$\int_D u \operatorname{div} \varphi dx = \lim_{n\to+\infty} \int_D u_n \operatorname{div} \varphi dx \leq \liminf_{n\to+\infty} TV_D(u_n).$$

Taking the supremum over φ yields the claim. □

When applied to characteristic functions, we obtain:

Corollary 3.23 *Let Ω_n, Ω be measurable subsets of D such that Ω_n converges to Ω in the sense of characteristic functions. Then $P_D(\Omega) \leq \liminf_{n\to+\infty} P_D(\Omega_n)$.*

3.3.4 Application

We revisit the two-phase conductivity problem of section 3.1.

Let D be an open, bounded and connected subset of $\mathbb{R}^N$ with boundary $\partial D = \Gamma_D \cup \Gamma_N$, $\Gamma_D \cap \Gamma_N = \emptyset$, $|\Gamma_D| > 0$. Let $\alpha > \beta > 0$ and recall that, for $\Omega \subset D$,

$$\sigma_\Omega = \chi_\Omega \alpha + (1 - \chi_\Omega)\beta.$$

Given $f \in L^2(D)$, $g \in H^{-1/2}(\partial D)$, we consider the problem

$$\begin{cases} -\operatorname{div}(\sigma_\Omega \nabla u_\Omega) = f & \text{in } D \\ u_\Omega = 0 & \text{on } \Gamma_D \\ \sigma_\Omega \frac{\partial u_\Omega}{\partial n} = g & \text{on } \Gamma_N. \end{cases} \tag{3.7}$$

Recall the work space

$$H := \left\{ v \in H^1(D) : \gamma_0 v = 0 \text{ on } \Gamma_D \right\}.$$

Theorem 3.24 *Let $\sigma_n, \sigma \in L^\infty(D, [\beta, \alpha])$. If $\sigma_n \to \sigma$ a.e. in D then $u_n \to u$ in $H^1(\Omega)$, where u_n, resp. u, are the solutions of* (3.7) *with conductivities σ_n, resp. σ.*

PROOF. Step 1. We first note that, due to the uniform coercivity and continuity constants, the sequence (u_n) is bounded in the Hilbert space H. Therefore, up to a subsequence, u_n weakly converges in H to some $u \in H$, in particular ∇u_n weakly converges to ∇u in $L^2(D)^N$. Moreover, the sequence $(\tau_n := \sigma_n \nabla u_n)$ is bounded in $L^2(D)$. Hence, for a possible further subsequence, τ_n weakly converges in $L^2(D)^N$ to some $\tau \in L^2(D)^N$. We have for all $\varphi \in \mathcal{C}^\infty_c(D)^N$

$$\int_D (\tau_n - \sigma \nabla u) \cdot \varphi dx = \int_D \sigma_n \nabla u_n \cdot \varphi dx - \int_D \sigma \nabla u \cdot \varphi dx = \int_D (\sigma_n - \sigma) \nabla u_n \cdot \varphi dx + \int_D \sigma (\nabla u_n - \nabla u) \cdot \varphi dx.$$

Therefore,

$$\left|\int_D (\tau_n - \sigma\nabla u)\cdot\varphi dx\right| \leq \|\sigma_n - \sigma\|_{L^2(D)}\|\nabla u_n\|_{L^2(D)}\|\varphi\|_{L^\infty(D)} + \int_D (\nabla u_n - \nabla u)\cdot(\sigma\varphi)dx.$$

We have $\|\sigma_n - \sigma\|_{L^2(D)} \to 0$ by dominated convergence, and $\|\nabla u_n\|_{L^2(D)}$ is uniformly bounded. Thus the first term of the right hand side goes to 0. The second term goes to 0 by weak convergence. We arrive at

$$\int_D (\tau - \sigma\nabla u)\cdot\varphi dx = 0 \ \forall\varphi \in \mathcal{C}_c^\infty(D)^N,$$

hence $\tau = \sigma\nabla u$. Consider now $\varphi \in H$. We have

$$\int_D f\varphi dx + \int_{\partial D} g\gamma_0\varphi ds = \int_D \sigma_n\nabla u_n\cdot\nabla\varphi dx = \int_D \tau_n\cdot\nabla\varphi dx \to \int_D \sigma\nabla u\cdot\nabla\varphi dx,$$

meaning that u is indeed solution of the boundary value problem with conductivity σ. By uniqueness of this solution, we infer that the full sequence (u_n) weakly converges to u in H, as well as the full sequence (τ_n) weakly converges to τ in $L^2(D)^N$.

Step 2. We now show that the convergence is strong. Passing to the limit in

$$\int_D \sigma_n\nabla u_n\cdot\nabla u_n dx = \int_D f u_n dx + \int_{\partial D} g\gamma_0 u_n ds$$

we obtain

$$\int_D \sigma_n\nabla u_n\cdot\nabla u_n dx \to \int_D f u dx + \int_{\partial D} g\gamma_0 u ds = \int_D \sigma\nabla u\cdot\nabla u dx.$$

We denote $\xi_n = \sigma_n^{1/2}\nabla u_n$ and $\xi = \sigma^{1/2}\nabla u$. We have just shown that $\|\xi_n\|_{L^2(D)} \to \|\xi\|_{L^2(D)}$. Yet we have for all $\varphi \in \mathcal{C}_c^\infty(D)^N$

$$\int_D \xi_n\cdot\varphi dx = \int_D \sigma_n^{-1/2}\tau_n\cdot\varphi dx = \int_D \tau_n\cdot\sigma^{-1/2}\varphi dx + \int_D (\sigma_n^{-1/2} - \sigma^{-1/2})\tau_n\cdot\varphi dx.$$

Passing to the limit using that $\|\sigma_n^{-1/2} - \sigma^{-1/2}\|_{L^2(D)} \to 0$ and $\|\tau_n\|_{L^2(D)}$ is bounded yields

$$\int_D \xi_n\cdot\varphi dx \to \int_D \tau\cdot\sigma^{-1/2}\varphi dx = \int_D \xi\cdot\varphi dx.$$

As the sequence (ξ_n) is weakly relatively compact in $L^2(D)^N$, as it is bounded, we infer that $\xi_n \to \xi$ weakly in $L^2(D)^N$. We now use a classical argument:

$$\|\xi_n - \xi\|_{L^2(D)}^2 = \|\xi_n\|_{L^2(D)}^2 + \|\xi\|_{L^2(D)}^2 - 2\int_D \xi_n\cdot\xi dx \to \|\xi\|_{L^2(D)}^2 + \|\xi\|_{L^2(D)}^2 - 2\|\xi\|_{L^2(D)}^2 = 0.$$

Eventually passing to the limit in

$$\|\nabla u_n - \nabla u\|_{L^2(D)} = \|\sigma_n^{-1/2}\xi_n - \sigma^{-1/2}\xi\|_{L^2(D)} \leq \|\sigma_n^{-1/2}(\xi_n - \xi)\|_{L^2(D)} + \|(\sigma_n^{-1/2} - \sigma^{-1/2})\xi\|_{L^2(D)}$$

results in $\|\nabla u_n - \nabla u\|_{L^2(D)} \to 0$. The Poincaré inequality permits to conclude that $\|u_n - u\|_{H^1(D)} \to 0$. □

To illustrate our findings we consider (but this is only an example) the compliance

$$J(\Omega) = \int_D f u_\Omega dx + \int_{\partial D} g\gamma_0 u_\Omega ds.$$

We address the problem

$$\underset{\Omega\in\mathcal{U}}{\text{minimize}}\ J(\Omega) + \eta P_D(\Omega), \tag{3.8}$$
$$\text{with} \quad \mathcal{U} = \{\Omega \subset D \text{ measurable} : |\Omega| = V\},$$

given a target volume $0 \leq V \leq |D|$ and a coefficient $\eta > 0$. This η is often seen as a regularization parameter. It is also possible to prescribe a perimeter inequality constraint, together with the volume constraint, provided attention is paid to checking that $\mathcal{U}$ is nonempty.

Theorem 3.25 *Problem* (3.8) *admits at least a solution.*

Proof. We develop the direct method of the calculus of variations (see section 1.4) for the convergence in the sense of characteristic functions. Since $J(\Omega) \geq 0$ and $P_D(\Omega) \geq 0$ for every $\Omega \in \mathcal{U}$, $m := \inf_{\Omega\in\mathcal{U}} J(\Omega)+\eta P_D(\Omega) \geq 0$. Let (Ω_n) be a minimizing sequence. By definition, $J(\Omega_n)+\eta P_D(\Omega_n) \to m$. In particular, the sequence $(P_D(\Omega_n))$ is bounded. By Theorem 3.19, Ω_n converges to some measurable set $\Omega \subset D$, up to a non-relabeled subsequence, in the sense of characteristic functions. We now argue that Ω is a minimizer. First, let us check that $\Omega \in \mathcal{U}$. This simply stems from Proposition 3.7, since by definition $\Omega_n \in \mathcal{U}$. We now examine the cost function. By construction, we have $\sigma_{\Omega_n} \to \sigma_\Omega$ a.e. in D. Using Theorem 3.24 we infer that $J(\Omega_n) \to J(\Omega)$. On the other hand, Corollary 3.23 yields $P_D(\Omega) \leq \liminf_{n\to+\infty} P_D(\Omega_n)$. Up to a possible further subsequence, we assume that $\liminf_{n\to+\infty} P_D(\Omega_n) = \lim_{n\to+\infty} P_D(\Omega_n)$. We arrive at

$$J(\Omega) + \eta P_D(\Omega) \leq \lim_{n\to+\infty} J(\Omega_n) + \eta \lim_{n\to+\infty} P_D(\Omega_n) = \lim_{n\to+\infty} J(\Omega_n) + \eta P_D(\Omega_n) = m.$$

This proves that Ω is a minimizer. □

To illustrate the role of the perimetric regularization we display in Figure 3.3 an example known as the optimal heater problem. The cost function is the thermal compliance augmented with a perimetric contribution as described above (here $f = 0$). The conductivities of the phases are $\alpha = 1$, $\beta = 10^{-3}$.

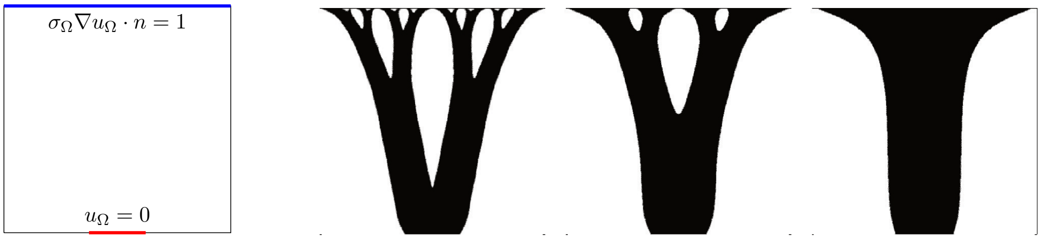


Figure 3.3: Optimal heater: boundary conditions ($\nabla u_\Omega \cdot n = 0$ when non specified) and optimal designs for increasing values of η. Here there is no true volume constraint but a fixed penalty $\lambda|\Omega|$ added to the cost.

3.4 Other regularity criteria

We have seen through the example of section 3.1.2 that it was crucial in order to enhance the existence of optimal shapes to control their regularity. We have shown that the perimeter was a good notion for that. There are other options, also associated with existence theorems. We refer to [1, 16], and only give a very brief overview.

Cone condition

Let $y \in \mathbb{R}^N$, h a unit vector and $\varepsilon > 0$. We define the truncated and unpointed cone

$$C(y, h, \varepsilon) = \{z \in \mathbb{R}^N : (z - y) \cdot h \geq \cos(\varepsilon)|z - y| \text{ and } 0 < |z - y| < \varepsilon\}.$$

We say that an open set $\Omega \subset \mathbb{R}^N$ satisfies the ε-cone condition if

$$\forall x \in \partial\Omega \ \exists h \text{ unitary s.t. } \forall y \in \bar{\Omega} \cap B(x, \varepsilon),\ C(y, h, \varepsilon) \subset \Omega.$$

It is important to note that in this definition ε is uniform with respect to $x \in \partial\Omega$.

Topological constraint

In dimension 2, it consists in prescribing an upper bound on the number of connected components of $D \setminus \Omega$, if D is a fixed bounded hold-all domain.

Comparison with a reference shape

It consists in working with the class of shapes $\Omega = T(\Omega_0)$, where Ω_0 is a fixed reference shape and T satisfies

$$\|T - \mathrm{Id}\,\|_{W^{1,\infty}(\mathbb{R}^N,\mathbb{R}^N)} + \|T^{-1} - \mathrm{Id}\,\|_{W^{1,\infty}(\mathbb{R}^N,\mathbb{R}^N)} \leq R.$$

We will see later why the norm of $W^{1,\infty}$ is appropriate to measure deformation fields.

Chapter 4

Direct and adjoint methods for the computation of derivatives

In this chapter we discuss the efficient computation of derivatives in view of their usage within optimization methods. In a first step we present general considerations in the framework of abstract optimal control problems. In a second step we specialize to governing equations in the form of elliptic boundary value problems.

4.1 Introduction

4.1.1 Definitions

An optimal control problem is typically defined with the help of the following ingredients:

- A design space X containing the admissible set $\mathcal{U}$. The elements of X are called control (or design) variables.
- A control-to state mapping $\xi \in X \mapsto u(\xi)$. It is usually defined implicitly through differential equations.
- A cost function of the form $j(\xi) = J(\xi, u(\xi))$.

We are interested in computing the Fréchet derivative of the cost function $dj(\xi)\hat{\xi}$. We recall the related notion of gradient.

Definition 4.1 *If X is a Hilbert space and $j : U \subset X \to \mathbb{R}$ is differentiable at a point ξ_0, the gradient of j at ξ_0 is the Riesz representative of the linear form $dj(\xi_0)$. It is denoted by $\nabla j(\xi_0)$ and it satisfies*

$$\langle \nabla j(\xi_0), \hat{\xi} \rangle_H = dj(\xi)\hat{\xi} \qquad \forall \hat{\xi} \in X.$$

The gradient is used to define descent directions in optimization methods. In case X is not a Hilbert space the concept of gradient is not defined. Instead we work directly with the Fréchet derivative, but we insist on the fact that it is important to have a knowledge of the full map $\hat{\xi} \mapsto dj(\xi)\hat{\xi}$, and not of a single directional derivative.

4.1.2 Examples

1. Let $A \in GL_N(\mathbb{R})$ and $F : \mathbb{R}^M \to \mathbb{R}^N$, $J : \mathbb{R}^N \to \mathbb{R}$ two (Fréchet) differentiable functions. For all $\xi \in \mathbb{R}^M$ we define

$$u(\xi) = A^{-1}F(\xi).$$

Of course, when N is large, we do not compute A^{-1} and in order to compute $u(\xi)$ we solve the linear system $Au(\xi) = F(\xi)$. It is important to keep in mind that A^{-1} is a mere mathematical object, usually almost impossible to compute and to store, since it may be full while A is sparse. We investigate the sensitivity of the cost

$$j(\xi) = J(u(\xi)).$$

This is easy to do: we simply apply the chain rule. We first differentiate the state:

$$du(\xi)\hat{\xi} = A^{-1}(dF(\xi)\hat{\xi}) = A^{-1}DF(\xi)\hat{\xi},$$

with the Jacobian matrix $DF(\xi)$. Then

$$dj(\xi)\hat{\xi} = dJ(u(\xi))(du(\xi)\hat{\xi}) = \langle \nabla J(u(\xi)), du(\xi)\hat{\xi}\rangle_{\mathbb{R}^N}, \tag{4.1}$$

using the gradient of J relative to the canonical inner product $\langle\cdot,\cdot\rangle_{\mathbb{R}^N}$ of $\mathbb{R}^N$. From the numerical point of view, this expression is straightforward to compute once the quantity $du(\xi)\hat{\xi}$ is known. But to obtain this latter one, one needs to solve a linear system of matrix A. This is perfectly doable, but in order to derive the full gradient of j one needs to do that for a family of vectors $\hat{\xi}$ spanning $\mathbb{R}^M$, preferably the canonical basis $(e_1, \cdots, e_M)$.

A better idea is to rearrange the calculations as follows:

$$dj(\xi)\hat{\xi} = \langle \nabla J(u(\xi)), A^{-1}DF(\xi)\hat{\xi}\rangle_{\mathbb{R}^N} = \langle A^{-\top}\nabla J(u(\xi)), DF(\xi)\hat{\xi}\rangle_{\mathbb{R}^N}.$$

Introducing the adjoint state

$$v(\xi) = A^{-\top}\nabla J(u(\xi))$$

we obtain

$$dj(\xi)\hat{\xi} = \langle v(\xi), DF(\xi)\hat{\xi}\rangle_{\mathbb{R}^N} = \langle DF(\xi)^\top v(\xi), \hat{\xi}\rangle_{\mathbb{R}^M}.$$

We identify the gradient

$$\nabla j(\xi) = DF(\xi)^\top v(\xi).$$

To compute this gradient, it is enough to solve the direct system for $u(\xi)$ and the adjoint system with matrix $A^\top$ for $v(\xi)$.

Suppose now that we have n functions $j_i(\xi) = J_i(u(\xi))$, $i = 1\cdots n$. In order to compute the n gradients $\nabla j_i(\xi) = J_i(u(\xi))$ by the adjoint method, one needs to compute the n adjoint states $v_i(\xi) = A^{-T}\nabla J_i(u(\xi))$. In contrast, the effort of the direct method remains dominated by the computation of the M derivatives $du(\xi)e_i$. In conclusion, the choice of the method is driven by the comparison between the number of design variables and the number of output variables. When the design space is a discrete approximation of an infinite dimensional space whereas the objective is single-valued, or even a vector of "small" dimension, the adjoint method is undoubtedly the method of choice.

2. Consider now a state defined by

$$u(\xi) = A(\xi)^{-1}F$$

where $F \in \mathbb{R}^N$ is fixed and $\xi \in U \subset \mathbb{R}^M \to A(\xi) \in GL_N(\mathbb{R})$ is a differentiable map, with U an open subset of $\mathbb{R}^M$. By the chain rule and (1.1) we obtain

$$du(\xi)\hat{\xi} = -A(\xi)^{-1}(dA(\xi)\hat{\xi})A(\xi)^{-1}F = -A(\xi)^{-1}(dA(\xi)\hat{\xi})u(\xi).$$

Here again, this computation requires to solve a specific linear system for each $\hat{\xi}$. Plugging this expression into (4.1) leads to

$$dj(\xi)\hat{\xi} = \langle \nabla J(u(\xi)), du(\xi)\hat{\xi}\rangle_{\mathbb{R}^N} = -\langle \nabla J(u(\xi)), A(\xi)^{-1}(dA(\xi)\hat{\xi})u(\xi)\rangle_{\mathbb{R}^N}.$$

Using the adjoint state

$$v(\xi) = -A(\xi)^{-\top}\nabla J(u(\xi))$$

this rewrites

$$dj(\xi)\hat{\xi} = \langle v(\xi), (dA(\xi)\hat{\xi})u(\xi)\rangle_{\mathbb{R}^N}.$$

In this form, there is no more additional system to solve to compute $dj(\xi)\hat{\xi}$ for every $\hat{\xi} \in \mathbb{R}^M$. Note that here further information on $A(\xi)$ is needed to identify the gradient $\nabla j(\xi)$.

4.1.3 General approach

We place ourselves in the framework of the implicit function theorem, with a control-to-state mapping defined through the relation

$$F(\xi, u(\xi)) = 0.$$

We denote by

$$d_uF(\xi, u) : \hat{u} \mapsto dF(\xi, u)(0, \hat{u}), \qquad d_\xi F(\xi, u) : \hat{\xi} \mapsto dF(\xi, u)(\hat{\xi}, 0)$$

the partial Fréchet derivatives of F with respect to u and ξ, respectively.

Proposition 4.2 *Let X, Y, Z be Banach spaces, $\mathcal{O}$ be an open subset of $X \times Y$, $F : \mathcal{O} \to Z$ be continuously Fréchet differentiable. Let $(\xi_0, u_0) \in \mathcal{O}$ such that $F(\xi_0, u_0) = 0$ and the partial Fréchet derivative $d_uF(\xi_0, u_0)$ is an isomorphism. There exists open neighborhoods U and V of ξ_0 and u_0, respectively, and a Fréchet differentiable function $\xi \in U \mapsto u(\xi) \in V$ such that*

$$\forall (\xi, u) \in U \times V, \qquad F(\xi, u) = 0 \Leftrightarrow u = u(\xi).$$

We have

$$du(\xi_0)\hat{\xi} = -(d_uF(\xi_0, u_0))^{-1}(d_\xi F(\xi_0, u_0)\hat{\xi}).$$

If in addition Y and Z are Hilbert spaces, $J : U \times V \to \mathbb{R}$ is a Fréchet differentiable function and $j(\xi) = J(\xi, u(\xi))$ then j is Fréchet differentiable at ξ_0. Defining the adjoint state

$$v_0 = -d_uF(\xi_0, u_0)^{-*}\nabla_u J(\xi_0, u_0)$$

we have

$$dj(\xi_0)\hat{\xi} = d_\xi J(\xi_0, u_0)\hat{\xi} + \langle v_0, d_\xi F(\xi_0, u_0)\hat{\xi}\rangle_Z.$$

Proof. The existence of the map $\xi \mapsto u(\xi)$ is a direct application of the implicit function theorem (Theorem 1.8). Differentiating the relation

$$F(\xi, u(\xi)) = 0$$

provides

$$d_\xi F(\xi, u(\xi))\hat{\xi} + d_uF(\xi, u(\xi))(du(\xi)\hat{\xi}) = 0,$$

from which we infer the derivative of the state. Now we differentiate the cost by

$$\begin{aligned}
dj(\xi)\hat{\xi} &= d_\xi J(\xi, u(\xi))\hat{\xi} + d_uJ(\xi, u(\xi))(du(\xi)\hat{\xi}) \\
&= d_\xi J(\xi, u(\xi))\hat{\xi} + \langle \nabla_u J(\xi, u(\xi)), du(\xi)\hat{\xi}\rangle_Y \\
&= d_\xi J(\xi, u(\xi))\hat{\xi} - \langle \nabla_u J(\xi, u(\xi)), d_uF(\xi, u(\xi))^{-1}(d_\xi F(\xi, u(\xi))\hat{\xi})\rangle_Y \\
&= d_\xi J(\xi, u(\xi))\hat{\xi} - \langle d_uF(\xi, u(\xi))^{-*}\nabla_u J(\xi, u(\xi)), d_\xi F(\xi, u(\xi))\hat{\xi}\rangle_Z.
\end{aligned}$$

□

It is remarkable that, although the (direct) state solves an a priori nonlinear problem, the adjoint state is always defined as the solution of a linear problem, namely

$$d_uF(\xi_0, u_0)^* v_0 = -\nabla_u J(\xi_0, u_0).$$

This is actually not always a good news: the adjoint problem may not have any physical meaning, whereby appropriate solvers may be unavailable. This is a severe limitation as to the application of the adjoint approach in industrial applications.

In the preceding calculation, involving the adjoint state seemed very natural. But in order to obtain computable expressions when the function F encodes boundary value problems, some efforts remain to be made. To prepare to this, we introduce the very convenient concept of **Lagrangian**. The Lagrangian function is defined by

$$\mathcal{L} : (\xi, u, v) \in U \times V \times Z \mapsto J(\xi, u) + \langle F(\xi, u), v\rangle_Z. \tag{4.2}$$

Then we immediately check the identity

$$dj(\xi_0)\hat{\xi} = d_\xi \mathcal{L}(\xi_0, u_0, v_0)\hat{\xi}.$$

In addition, we observe that the direct and adjoint problems are equivalent to the stationarity relations $d_v\mathcal{L}(\xi_0, u_0, v_0) = 0$ and $d_u\mathcal{L}(\xi_0, u_0, v_0) = 0$, respectively. We see in the construction that the adjoint state plays the role of Lagrange multiplier for the constraint $F(\xi, u) = 0$, however it is not assumed here that ξ is optimal.

4.2 Direct and adjoint methods for elliptic boundary value problems

4.2.1 A prototype problem

We consider a parametric family of problems of the form

$$\begin{array}{lll} u(\xi) \in H & \forall \xi \in \mathcal{O} & \\ a(\xi, u(\xi), \varphi) = l(\xi, \varphi) & \forall \varphi \in H & \forall \xi \in \mathcal{O} \end{array}$$

where

- H is a Hilbert space,
- $\mathcal{O}$ is an open subset of a Banach space X,
- for all $\xi \in \mathcal{O}$ the map $a(\xi, \cdot, \cdot)$ is a continuous and coercive bilinear form on H, and the map $l(\xi, \cdot)$ is a continuous linear form on H.

For all $\xi \in \mathcal{O}$ we define the maps $A(\xi) \in \mathcal{L}(H, H')$ and $L(\xi) \in H'$ by

$$\langle A(\xi)\psi, \varphi\rangle_{H',H} = a(\xi, \psi, \varphi), \qquad \langle L(\xi), \varphi\rangle_{H',H} = l(\xi, \varphi) \qquad \forall \varphi, \psi \in H.$$

By Lax-Milgram's theorem, $A(\xi)$ is an isomorphism and we have

$$u(\xi) = A(\xi)^{-1}L(\xi). \tag{4.3}$$

Proposition 4.3 *If the maps $\xi \mapsto A(\xi)$ and $\xi \mapsto L(\xi)$ are Fréchet differentiable in $\mathcal{O}$ then the map $\xi \mapsto u(\xi)$ is Fréchet differentiable in $\mathcal{O}$.*

PROOF. It directly follows from the differentiability of the map $A \in \text{isom}(H, H') \mapsto A^{-1}$ (Proposition 1.9). □

Observe that it is in fact sufficient that the maps $\xi \mapsto A(\xi)$ and $\xi \mapsto L(\xi)$ be differentiable at some point $\xi_0 \in \mathcal{O}$ to get the differentiability of the map $\xi \mapsto u(\xi)$ at ξ_0. This is in contrast with an argument based on the implicit function theorem, which would also require continuous differentiability.

Consider now a cost function

$$j(\xi) = J(\xi, u(\xi)),$$

with J differentiable on $\mathcal{O} \times H$. Of course, j is differentiable in $\mathcal{O}$ by composition. Our purpose is to obtain a convenient expression of the derivative.

4.2.2 Direct method

The direct method simply consists in applying the chain rule:

$$dj(\xi)\hat{\xi} = d_\xi J(\xi, u(\xi))\hat{\xi} + d_u J(\xi, u(\xi))(du(\xi)\hat{\xi}).$$

In order to determine the derivative of the state, an option is to differentiate the relation (4.3). More easily, we differentiate the relation

$$a(\xi, u(\xi), \varphi) = l(\xi, \varphi).$$

This entails

$$d_\xi a(\xi, u(\xi), \varphi)\hat{\xi} + a(\xi, du(\xi)\hat{\xi}, \varphi) = d_\xi l(\xi, \varphi)\hat{\xi}.$$

Therefore $du(\xi)\hat{\xi} \in H$ is the unique solution of

$$a(\xi, du(\xi)\hat{\xi}, \varphi) = d_\xi l(\xi, \varphi)\hat{\xi} - d_\xi a(\xi, u(\xi), \varphi)\hat{\xi} \qquad \forall \varphi \in H.$$

As pointed out in the examples, the drawback of this approach is that the above problem needs in principle to be solved for every $\hat{\xi}$, or at least on a spanning family, in order to have a complete knowledge of $du(\xi)$.

4.2.3 Adjoint method

In view of (4.2) we introduce the Lagrangian

$$\mathcal{L}(\xi, u, v) = J(\xi, u) + a(\xi, u, v) - l(\xi, v).$$

Theorem 4.4 *Under the assumptions of Proposition 4.3 we have for all* $\xi \in \mathcal{O}$

$$\boxed{dj(\xi)\hat{\xi} = d_\xi \mathcal{L}(\xi, u(\xi), v(\xi))\hat{\xi}}$$

where the direct state $u(\xi) \in H$ *and the adjoint state* $v(\xi) \in H$ *are unambiguously defined by*

$$d_v \mathcal{L}(\xi, u(\xi), v(\xi)) = d_u \mathcal{L}(\xi, u(\xi), v(\xi)) = 0.$$

Proof. The first stationarity relation reads

$$0 = d_v \mathcal{L}(\xi, u(\xi), v(\xi))\hat{v} = a(\xi, u(\xi), \hat{v}) - l(\xi, \hat{v}) \qquad \forall \hat{v} \in H,$$

which is the variational formulation for $u(\xi)$. The second stationarity relation reads

$$0 = d_u \mathcal{L}(\xi, u(\xi), v(\xi))\hat{u} = d_u J(\xi, u)\hat{u} + a(\xi, \hat{u}, v(\xi)) \qquad \forall \hat{u} \in H,$$

which admits a unique solution $v(\xi)$ by the Lax-Milgram theorem. We now note that

$$j(\xi) = \mathcal{L}(\xi, u(\xi), \varphi) \qquad \forall \varphi \in H,\ \forall \xi \in \mathcal{O}.$$

This results in

$$dj(\xi)\hat{\xi} = d_\xi \mathcal{L}(\xi, u(\xi), \varphi)\hat{\xi} + d_u \mathcal{L}(\xi, u(\xi), \varphi)(du(\xi)\hat{\xi}).$$

Choosing $\varphi = v(\xi)$ yields the claim by cancellation of the last term. □

Remark 4.5 *If we consider a non-homogeneous Dirichlet problem of the form*

$$\begin{array}{ll} u(\xi) \in \{w\} + H & \forall \xi \in \mathcal{O} \\ a(\xi, u(\xi), \varphi) = l(\xi, \varphi) & \forall \varphi \in H \qquad \forall \xi \in \mathcal{O}, \end{array}$$

where H *is a closed linear subspace of a Hilbert space* $\mathcal{H}$ *and* $w \in \mathcal{H}$, *then it is immediately seen that the above procedure remains almost unchanged. The adjoint state is still defined in* H *and the stationarity conditions remain the same.*

4.2.4 Example

Consider the membrane problem (2.1). Here the variational formulation is defined by

$$a(h,u,v)=\int_\Omega h\nabla u\cdot\nabla v dx,\qquad l(v)=\int_\Omega fv dx,\qquad u,v\in H=H_0^1(\Omega).$$

The thickness h is supposed to belong to $L^\infty(\Omega)$, and more precisely to the open subset

$$\begin{aligned}\mathcal{O} &= \{h\in L^\infty(\Omega) : \operatorname{essinf} h>0\}\\ &= \{h\in L^\infty(\Omega) : \exists\alpha>0, h(x)\geq\alpha \text{ a.e. } x\in\Omega\}.\end{aligned}$$

In this case the right hand side l is independent of h, and the left hand side a is a trilinear form such that $\|A(h)\|_{\mathcal{L}(H,H')}=\|h\|_{L^\infty(D)}$. It is then straightforward to see that Proposition 4.3 applies.

As cost function, let us consider the compliance

$$j(h)=J(u(h))=\int_\Omega fu(h)dx.$$

We assemble the Lagrangian

$$\mathcal{L}(h,u,v)=\int_\Omega fudx+\int_\Omega h\nabla u\cdot\nabla vdx-\int_\Omega fvdx.$$

It admits the derivatives

$$d_h\mathcal{L}(h,u,v)\hat{h}=\int_\Omega \hat{h}\nabla u\cdot\nabla vdx,$$

$$d_u\mathcal{L}(h,u,v)\hat{u}=\int_\Omega f\hat{u}dx+\int_\Omega h\nabla\hat{u}\cdot\nabla vdx,\qquad d_v\mathcal{L}(h,u,v)\hat{v}=\int_\Omega h\nabla u\cdot\nabla\hat{v}dx-\int_\Omega f\hat{v}dx.$$

We obtain by Theorem 4.4 the derivative

$$dj(h)\hat{h}=d_h\mathcal{L}(h,u(h),v(h))\hat{h}=\int_\Omega \hat{h}\nabla u(h)\cdot\nabla v(h)dx,$$

with the adjoint state $v(h)$ solution of

$$d_u\mathcal{L}(h,u(h),v(h))\hat{u}=\int_\Omega f\hat{u}dx+\int_\Omega h\nabla\hat{u}\cdot\nabla v(h)dx=0\qquad\forall\hat{u}\in H.$$

Rewriting this latter equation as

$$\int_\Omega h\nabla v(h)\cdot\nabla\hat{u}dx=-\int_\Omega f\hat{u}dx\qquad\forall\hat{u}\in H,$$

we infer by uniqueness that $v(h)=-u(h)$. We say that it is a self-adjoint problem. This is typical of the compliance. Altogether we arrive at

$$dj(h)\hat{h}=-\int_\Omega \hat{h}|\nabla u(h)|^2dx.$$

Observe that this quantity has a sign: the compliance decreases when the thickness increases. That was expected.

4.3 Case of eigenvalues

4.3.1 General framework

We now consider a parametric family of generalized eigenvalue problems of the form

$$\begin{array}{l} (\lambda(\xi), u(\xi)) \in \mathbb{R} \times H \qquad \forall \xi \in \mathcal{O} \\ a(\xi, u(\xi), \varphi) = \lambda(\xi) b(\xi, u(\xi), \varphi) \qquad \forall \varphi \in H, \qquad \forall \xi \in \mathcal{O} \\ b(\xi, u(\xi), u(\xi)) = 1 \qquad \forall \xi \in \mathcal{O} \end{array}$$

where

- H is a Hilbert space,
- $\mathcal{O}$ is an open subset of a Banach space X,
- for all $\xi \in \mathcal{O}$ the map $a(\xi, \cdot, \cdot)$ is a symmetric, continuous and coercive bilinear form in H, and the map $b(\xi, \cdot, \cdot)$ is a symmetric, continuous bilinear form on H.

This is called a generalized eigenvalue problem because of the right hand side b that can take various forms. This b is also used here to normalize eigenvectors.

For all $\xi \in \mathcal{O}$ we define the maps $A(\xi), B(\xi) \in \mathcal{L}(H, H')$ by

$$\langle A(\xi)\psi, \varphi\rangle_{H',H} = a(\xi, \psi, \varphi), \qquad \langle B(\xi)\psi, \varphi\rangle_{H',H} = b(\xi, \psi, \varphi) \qquad \forall \varphi, \psi \in H.$$

Theorem 4.6 *Suppose that the maps $\xi \mapsto A(\xi)$ and $\xi \mapsto B(\xi)$ are continuously Fréchet differentiable in $\mathcal{O}$, and that $B(\xi)$ is compact for all $\xi \in \mathcal{O}$. Let $(\xi_0, \lambda_0, u_0) \in \mathcal{O} \times \mathbb{R} \times H$ be such that*

$$\left\{ \begin{array}{l} b(\xi_0, u_0, u_0) = 1 \\ a(\xi_0, u_0, \varphi) = \lambda_0 b(\xi_0, u_0, \varphi) \qquad \forall \varphi \in H, \end{array} \right.$$

and such that the eigenspace

$$E_0 = \{u \in H \ \text{ s.t. } \ a(\xi_0, u, \varphi) = \lambda_0 b(\xi_0, u, \varphi) \quad \forall \varphi \in H\}$$

is of dimension 1 (simple eigenvalue). There exists a neighborhood $\mathcal{O}' \subset \mathcal{O}$ of ξ_0 and continuously differentiable functions

$$\xi \in \mathcal{O}' \mapsto \lambda(\xi) \in \mathbb{R}, \qquad \xi \in \mathcal{O}' \mapsto u(\xi) \in H,$$

such that

$$\left\{ \begin{array}{l} b(\xi, u(\xi), u(\xi)) = 1 \qquad \forall \xi \in \mathcal{O}' \\ a(\xi, u(\xi), \varphi) = \lambda(\xi) b(\xi, u(\xi), \varphi) \qquad \forall \varphi \in H \qquad \forall \xi \in \mathcal{O}'. \end{array} \right.$$

Proof. We define the map

$$\begin{array}{rcl} F : \mathcal{O} \times \mathbb{R} \times H & \to & \mathbb{R} \times H' \\ (\xi, \lambda, u) & \mapsto & \big(\langle B(\xi)u, u\rangle_{H',H} - 1,\ A(\xi)u - \lambda B(\xi)u\big)\,. \end{array}$$

We have by construction $F(\xi_0, \lambda_0, u_0) = 0$. Moreover, F is continuously differentiable and we have

$$\begin{array}{l} dF(\xi, u, \lambda)(\hat{\xi}, \hat{\lambda}, \hat{u}) \\ = \Big(\langle (d_\xi B(\xi)\hat{\xi})u, u\rangle_{H',H} + 2\langle B(\xi)u, \hat{u}\rangle_{H',H},\ (dA(\xi)\hat{\xi})u - \lambda(dB(\xi)\hat{\xi})u - \hat{\lambda}B(\xi)u + A(\xi)\hat{u} - \lambda B(\xi)\hat{u}\Big)\,. \end{array}$$

Call

$$\begin{array}{rl} M(\hat{\lambda}, \hat{u}) = d_{(\lambda,u)}F(\xi_0, u_0, \lambda_0)(\hat{\lambda}, \hat{u}) = dF(\xi_0, u_0, \lambda_0)(0, \hat{\lambda}, \hat{u}) & \\ = \Big(2\langle B(\xi_0)u_0, \hat{u}\rangle_{H',H},\ -\hat{\lambda}B(\xi_0)u_0 + A(\xi_0)\hat{u} - \lambda_0 B(\xi_0)\hat{u}\Big)\,. & \end{array}$$

Suppose that $M(\hat{\lambda}, \hat{u}) = 0$. This is equivalent to

$$\begin{cases} \langle B(\xi_0)u_0, \hat{u}\rangle_{H',H} = 0 \\ -\hat{\lambda}B(\xi_0)u_0 + A(\xi_0)\hat{u} - \lambda_0 B(\xi_0)\hat{u} = 0. \end{cases}$$

Evaluating the second row against u_0 yields, using symmetry and $\langle B(\xi_0)u_0, u_0\rangle = 1$, $\hat{\lambda} = 0$. This implies that

$$A(\xi_0)\hat{u} = \lambda_0 B(\xi_0)\hat{u},$$

i.e. $\hat{u} \in E_0$. By assumption, this space is spanned by u_0, thus $\hat{u}$ is colinear to u_0. Therefore the condition $\langle B(\xi_0)u_0, \hat{u}\rangle_{H',H} = 0$ entails $\hat{u} = 0$. We have shown that $d_{(\lambda,u)}F(\xi_0, u_0, \lambda_0)$ is injective.

We now prove surjectivity, which is more delicate but needed as long as H is not of finite dimension. Let $(r, S) \in \mathbb{R} \times H'$. By Lax-Milgram's theorem we know that $A(\xi_0)$ is an isomorphism. Define $T \in \mathcal{L}(H)$ by

$$T(\hat{u}) = A(\xi_0)^{-1}\big(\lambda_0 B(\xi_0)\hat{u} + \langle A(\xi_0)u_0, \hat{u}\rangle_{H',H} B(\xi_0)u_0\big).$$

It appears from this definition and the assumptions that T is compact. Suppose that $(\mathrm{Id}_H - T)\hat{u} = 0$. Then

$$A(\xi_0)\hat{u} - \lambda_0 B(\xi_0)\hat{u} - \langle A(\xi_0)u_0, \hat{u}\rangle_{H',H} B(\xi_0)u_0 = 0. \tag{4.4}$$

Evaluating against u_0 results in

$$\langle A(\xi_0)u_0 - \lambda_0 B(\xi_0)u_0, \hat{u}\rangle_{H',H} - \langle A(\xi_0)u_0, \hat{u}\rangle_{H',H} = 0,$$

hence

$$\lambda_0 \langle B(\xi_0)u_0, \hat{u}\rangle_{H',H} = 0.$$

This in turn yields $\langle A(\xi_0)u_0, \hat{u}\rangle_{H',H} = 0$, since $A(\xi_0)u_0 = \lambda_0 B(\xi_0)u_0$. Hence (4.4) entails $\hat{u} \in E_0$. As argued before, in view of $\lambda_0 = \langle A(\xi_0)u_0, u_0\rangle \neq 0$, this implies that $\hat{u} = 0$. We have shown that $\mathrm{Id}_H - T$ is injective. By Fredholm's alternative it is an isomorphism. Therefore there exists $\hat{u} \in H$ such that

$$(\mathrm{Id}_H - T)\hat{u} = A(\xi_0)^{-1}\left(S - \left(\frac{\lambda_0 r}{2} + \langle S, u_0\rangle_{H',H}\right) B(\xi_0)u_0\right).$$

This yields

$$A(\xi_0)\hat{u} - \lambda_0 B(\xi_0)\hat{u} - \langle A(\xi_0)u_0, \hat{u}\rangle_{H',H} B(\xi_0)u_0 = S - \left(\frac{\lambda_0 r}{2} + \langle S, u_0\rangle_{H',H}\right) B(\xi_0)u_0. \tag{4.5}$$

We set

$$\hat{\lambda} = \langle A(\xi_0)u_0, \hat{u}\rangle_{H',H} - \left(\frac{\lambda_0 r}{2} + \langle S, u_0\rangle_{H',H}\right),$$

so that

$$A(\xi_0)\hat{u} - \lambda_0 B(\xi_0)\hat{u} - \hat{\lambda} B(\xi_0)u_0 = S.$$

In addition, (4.5) against u_0 results in

$$-\lambda_0 \langle B(\xi_0)u_0, \hat{u}\rangle_{H',H} = -\frac{\lambda_0 r}{2}.$$

Since $\lambda_0 \neq 0$ we have obtained $M(\hat{\lambda}, \hat{u}) = (r, S)$.

We are now in position to apply the implicit function theorem to F, which provides the claim. □

We now proceed to the evaluation of the derivative $d\lambda(\xi)\hat{\xi}$.

Theorem 4.7 *Under the assumptions of Theorem 4.6 we have*

$$d\lambda(\xi_0)\hat{\xi} = d_\xi a(\xi_0, u_0, u_0)\hat{\xi} - \lambda_0 d_\xi b(\xi_0, u_0, u_0)\hat{\xi}.$$

PROOF. We differentiate the relation

$$a(\xi, u(\xi), \varphi) = \lambda(\xi) b(\xi, u(\xi), \varphi) \qquad \forall \varphi \in H, \qquad \forall \xi \in \mathcal{O}.$$

This entails

$$d_\xi a(\xi, u(\xi), \varphi)\hat{\xi} + a(\xi, du(\xi)\hat{\xi}, \varphi) = (d\lambda(\xi)\hat{\xi}) b(\xi, u(\xi), \varphi) + \lambda(\xi) \left(d_\xi b(\xi, u(\xi), \varphi)\hat{\xi} + b(\xi, du(\xi)\hat{\xi}, \varphi) \right).$$

Evaluating at $\xi = \xi_0$ and choosing $\varphi = u_0$ results in

$$d_\xi a(\xi_0, u_0, u_0)\hat{\xi} + a(\xi_0, du(\xi_0)\hat{\xi}, u_0) = d\lambda(\xi_0)\hat{\xi} + \lambda_0 \left(d_\xi b(\xi_0, u_0, u_0)\hat{\xi} + b(\xi_0, du(\xi_0)\hat{\xi}, u_0) \right).$$

By symmetry, this simplifies as

$$d_\xi a(\xi_0, u_0, u_0)\hat{\xi} = d\lambda(\xi_0)\hat{\xi} + \lambda_0 d_\xi b(\xi_0, u_0, u_0)\hat{\xi}.$$

□

Note that there is no need of adjoint state in order to differentiate an eigenvalue. We only need a corresponding eigenvector.

4.3.2 Example

Consider again the membrane problem, with eigenvalues / eigenfunctions solving

$$\int_\Omega h \nabla u(h) \cdot \nabla \varphi dx = \lambda(h) \int_\Omega h u(h) \varphi dx \qquad \forall \varphi \in H = H_0^1(\Omega),$$

for $u(h) \in H$, i.e. in strong form

$$\begin{cases} -\operatorname{div}(h \nabla u(h)) = \lambda(h) h u(h) \text{ in } \Omega \\ u(h) = 0 \text{ on } \partial\Omega. \end{cases}$$

We have incorporated the thickness in the inertial term, which is a reasonable modeling assumption. We set

$$a(h, u, v) = \int_\Omega h \nabla u \cdot \nabla v dx, \qquad b(h, u, v) = \int_\Omega h u v dx, \qquad h \in L^\infty(\Omega), \ u, v \in H.$$

Through identifying $L^2(\Omega)$ with its dual we have $B(h)u = hu$, hence $B(h) \in \mathcal{L}(L^2(\Omega))$. It follows from Rellich's theorem that $B(h) : H \to H'$ is compact. All the assumptions of Theorems 4.6 and 4.7 are satisfied and we obtain

$$d\lambda(h_0)\hat{h} = \int_\Omega \hat{h} |\nabla u_0|^2 dx - \lambda_0 \int_\Omega \hat{h} u_0^2 dx = \int_\Omega \hat{h} (|\nabla u_0|^2 - \lambda_0 u_0^2) dx,$$

where (λ_0, u_0) satisfy

$$\int_\Omega h_0 \nabla u_0 \cdot \nabla \varphi dx = \lambda_0 \int_\Omega h_0 u_0 \varphi dx \ \forall \varphi \in H, \qquad \int_\Omega h_0 |u_0|^2 dx = 1.$$

Chapter 5

Shape derivative

5.1 Deformations and definition of the shape derivative

5.1.1 Deformations

The approach we present here was introduced by Murat and Simon [17]. An alternative approach, the so-called speed method, is developed in [21] and briefly discussed in subsection 5.3.1.

Let Ω_0 be an open subset of $\mathbb{R}^N$. We want to represent our unknown domain Ω as a deformation of Ω_0, with a deformation smooth enough in order perform a rigorous sensitivity analysis. Without loss of generality we can write

$$\Omega = T(\Omega_0), \qquad T = \operatorname{Id} + \theta, \qquad \theta : \mathbb{R}^N \to \mathbb{R}^N,$$

$$\text{i.e.,} \qquad \Omega = \{T(x), x \in \Omega_0\}, \qquad T(x) = x + \theta(x).$$

This decomposition of the deformation is somewhat natural: θ is then the displacement field. We will later need to perform change of variables in integrals, therefore we require that T be injective. We will also perform differential calculus, hence we also need to work in an open subset of a normed vector space. These remarks will motivate our choice of deformation fields, in view of the following results.

We equip $\mathbb{R}^N$ with its Euclidean norm $|\cdot|$, $\mathcal{M}_N(\mathbb{R})$ with the induced norm $|M| = \sup_{|x|=1} |Mx|$, $L^\infty(\mathbb{R}^N, \mathbb{R}^N)$ with the norm $\|\theta\|_{L^\infty(\mathbb{R}^N,\mathbb{R}^N)} = \| \, |\theta| \, \|_{L^\infty(\mathbb{R}^N)}$, $L^\infty(\mathbb{R}^N, \mathcal{M}_N(\mathbb{R}))$ with the norm $\|M\|_{L^\infty(\mathbb{R}^N,\mathcal{M}_N(\mathbb{R}))} = \| \, |M| \, \|_{L^\infty(\mathbb{R}^N)}$, and $W^{1,\infty}(\mathbb{R}^N, \mathbb{R}^N)$ with the norm

$$\|\theta\|_{W^{1,\infty}(\mathbb{R}^N,\mathbb{R}^N)} = \|\theta\|_{L^\infty(\mathbb{R}^N,\mathbb{R}^N)} + \|D\theta\|_{L^\infty(\mathbb{R}^N,\mathcal{M}_N(\mathbb{R}))}.$$

Lemma 5.1 *For all $\theta \in W^{1,\infty}(\mathbb{R}^N, \mathbb{R}^N)$ it holds*

$$|\theta(x) - \theta(y)| \le \|D\theta\|_{L^\infty(\mathbb{R}^N,\mathcal{M}_N(\mathbb{R}))} |x - y| \qquad \text{a.e. } x, y \in \mathbb{R}^N.$$

Therefore, θ admits a Lipschitz-continuous representative.

Proof. Let B be an arbitrary open ball of $\mathbb{R}^N$. Consider first a function $u \in W^{1,\infty}(\mathbb{R}^N)$. We have for any unit vector e of $\mathbb{R}^N$

$$\int_B (|u| + |\nabla u \cdot e|) dx < +\infty.$$

Fubini's theorem yields that for a.e. $a \in e^\perp$

$$\int_{L_{e,a} \cap B} (|u| + |\nabla u \cdot e|) dx < +\infty,$$

where now we consider the integral on the line $L_{e,a} = a + \mathbb{R}e$. Let $u_{e,a}(t) = u(a + te)$ and $t_{e,a} \ge 0$ such that $L_{e,a} \cap B = \{a + te, -t_{e,a} < t < t_{e,a}\}$. We have by definition of the weak derivative

$$-\int_{e^\perp} \int_{\mathbb{R}} u(a + te) \operatorname{div} \Phi(a + te) dt da = \int_{e^\perp} \int_{\mathbb{R}} \nabla u(a + te) \cdot \Phi(a + te) dt da \qquad \forall \Phi \in \mathcal{C}_c^1(\mathbb{R}^N, \mathbb{R}^N).$$

Choosing test functions of the form $\Phi(a+te)=\eta(a)\psi(t)e$ shows that $u_{e,a}$ is weakly differentiable for a.e. $a\in e^\perp$ with $u'_{e,a}(t)=\nabla u(a+te)\cdot e$. Then (see e.g. [8] Thm VIII.2) $u_{a,e}$ is a.e. equal to an absolutely continuous function $\tilde{u}_{e,a}$ satisfying

$$\tilde{u}_{e,a}(t_2)-\tilde{u}_{e,a}(t_1)=\int_{t_1}^{t_2}u'_{e,a}(t)dt\qquad\forall t_1,t_2\in]-t_{e,a},t_{e,a}[.$$

This rewrites as

$$u(a+t_2e)-u(a+t_1e)=\int_{t_1}^{t_2}\nabla u(a+te)\cdot edt\qquad\text{a.e. } t_1,t_2\in]-t_{e,a},t_{e,a}[.$$

Let now $\theta\in W^{1,\infty}(\mathbb{R}^N,\mathbb{R}^N)$. Applying the previous equality to the function $\theta\cdot\hat{e}$, for an arbitrary unit vector $\hat{e}\in\mathbb{R}^N$, reveals that

$$\theta(a+t_2e)\cdot\hat{e}-\theta(a+t_1e)\cdot\hat{e}=\int_{t_1}^{t_2}D\theta(a+te)e\cdot\hat{e}dt\qquad\text{a.e. } t_1,t_2\in]-t_{e,a},t_{e,a}[.$$

We know that

$$|D\theta(a+te)e\cdot\hat{e}|\leq\|D\theta\|_{L^\infty(\mathbb{R}^N,\mathcal{M}_N(\mathbb{R}))}\qquad\text{a.e. }(a,t)\in e^\perp\times\mathbb{R}.$$

This results in

$$(\theta(a+t_2e)-\theta(a+t_1e))\cdot\hat{e}\leq\|D\theta\|_{L^\infty(\mathbb{R}^N,\mathcal{M}_N(\mathbb{R}))}|t_2-t_1|\qquad\text{a.e. } a\in e^\perp,\text{a.e. } t_1,t_2\in]-t_{e,a},t_{e,a}[,$$

which can be rewritten as

$$(\theta(x_2)-\theta(x_1))\cdot\hat{e}\leq\|D\theta\|_{L^\infty(\mathbb{R}^N,\mathcal{M}_N(\mathbb{R}))}|x_2-x_1|\qquad\text{a.e. } a\in e^\perp,\text{a.e. } x_1,x_2\in L_{e,a}.$$

As e is arbitrary, this holds true for a.e. $x_1,x_2\in B$. Taking now the supremum in $\hat{e}$ yields

$$|\theta(x_2)-\theta(x_1)|\leq\|D\theta\|_{L^\infty(\mathbb{R}^N,\mathcal{M}_N(\mathbb{R}))}|x_2-x_1|\qquad\text{a.e. } x_1,x_2\in B.$$

As B is arbitrary, it extends to $\mathbb{R}^N$ as countable union of balls. □

Lemma 5.2 *If $\theta\in L^\infty(\mathbb{R}^N,\mathbb{R}^N)$ satisfies for some $k\geq 0$*

$$|\theta(x)-\theta(y)|\leq k|x-y|\qquad\text{a.e. } x,y\in\mathbb{R}^N$$

then $\theta\in W^{1,\infty}(\mathbb{R}^N,\mathbb{R}^N)$ and we have $\|D\theta\|_{L^\infty(\mathbb{R}^N,\mathcal{M}_N(\mathbb{R}))}\leq k$.

PROOF. Without loss of generality we choose a continuous representative of θ. For any $\hat{e},e,a\in\mathbb{R}^N$, $|\hat{e}|=|e|=1$, set

$$\theta_{\hat{e},e,a}(t)=\theta(a+te)\cdot\hat{e}.$$

We have that $\theta_{\hat{e},e,a}$ is k-Lpischitz continuous, in particular it is absolutely continuous. Hence it is differentiable a.e. (see e.g. [19]) with

$$\theta_{\hat{e},e,a}(t_2)-\theta_{\hat{e},e,a}(t_1)=\int_{t_1}^{t_2}\theta'_{\hat{e},e,a}(t)dt\qquad\forall t_1,t_2.$$

In particular this tells us that $\theta\cdot\hat{e}$ admits partial derivatives almost everywhere.

Then we have

$$\left|\int_{t_1}^{t_2}\theta'_{\hat{e},e,a}(t)dt\right|\leq k|t_2-t_1|\qquad\forall t_1,t_2.$$

Now we write that

$$\theta'_{\hat{e},e,a}(t)=\frac{1}{2\varepsilon}\int_{t-\varepsilon}^{t+\varepsilon}\left(\theta'_{\hat{e},e,a}(t)-\theta'_{\hat{e},e,a}(s)\right)ds+\frac{1}{2\varepsilon}\int_{t-\varepsilon}^{t+\varepsilon}\theta'_{\hat{e},e,a}(s)ds,$$

and let $\varepsilon \searrow 0$. The first integral goes to 0 for a.e. t by Lebesgue's differentiation theorem, and we obtain

$$|\theta'_{\hat{e},e,a}(t)| \leq k \qquad \text{a.e. } t.$$

From $(\nabla(\theta \cdot \hat{e})) \cdot e = \theta'_{\hat{e},e,a}(t)$ we infer $|\nabla(\theta \cdot \hat{e})| \leq k$.

Lastly this a.e. gradient is also the gradient in the sense of weak derivatives, since on the one hand it is in $L^1_{\text{loc}}(\mathbb{R}^N)$, and on the other hand we have for any $\varphi \in \mathcal{C}^1_c(\mathbb{R}^N)$

$$\int_{\mathbb{R}^N} \frac{\partial(\theta \cdot \hat{e})}{\partial x_i} \varphi dx + \int_{\mathbb{R}^N} (\theta \cdot \hat{e}) \frac{\partial \varphi}{\partial x_i} dx = \int_{\mathbb{R}^N} \frac{\partial(\theta \cdot \hat{e}\varphi)}{\partial x_i} dx = 0,$$

since $\theta \cdot \hat{e}\varphi$ is absolutely continuous with respect to x_i. □

Up to choosing continuous representatives, as we will always implicitly do in the sequel, Lemmas 5.1 and 5.2 permit to identify $W^{1,\infty}(\mathbb{R}^N, \mathbb{R}^N)$ with the set of bounded Lipschitz vector fields.

Remark 5.3 *In the proof of Lemma 5.2 we established that θ admits partial derivatives almost everywhere. It can even be proven that θ is Fréchet differentiable almost everywhere. This is Rademacher's theorem, a proof of which can be found in [13].*

Proposition 5.4 *Let $\theta \in W^{1,\infty}(\mathbb{R}^N, \mathbb{R}^N)$ be such that $\|\theta\|_{W^{1,\infty}(\mathbb{R}^N,\mathbb{R}^N)} < 1$. Then $T = \text{Id} + \theta : \mathbb{R}^N \to \mathbb{R}^N$ is a bijection. Moreover we have $T^{-1} - \text{Id} \in W^{1,\infty}(\mathbb{R}^N, \mathbb{R}^N)$.*

Proof. To prove bijectivity let us fix an arbitrary $y \in \mathbb{R}^N$ and address the equation $T(x) = y$. We reformulate equivalently as $S(x) = x$ with $S(x) = y + x - T(x) = y - \theta(x)$. We show that S is a contraction:

$$|S(x) - S(\hat{x})| = |\theta(x) - \theta(\hat{x})| \leq \|D\theta\|_{L^\infty(\mathbb{R}^N, \mathcal{M}_N(\mathbb{R}))} |x - \hat{x}|,$$

by Lemma 5.1. By assumption we have $\|D\theta\|_{L^\infty(\mathbb{R}^N, \mathcal{M}_N(\mathbb{R}))} < 1$. By the Banach fixed point theorem S admits a unique fixed point. We infer that T is a bijection.

We now prove that $T^{-1} - \text{Id} \in W^{1,\infty}(\mathbb{R}^N, \mathbb{R}^N)$. We have for any $y \in \mathbb{R}^N$, denoting $x = T^{-1}(y)$

$$(T^{-1} - \text{Id})(y) = T^{-1}(y) - y = x - T(x) = -\theta(x).$$

It follows that $T^{-1} - \text{Id} \in L^\infty(\mathbb{R}^N, \mathbb{R}^N)$, with $\|T^{-1} - \text{Id}\|_{L^\infty(\mathbb{R}^N,\mathbb{R}^N)} \leq \|\theta\|_{L^\infty(\mathbb{R}^N,\mathbb{R}^N)} < 1$. Consider now two points $y, \hat{y} \in \mathbb{R}^N$ and denote $x = T^{-1}(y)$, $\hat{x} = T^{-1}(\hat{y})$. We have

$$|(T^{-1} - \text{Id})(y) - (T^{-1} - \text{Id})(\hat{y})| = |\theta(\hat{x}) - \theta(x)| \leq \|D\theta\|_{L^\infty(\mathbb{R}^N, \mathcal{M}_N(\mathbb{R}))} |x - \hat{x}|.$$

This yields

$$|(T^{-1} - \text{Id})(y) - (T^{-1} - \text{Id})(\hat{y})| \leq \|D\theta\|_{L^\infty(\mathbb{R}^N, \mathcal{M}_N(\mathbb{R}))} \left|(T^{-1} - \text{Id})(y) - (T^{-1} - \text{Id})(\hat{y}) + (y - \hat{y})\right|,$$

from which we infer

$$|(T^{-1} - \text{Id})(y) - (T^{-1} - \text{Id})(\hat{y})| \leq \frac{\|D\theta\|_{L^\infty(\mathbb{R}^N, \mathcal{M}_N(\mathbb{R}))}}{1 - \|D\theta\|_{L^\infty(\mathbb{R}^N, \mathcal{M}_N(\mathbb{R}))}} |y - \hat{y}|.$$

From Lemma 5.2 this implies that $T^{-1} - \text{Id} \in W^{1,\infty}(\mathbb{R}^N, \mathbb{R}^N)$. □

We denote by $\text{Hom}(\mathbb{R}^N, \mathbb{R}^N)$ the set of homeomorphisms from $\mathbb{R}^N$ into itself, i.e. the set of continuous bijective maps with continuous inverse. Let

$$\mathcal{T}_N = \left\{T \in \text{Hom}(\mathbb{R}^N, \mathbb{R}^N) : T - \text{Id} \in W^{1,\infty}(\mathbb{R}^N, \mathbb{R}^N), T^{-1} - \text{Id} \in W^{1,\infty}(\mathbb{R}^N, \mathbb{R}^N)\right\}$$

and, given an open set $\Omega_0 \subset \mathbb{R}^N$,

$$\mathcal{A}(\Omega_0) = \{T(\Omega_0), T \in \mathcal{T}_N\}.$$

This will be our set of admissible shapes for the subsequent analysis. Denote by $B_{1,\infty}(\mathbb{R}^N, \mathbb{R}^N)$ the open unit ball of $W^{1,\infty}(\mathbb{R}^N, \mathbb{R}^N)$. By Proposition 5.4 and Lemma 5.1 we have

$$\theta \in B_{1,\infty}(\mathbb{R}^N, \mathbb{R}^N) \Rightarrow \text{Id} + \theta \in \mathcal{T}_N. \tag{5.1}$$

Remark 5.5 *We have obtained in the proof of Proposition 5.4 the following bounds valid for every* $\theta \in B_{1,\infty}(\mathbb{R}^N, \mathbb{R}^N)$

$$\|(\mathrm{Id} + \theta)^{-1} - \mathrm{Id}\,\|_{L^\infty(\mathbb{R}^N,\mathbb{R}^N)} \leq \|\theta\|_{L^\infty(\mathbb{R}^N,\mathbb{R}^N)},$$

$$\|D((\mathrm{Id} + \theta)^{-1} - \mathrm{Id})\|_{L^\infty(\mathbb{R}^N,\mathbb{R}^N)} \leq \frac{\|D\theta\|_{L^\infty(\mathbb{R}^N,\mathcal{M}_N(\mathbb{R}))}}{1 - \|D\theta\|_{L^\infty(\mathbb{R}^N,\mathcal{M}_N(\mathbb{R}))}}.$$

This in particular shows the continuity at 0 *of the map* $\theta \in B_{1,\infty}(\mathbb{R}^N, \mathbb{R}^N) \mapsto (\mathrm{Id} + \theta)^{-1} - \mathrm{Id} \in W^{1,\infty}(\mathbb{R}^N, \mathbb{R}^N)$.

5.1.2 Definition of the shape derivative

Consider a shape functional

$$\Omega \in \mathcal{A}(\Omega_0) \mapsto \mathcal{J}(\Omega) \in \mathbb{R}.$$

Definition 5.6 *We say that* $\mathcal{J}$ *admits a shape derivative at* Ω_0 *if the map*

$$\theta \in B_{1,\infty}(\mathbb{R}^N, \mathbb{R}^N) \mapsto j(\theta) := \mathcal{J}((\mathrm{Id} + \theta)(\Omega_0))$$

is Fréchet differentiable at 0*. The shape derivative is the map*

$$\tilde{\theta} \in W^{1,\infty}(\mathbb{R}^N, \mathbb{R}^N) \mapsto d_S\mathcal{J}(\Omega_0, \tilde{\theta}) := dj(0)\tilde{\theta}.$$

5.2 Calculus in $W^{1,\infty}(\mathbb{R}^N, \mathbb{R}^N)$

The framework of Lipschitz deformations is natural in that it is the most general one that allows to perform rigorous derivations. Nevertheless it brings some technical complications compared with the restriction to smooth deformation fields. In the first reading the present section may be skipped, and deformations in $\mathcal{C}^1_b(\mathbb{R}^N, \mathbb{R}^N)$ may be considered.

5.2.1 Change of variables in integrals

We will use the following change of variables formula. A proof, in an even more general setting consequence of the area formula of geometric measure theory, can be found in [13].

Theorem 5.7 *Let* $T : \mathbb{R}^N \to \mathbb{R}^N$ *be Lipschitz continuous. For a.e.* $y \in \mathbb{R}^N$ *the set* $T^{-1}(\{y\})$ *is at most countable, and it holds for all function* $g \in L^1(\mathbb{R}^N)$

$$\int_{\mathbb{R}^N} g(x)|\det DT(x)|dx = \int_{\mathbb{R}^N} \left(\sum_{x \in T^{-1}(\{y\})} g(x) \right) dy.$$

Corollary 5.8 *Let* $T : \mathbb{R}^N \to \mathbb{R}^N$ *be Lipschitz continuous. Then for every measurable set* $A \subset \mathbb{R}^N$*it holds*

$$|A| = 0 \Rightarrow |T(A)| = 0.$$

Proof. It stems from

$$\int_{\mathbb{R}^N} \left(\sum_{x \in T^{-1}(\{y\})} \chi_A(x) \right) dy = \int_A |\det DT(x)|dx,$$

valid as soon as $|A| < +\infty$, where we used Theorem 5.7 with $g = \chi_A$, and

$$\chi_{T(A)}(y) \leq \sum_{x \in T^{-1}(\{y\})} \chi_A(x).$$

□

Corollary 5.8 is useful to give a meaning to composite functions of form $f \circ T$, where f is only defined almost everywhere: it is then required that $|T^{-1}(A)| = 0$ whenever $|A| = 0$. It will be achieved when $T \in \mathcal{T}_N$.

This precaution being taken, we will merely use the following corollary, which is a Lipschitz extension of the classical change of variables formula.

Corollary 5.9 *Let Ω be an open subset of $\mathbb{R}^N$ and $T \in \mathcal{T}_N$. Then $f \in L^1(T(\Omega))$ if and only if $f \circ T \in L^1(\Omega)$, in which case*

$$\int_\Omega f(T(x))|\det DT(x)|dx = \int_{T(\Omega)} f(y)dy,$$

$$\int_\Omega f(T(x))dx = \int_{T(\Omega)} f(y)|\det D(T^{-1})(y)|dy.$$

Proof. Applying Theorem 5.7 first to $g = |f \circ T|\chi_\Omega$ yields

$$\int_\Omega |f(T(x))||\det DT(x)|dx = \int_{T(\Omega)} |f(y)|dy.$$

This proves the implication $f \circ T \in L^1(\Omega) \Rightarrow f \in L^1(T(\Omega))$. Conversely,

$$f \in L^1(T(\Omega)) \Rightarrow (f \circ T) \circ T^{-1} \in L^1(T(\Omega)) \Rightarrow f \circ T \in L^1(T^{-1}(T(\Omega))) = L^1(\Omega).$$

Applying now Theorem 5.7 to $g = (f \circ T)\chi_\Omega$ yields the first formula. Applying Theorem 5.7 to $g = f\chi_{T(\Omega)}$ and with T^{-1} substituted for T yields

$$\int_{T(\Omega)} f(x)|\det D(T^{-1})(x)|)dx = \int_\Omega f(T(y))dy,$$

which is the second formula. □

5.2.2 Chain rule

Lemma 5.10 *1) Let $f \in \mathcal{C}^1(\mathbb{R}^N)$ with $\nabla f \in L^\infty(\mathbb{R}^N)^N$, and $T \in W^{1,1}_{\text{loc}}(\mathbb{R}^N, \mathbb{R}^N)$. Then we have the chain rule*

$$\nabla(f \circ T)(x) = DT(x)^\top \nabla f(T(x)) \qquad a.e.\ x \in \mathbb{R}^N. \tag{5.2}$$

2) If $f \in W^{1,\infty}(\mathbb{R}^N)$ and $T \in \mathcal{T}_N$ then $f \circ T \in W^{1,\infty}(\mathbb{R}^N)$ and (5.2) holds true.

Proof. 1) Let $\phi \in \mathcal{C}^1_c(\mathbb{R}^N, \mathbb{R}^N)$. Let ω be an open and bounded subset of $\mathbb{R}^N$ such that $\operatorname{supp}\phi \subset \omega$. Consider a sequence $T_n \in \mathcal{C}^\infty_c(\mathbb{R}^N, \mathbb{R}^N)$ such that $T_n \to T$ in $W^{1,1}(\omega, \mathbb{R}^N)$ and (see e.g. [8] thm IX.2). Since $f \circ T_n$ is differentiable we have by integration by parts and the standard chain rule

$$-\int_{\mathbb{R}^N} f(T_n(x)) \operatorname{div}\phi(x)dx = \int_{\mathbb{R}^N} DT_n(x)^\top \nabla f(T_n(x)) \cdot \phi(x)dx.$$

Passing to the limit by dominated convergence yields

$$-\int_{\mathbb{R}^N} f(T(x)) \operatorname{div}\phi(x)dx = \int_{\mathbb{R}^N} DT(x)^\top \nabla f(T(x)) \cdot \phi(x)dx,$$

which gives the desired weak derivative.
2) Let $\phi \in \mathcal{C}^1_c(\mathbb{R}^N, \mathbb{R}^N)$. Let ω be an open and bounded subset of $\mathbb{R}^N$ such that $\operatorname{supp}\phi \subset \omega$. Let $f_n \in \mathcal{C}^\infty_c(\mathbb{R}^N)$ such that $f_n \to f$ in $W^{1,1}(T^{-1}(\omega))$. From 1) we have

$$-\int_{\mathbb{R}^N} f_n(T(x)) \operatorname{div}\phi(x)dx = \int_{\mathbb{R}^N} DT(x)^\top \nabla f_n(T(x))\phi(x)dx.$$

By Corollary 5.9 we infer that $f_n \circ T \to f \circ T$ in $L^1(\omega)$, as well as $\nabla f_n \circ T \to \nabla f \circ T$ in $L^1(\omega)$. This allows to pass to the limit to obtain

$$-\int_{\mathbb{R}^N} f(T(x)) \operatorname{div} \phi(x) dx = \int_{\mathbb{R}^N} DT(x)^\top \nabla f(T(x)) \phi(x) dx.$$

This proves (5.2). It is then immediate that $f \circ T \in W^{1,\infty}(\mathbb{R}^N)$. □

Lemma 5.11 *Let $T_1 \in W^{1,\infty}(\mathbb{R}^N, \mathbb{R}^N)$, $T_2 \in \mathcal{T}_N$. Then $T_1 \circ T_2 \in W^{1,\infty}(\mathbb{R}^N, \mathbb{R}^N)$ and we have*

$$D(T_1 \circ T_2) = DT_1 \circ T_2 \ DT_2. \tag{5.3}$$

If $T \in \mathcal{T}_N$ then DT is a.e. invertible and we have

$$D(T^{-1}) \circ T = (DT)^{-1}. \tag{5.4}$$

Proof. The first assertion is only the vector-valued extension of Lemma 5.10. To obtain (5.4) we simply differentiate the relation

$$T^{-1} \circ T = \operatorname{Id}.$$

□

5.3 Flow and structure of shape derivatives

5.3.1 Displacement field vs flow

Given $\theta \in W^{1,\infty}(\mathbb{R}^N, \mathbb{R}^N)$ and $x \in \mathbb{R}^N$, let $t \in \mathbb{R} \mapsto \Phi_\theta(t, x)$ be the solution of the ODE

$$\frac{\partial \Phi_\theta}{\partial t}(t, x) = \theta(\Phi_\theta(t, x)), \qquad \Phi_\theta(0, x) = x.$$

By the global Cauchy-Lipschitz theorem, we know that this ODE admits a unique solution $\Phi_\theta(\cdot, x) \in \mathcal{C}^1(\mathbb{R}, \mathbb{R}^N)$. For further analysis we need the following technical lemma.

Lemma 5.12 *Suppose that $\theta \in W^{2,\infty}(\mathbb{R}^N, \mathbb{R}^N)$. There exists $t_0 > 0$ such that*

- *$\Phi_\theta(t, \cdot) - \operatorname{Id} \in B_{1,\infty}(\mathbb{R}^N, \mathbb{R}^N)$ for all $t \in]-t_0, t_0[$,*
- *the map $t \in]-t_0, t_0[\mapsto \Phi_\theta(t, \cdot) - \operatorname{Id} \in W^{1,\infty}(\mathbb{R}^N, \mathbb{R}^N)$ is of class $\mathcal{C}^1$.*

Proof. Set $\Psi_\theta(t, x) = \Phi_\theta(t, x) - x$. We consider the ODE in the space $W^{1,\infty}(\mathbb{R}^N, \mathbb{R}^N)$:

$$\frac{\partial \Psi_\theta}{\partial t}(t, \cdot) = \theta \circ (\operatorname{Id} + \Psi_\theta(t, \cdot)), \qquad \Psi_\theta(0, \cdot) = 0.$$

To apply the local Cauchy-Lipschitz theorem we check that the map

$$T \in B_{1,\infty}(\mathbb{R}^N, \mathbb{R}^N) \mapsto \theta \circ (\operatorname{Id} + T) \in W^{1,\infty}(\mathbb{R}^N, \mathbb{R}^N)$$

is Lipschitz. This is easily achieved using Lemma 5.1 and Lemma 5.11, since we have for every $T_1, T_2 \in B_{1,\infty}(\mathbb{R}^N, \mathbb{R}^N)$

$$\|\theta \circ (\operatorname{Id} + T_1) - \theta \circ (\operatorname{Id} + T_2)\|_{L^\infty(\mathbb{R}^N)} \leq \|D\theta\|_{L^\infty(\mathbb{R}^N, \mathcal{M}_n(\mathbb{R}))} \|T_1 - T_2\|_{L^\infty(\mathbb{R}^N)},$$

and

$$\begin{aligned}
&\|D(\theta \circ (\operatorname{Id} + T_1) - \theta \circ (\operatorname{Id} + T_2))\|_{L^\infty(\mathbb{R}^N)} \\
&= \|D\theta \circ (\operatorname{Id} + T_1)(I + DT_1) - D\theta \circ (\operatorname{Id} + T_2)(I + DT_2)\| \\
&\leq \|D\theta \circ (\operatorname{Id} + T_1) - D\theta \circ (\operatorname{Id} + T_2)\|_{L^\infty(\mathbb{R}^N, \mathcal{M}_N(\mathbb{R}))} \|I + DT_1\|_{L^\infty(\mathbb{R}^N, \mathcal{M}_N(\mathbb{R}))} \\
&+ \|D\theta \circ (\operatorname{Id} + T_2)\|_{L^\infty(\mathbb{R}^N, \mathcal{M}_N(\mathbb{R}))} \|DT_1 - DT_2\|_{L^\infty(\mathbb{R}^N, \mathcal{M}_N(\mathbb{R}))} \\
&\leq \|D^2\theta\|_{L^\infty(\mathbb{R}^N, \mathcal{L}(\mathcal{M}_N(\mathbb{R})))} \|T_1 - T_2\|_{L^\infty(\mathbb{R}^N, \mathcal{M}_N(\mathbb{R}))} \|I + DT_1\|_{L^\infty(\mathbb{R}^N, \mathcal{M}_N(\mathbb{R}))} \\
&+ \|D\theta\|_{L^\infty(\mathbb{R}^N, \mathcal{M}_N(\mathbb{R}))} \|DT_1 - DT_2\|_{L^\infty(\mathbb{R}^N, \mathcal{M}_N(\mathbb{R}))}.
\end{aligned}$$

We conclude the existence of a solution $t \in]-t_0, t_0[\mapsto \Psi_\theta(t, \cdot) \in B_{1,\infty}(\mathbb{R}^N, \mathbb{R}^N)$ of class $\mathcal{C}^1$. □

Proposition 5.13 *If $\mathcal{J}$ admits a shape derivative at Ω_0 then*

$$d_S\mathcal{J}(\Omega_0,\theta) = \frac{d}{dt}\left[\mathcal{J}(\Phi_\theta(t,\Omega_0))\right]_{|t=0} \qquad \forall\theta \in W^{2,\infty}(\mathbb{R}^N,\mathbb{R}^N). \tag{5.5}$$

Proof. Set

$$\Theta(t) = \Phi_\theta(t,\cdot) - \mathrm{Id}\,.$$

By Lemma 5.12 we know that $\Theta(t) \in B_{1,\infty}(\mathbb{R}^N,\mathbb{R}^N)$ for $|t|$ small enough, with Θ is of class $\mathcal{C}^1$, and that, together with (5.1), $\mathrm{Id}+\Theta(t) \in \mathcal{T}_N$ for such t. We have with the notation of Definition 5.6

$$\mathcal{J}(\Phi_\theta(t,\Omega_0)) = \mathcal{J}((\mathrm{Id}+\Theta(t))(\Omega_0)) = j(\Theta(t)).$$

The chain rule yields

$$\frac{d}{dt}\left[\mathcal{J}(\Phi_\theta(t,\Omega_0))\right]_{|t=0} = dj(\Theta(0))\Theta'(0) = dj(0)\theta = d_S\mathcal{J}(\Omega_0,\theta).$$

□

Remark 5.14 *The right hand side of* (5.5) *is used to define the shape derivative in the framework of the speed method. In this approach it is also possible to consider time-dependent velocity fields, i.e. a non-autonomous ODE.*

5.3.2 Structure of shape derivatives

Theorem 5.15 *Suppose that Ω_0 is bounded and of class $\mathcal{C}^1$, and that $\mathcal{J}$ admits a shape derivative at Ω_0. If $\theta_1,\theta_2 \in W^{1,\infty}(\mathbb{R}^N,\mathbb{R}^N)$ are such that $\theta_2-\theta_1 \in W^{2,\infty}(\mathbb{R}^N,\mathbb{R}^N)$ and $\theta_1\cdot n = \theta_2\cdot n$ on $\partial\Omega_0$ then*

$$d_S\mathcal{J}(\Omega_0,\theta_1) = d_S\mathcal{J}(\Omega_0,\theta_2).$$

We first establish a preliminary result related to the flow approach.

Lemma 5.16 *Suppose that Ω_0 is bounded and of class $\mathcal{C}^1$. Let $\theta \in W^{1,\infty}(\mathbb{R}^N,\mathbb{R}^N)$ be such that $\theta\cdot n = 0$ on $\partial\Omega_0$. Then*

$$\Phi_\theta(t,\Omega_0) = \Omega_0 \qquad \forall t \in \mathbb{R}.$$

Proof. Consider first some $x \in \partial\Omega_0$. We locally represent $\partial\Omega_0$ as $\partial\Omega_0 = \{\psi = 0\}$, with $\psi : \mathbb{R}^N \to \mathbb{R}$ of class $\mathcal{C}^1$. A unit normal to $\partial\Omega_0$ is given by the vector $n = \nabla\psi/|\nabla\psi|$, defined in a neighborhood of x. Let $\tilde\theta = \theta - (\theta\cdot n)n$. We have for all t in some interval $]-\varepsilon,\varepsilon[$

$$\frac{d}{dt}\left[\psi(\Phi_{\tilde\theta}(t,x))\right] = \nabla\psi(\Phi_{\tilde\theta}(t,x))\cdot\tilde\theta(\Phi_{\tilde\theta}(t,x)) = 0.$$

Since $\psi(\Phi_{\tilde\theta}(0,x)) = \psi(x) = 0$, we infer that $\psi(\Phi_{\tilde\theta}(t,x)) = \psi(x) = 0$ for all $t \in]-\varepsilon,\varepsilon[$. This means that $\Phi_{\tilde\theta}(t,x) \in \partial\Omega_0$ for all $t \in]-\varepsilon,\varepsilon[$. Subsequently, as $\tilde\theta = \theta$ on $\partial\Omega_0$, $\Phi_\theta(t,x) \in \partial\Omega_0$ for all $t \in]-\varepsilon,\varepsilon[$.

Let $x_0 \in \partial\Omega_0$. By continuity we know that $S := \Phi_\theta(\cdot,x_0)^{-1}(\partial\Omega_0) = \{t \in \mathbb{R} : \Phi_\theta(t,x_0) \in \partial\Omega_0\}$ is a closed subset of $\mathbb{R}$ containing 0. The previous argument applied at $x = \Phi_\theta(t,x_0)$ for any $t \in S$ shows that S is also open. We conclude that $S = \mathbb{R}$, i.e. $\Phi_\theta(t,x_0) \in \partial\Omega_0$ for all $t \in \mathbb{R}$.

Let now $x \in \Omega_0$, $t > 0$. From what precedes the trajectory $\{\Phi_\theta(s,x), 0 \le s \le t\}$ cannot intersect $\partial\Omega_0$. By connectedness[1] we infer that $\{\Phi_\theta(s,x), 0 \le s \le t\} \subset \Omega_0$, in particular $\Phi_\theta(t,x) \in \Omega_0$. Of course the same arguments hold for $t < 0$, hence $\Phi_\theta(t,\Omega_0) \subset \Omega_0$ for all $t \in \mathbb{R}$. If now $x \in \Omega_0$, writing $\Phi_\theta(t,\Phi_\theta(-t,x)) = x$ shows that $x \in \Phi_\theta(t,\Omega_0)$. □

Proof of Theorem 5.15. Set $\theta = \theta_2 - \theta_1$. By Lemma 5.16 we know that $\Phi_\theta(t,\Omega_0) = \Omega_0$ forall $t \in \mathbb{R}$. This yields

$$\mathcal{J}(\Phi_\theta(t,\Omega_0)) = \mathcal{J}(\Omega_0) \qquad \forall t \in \mathbb{R}.$$

In view of Proposition 5.13 we derive that $d_S\mathcal{J}(\Omega_0,\theta) = 0$. The claim follows by linearity of the Fréchet derivative. □

[1] "théorème du passage à la douane": a connected set that meets a set C and its complementary set meets ∂C. Indeed, if $A\cap\partial C = \emptyset$ then $A = (A\cap\mathrm{int}C)\cup(A\cap\mathrm{ext}C)$, which is only possible for A connected if $A\cap\mathrm{int}C = \emptyset$ or $A\cap\mathrm{ext}C = \emptyset$.

5.4 Shape derivative of integral functionals

5.4.1 Shape derivative of volume integrals

Lemma 5.17 *The map*

$$\Phi : \Theta \in L^\infty(\mathbb{R}^N, \mathcal{M}_N(\mathbb{R})) \mapsto \det(I + \Theta) \in L^\infty(\mathbb{R}^N)$$

is differentiable at 0 *with derivative*

$$d\Phi(0)\tilde{\Theta} = \operatorname{tr} \tilde{\Theta}.$$

PROOF. We proceed as in the proof of Proposition 1.10: the determinant is N-linear with respect to the columns, which provides the differentiability and the formula. □

Lemma 5.18 *Let $f \in W^{1,p}(\mathbb{R}^N)$, $1 \le p < +\infty$. The map*

$$\Phi : \theta \in B_{1,\infty}(\mathbb{R}^N, \mathbb{R}^N) \mapsto f \circ (\operatorname{Id} + \theta) \in L^p(\mathbb{R}^N)$$

is well-defined and differentiable at 0 *with*

$$d\Phi(0)\tilde{\theta} : x \mapsto \nabla f(x) \cdot \tilde{\theta}(x).$$

PROOF. We present the proof for $p = 1$ and leave the adaptation to the general case to the reader.
Step 1. If $\theta \in B_{1,\infty}(\mathbb{R}^N, \mathbb{R}^N)$ we know from (5.1) that $\operatorname{Id} + \theta \in \mathcal{T}_N$, hence $f \circ (\operatorname{Id} + \theta)$ is well-defined by Corollary 5.8, and belongs to $L^1(\mathbb{R}^N)$ by Corollary 5.9.
Step 2. We establish a preliminary estimate. Let $\theta \in B_{1,\infty}(\mathbb{R}^N, \mathbb{R}^N)$ and $g \in \mathcal{C}_c^1(\mathbb{R}^N)$. We have

$$\begin{aligned}
\int_{\mathbb{R}^N} |g(x + \theta(x)) - g(x)| dx &= \int_{\mathbb{R}^N} \left| \int_0^1 \nabla g(x + t\theta(x)) \cdot \theta(x) dt \right| dx \\
&\le \int_0^1 \int_{\mathbb{R}^N} |\nabla g(x + t\theta(x))| |\theta(x)| dx dt \\
&\le \|\theta\|_{L^\infty(\mathbb{R}^N)} \int_0^1 \int_{\mathbb{R}^N} |\nabla g(x + t\theta(x))| dx dt.
\end{aligned}$$

By the change of variable $y = (\operatorname{Id} + t\theta)(x)$ within Corollary 5.9 and by continuity at 0 of the map $\theta \in W^{1,\infty}(\mathbb{R}^N, \mathbb{R}^N) \mapsto \det(D(\operatorname{Id} + \theta)^{-1}) \in L^\infty(\mathbb{R}^N)$, consequence of Lemma 5.17 and Remark 5.5, we infer that

$$\int_{\mathbb{R}^N} |g(x + \theta(x)) - g(x)| dx \le 2\|\theta\|_{L^\infty(\mathbb{R}^N)} \|\nabla g\|_{L^1(\mathbb{R}^N)},$$

as soon as $\|\theta\|_{W^{1,\infty}(\mathbb{R}^N, \mathbb{R}^N)} \le M$ for some constant M independent of g. The inequality then extends by density to all $g \in W^{1,1}(\mathbb{R}^N)$.
Step 3. Set

$$E(\theta) = \int_{\mathbb{R}^N} |f(x + \theta(x)) - f(x) - \nabla f(x) \cdot \theta(x)| \, dx.$$

a) Suppose first that $f \in \mathcal{C}_c^2(\mathbb{R}^N)$. We have for every $x \in \mathbb{R}^N$

$$f(x + \theta(x)) - f(x) = \int_0^1 \nabla f(x + t\theta(x)) \cdot \theta(x) dt,$$

hence

$$f(x + \theta(x)) - f(x) - \nabla f(x) \cdot \theta(x) = \int_0^1 (\nabla f(x + t\theta(x)) - \nabla f(x)) \cdot \theta(x) dt.$$

It follows that

$$\begin{aligned} E(\theta) &\leq \int_0^1 \int_{\mathbb{R}^N} |\nabla f(x+t\theta(x)) - \nabla f(x)|\, |\theta(x)| dx dt \\ &\leq \|\theta\|_{L^\infty(\mathbb{R}^N)} \int_{\mathbb{R}^N} R(\theta)(x) dx, \qquad R(\theta)(x) = \int_0^1 |\nabla f(x+t\theta(x)) - \nabla f(x)|\, dt. \end{aligned}$$

As ∇f is compactly supported and Lipschitz we have for some constant $C(f) > 0$

$$\|R(\theta)\|_{L^1(\mathbb{R}^N)} \leq C(f) \|\theta\|_{L^\infty(\mathbb{R}^N)}, \qquad E(\theta) \leq C(f) \|\theta\|^2_{L^\infty(\mathbb{R}^N)}.$$

b) Suppose now that $f \in W^{1,1}(\mathbb{R}^N)$. Let $\varepsilon > 0$ and $f_\varepsilon \in \mathcal{C}^2_c(\mathbb{R}^N)$ such that $\|f_\varepsilon - f\|_{W^{1,1}(\mathbb{R}^N)} \leq \varepsilon$. We have

$$\begin{aligned} E(\theta) \leq \int_{\mathbb{R}^N} |(f - f_\varepsilon)(x + \theta(x)) + (f - f_\varepsilon)(x)|\, dx + \int_{\mathbb{R}^N} |\nabla (f - f_\varepsilon)(x)| |\theta(x)| dx \\ + \int_{\mathbb{R}^N} |f_\varepsilon(x + \theta(x)) - f_\varepsilon(x) - \nabla f_\varepsilon(x) \cdot \theta(x)|\, dx. \end{aligned}$$

Suppose that $\|\theta\|_{L^\infty(\mathbb{R}^N)} \leq \varepsilon C(f_\varepsilon)^{-1}$. We infer from a) that

$$E(\theta) \leq \int_{\mathbb{R}^N} |(f - f_\varepsilon)(x + \theta(x)) + (f - f_\varepsilon)(x)|\, dx + 2\varepsilon \|\theta\|_{L^\infty(\mathbb{R}^N)}.$$

Using now step 2 with $g = f - f_\varepsilon$ we arrive at

$$E(\theta) \leq 4\varepsilon \|\theta\|_{L^\infty(\mathbb{R}^N)}$$

as soon as $\|\theta\|_{W^{1,\infty}(\mathbb{R}^N, \mathbb{R}^N)}$ is small enough. □

We will also need the following variant.

Lemma 5.19 *Let $f \in W^{1,p}(\mathbb{R}^N)$, $1 \leq p < +\infty$. The map*

$$\Phi : \theta \in B_{1,\infty}(\mathbb{R}^N, \mathbb{R}^N) \mapsto f \circ (\mathrm{Id} + \theta)^{-1} \in L^p(\mathbb{R}^N)$$

is well-defined and differentiable at 0 with

$$d\Phi(0)\tilde{\theta} : x \mapsto -\nabla f(x) \cdot \tilde{\theta}(x).$$

Proof. Here also we display the proof for $p = 1$. The function is well-defined by the same arguments as in Lemma 5.18. Define

$$E(\theta) = \int_{\mathbb{R}^N} \left| f \circ (\mathrm{Id} + \theta)^{-1} - f + \nabla f \cdot \theta \right| dx.$$

By corollary 5.9 we have

$$E(\theta) = \int_{\mathbb{R}^N} |f - f \circ (\mathrm{Id} + \theta) + \nabla f \circ (\mathrm{Id} + \theta) \cdot \theta \circ (\mathrm{Id} + \theta)|\, |\det D(\mathrm{Id} + \theta)| dx.$$

From Lemma 5.17 and Remark 5.5 we infer that for $\|\theta\|_{W^{1,\infty}(\mathbb{R}^N, \mathbb{R}^N)}$ small enough

$$E(\theta) \leq 2 \int_{\mathbb{R}^N} |f - f \circ (\mathrm{Id} + \theta) + \nabla f \circ (\mathrm{Id} + \theta) \cdot \theta \circ (\mathrm{Id} + \theta)|\, dx.$$

We decompose as

$$E(\theta) \leq 2 \left(E_1(\theta) + E_2(\theta) + E_3(\theta) \right)$$

with

$$E_1(\theta) = \int_{\mathbb{R}^N} |f - f \circ (\mathrm{Id} + \theta) + \nabla f \cdot \theta| \, dx,$$

$$E_2(\theta) = \int_{\mathbb{R}^N} |(\nabla f \circ (\mathrm{Id} + \theta) - \nabla f) \cdot \theta \circ (\mathrm{Id} + \theta)| \, dx,$$

$$E_3(\theta) = \int_{\mathbb{R}^N} |\nabla f \cdot (\theta \circ (\mathrm{Id} + \theta) - \theta)| \, dx.$$

Lemma 5.18 yields $E_1(\theta) = o(\|\theta\|_{W^{1,\infty}(\mathbb{R}^N,\mathbb{R}^N)})$. Then we use

$$E_2(\theta) \leq \|\theta\|_{W^{1,\infty}(\mathbb{R}^N,\mathbb{R}^N)} \int_{\mathbb{R}^N} |\nabla f \circ (\mathrm{Id} + \theta) - \nabla f| \, dx.$$

Let $\varepsilon > 0$ and $f_\varepsilon \in \mathcal{C}_c^2(\mathbb{R}^N)$ such that $\|f - f_\varepsilon\|_{W^{1,1}(\mathbb{R}^N)} \leq \varepsilon$. Then

$$E_2(\theta) \leq \|\theta\|_{W^{1,\infty}(\mathbb{R}^N,\mathbb{R}^N)} \bigg(\int_{\mathbb{R}^N} |\nabla f_\varepsilon \circ (\mathrm{Id} + \theta) - \nabla f_\varepsilon| \, dx + \int_{\mathbb{R}^N} |\nabla (f - f_\varepsilon) \circ (\mathrm{Id} + \theta)| \, dx + \int_{\mathbb{R}^N} |\nabla (f - f_\varepsilon)| \, dx \bigg).$$

Since ∇f_ε is Lipschitz, choosing θ small enough permits to have the first integral bounded by ε. By change of variables the second integral can be bounded by 2ε. We arrive at

$$E_2(\theta) \leq 4\varepsilon \|\theta\|_{W^{1,\infty}(\mathbb{R}^N,\mathbb{R}^N)},$$

meaning that $E_2(\theta) = o(\|\theta\|_{W^{1,\infty}(\mathbb{R}^N,\mathbb{R}^N)})$. Lastly, using

$$|\theta(x + \theta(x)) - \theta(x)| \leq \|\theta\|_{W^{1,\infty}(\mathbb{R}^N,\mathbb{R}^N)} |\theta(x)| \leq \|\theta\|^2_{W^{1,\infty}(\mathbb{R}^N,\mathbb{R}^N)}$$

we get $E_3(\theta) = O(\|\theta\|^2_{W^{1,\infty}(\mathbb{R}^N,\mathbb{R}^N)})$. □

Lemma 5.20 *If $f \in W^{1,p}(\mathbb{R}^N)$, $1 \leq p \leq +\infty$, and $\theta \in W^{1,\infty}(\mathbb{R}^N, \mathbb{R}^N)$ then*

$$\mathrm{div}(f\theta) = \nabla f \cdot \theta + f \, \mathrm{div} \, \theta \in L^p(\mathbb{R}^N).$$

Proof. Let $\varphi, \psi \in \mathcal{C}_c^1(\mathbb{R}^N)$. We have

$$-\int_{\mathbb{R}^N} \theta\varphi \cdot \nabla\psi dx = \int_{\mathbb{R}^N} \theta \cdot (-\nabla(\varphi\psi) + \nabla\varphi\psi) \, dx = \int_{\mathbb{R}^N} \mathrm{div} \, \theta\varphi\psi dx + \int_{\mathbb{R}^N} \theta \cdot \nabla\varphi\psi dx,$$

leading to $\mathrm{div}(\theta\varphi) = \mathrm{div} \, \theta\varphi + \theta \cdot \nabla\varphi$. This yields

$$\begin{aligned} -\int_{\mathbb{R}^N} f\theta \cdot \nabla\varphi dx &= \int_{\mathbb{R}^N} f \left(-\mathrm{div}(\theta\varphi) + \mathrm{div} \, \theta\varphi \right) dx \\ &= \int_{\mathbb{R}^N} \nabla f \cdot \theta\varphi dx + \int_{\mathbb{R}^N} f \, \mathrm{div} \, \theta\varphi dx, \end{aligned}$$

whereby we infer that $\mathrm{div}(f\theta) = \nabla f \cdot \theta + f \, \mathrm{div} \, \theta$. □

Theorem 5.21 *Let $f \in W^{1,1}(\mathbb{R}^N)$, Ω_0 be an open subset of $\mathbb{R}^N$ and consider the shape functional*

$$\mathcal{J} : \Omega \in \mathcal{A}(\Omega_0) \mapsto \int_\Omega f(x) dx.$$

Then $\mathcal{J}$ admits a shape derivative at Ω_0 given by

$$d_S\mathcal{J}(\Omega_0, \tilde{\theta}) = \int_{\Omega_0} \mathrm{div}(f\tilde{\theta}) dx.$$

If Ω_0 is bounded and of class $\mathcal{C}^1$ then we have the boundary expression

$$\boxed{d_S\mathcal{J}(\Omega_0, \tilde{\theta}) = \int_{\partial\Omega_0} \gamma_0(f\tilde{\theta}) \cdot n ds.}$$

PROOF. We use Corollary 5.9 and $D(\mathrm{Id}+\theta) = I + D\theta$ to write

$$\mathcal{J}((\mathrm{Id}+\theta)(\Omega_0)) = \int_{\Omega_0} f((\mathrm{Id}+\theta)(x))|\det(I + D\theta(x))|dx.$$

Lemma 5.18 yields

$$f((\mathrm{Id}+\theta)(x)) = f(x) + \nabla f(x)\cdot\theta(x) + R_1(x), \qquad \lim_{\theta\to 0}\frac{\|R_1\|_{L^1(\mathbb{R}^N)}}{\|\theta\|_{W^{1,\infty}(\mathbb{R}^N,\mathbb{R}^N)}} = 0.$$

Lemma 5.17 yields

$$\det(I + D\theta(x)) = 1 + \operatorname{div}\theta(x) + R_2(x), \qquad \lim_{\theta\to 0}\frac{\|R_2\|_{L^\infty(\mathbb{R}^N)}}{\|\theta\|_{W^{1,\infty}(\mathbb{R}^N,\mathbb{R}^N)}} = 0.$$

Combining the two above estimates results in

$$\mathcal{J}((\mathrm{Id}+\theta)(\Omega_0)) = J(\Omega_0) + \int_{\Omega_0}\big(f\operatorname{div}\theta + \nabla f\cdot\theta + \nabla f\cdot\theta\operatorname{div}\theta$$
$$+ R_1(1+\operatorname{div}\theta) + R_2(f + \nabla f\cdot\theta) + R_1R_2\big)dx.$$

Using Lemma 5.20 we obtain

$$\mathcal{J}((\mathrm{Id}+\theta)(\Omega_0)) = \mathcal{J}(\Omega_0) + \int_{\Omega_0}(\operatorname{div}(f\theta) + R)\,dx, \qquad \lim_{\theta\to 0}\frac{\|R\|_{L^1(\mathbb{R}^N)}}{\|\theta\|_{W^{1,\infty}(\mathbb{R}^N,\mathbb{R}^N)}} = 0.$$

This provides the first expression of the shape derivative. The second expression results from the divergence formula (1.3). □

5.4.2 Shape derivative of boundary integrals

We will restrict ourselves here to $\mathcal{C}^1$-diffeomorphisms in order to preserve the smoothness of the boundary. Here also, our analysis will extensively rely on a change of variables formula.

Theorem 5.22 *Let Ω_0 be a bounded open subset of $\mathbb{R}^N$ of class $\mathcal{C}^1$. Let T be a $\mathcal{C}^1$-diffeomorphism of $\mathbb{R}^N$. If $f\in L^1(\partial T(\Omega_0))$ then $f\circ T\in L^1(\partial\Omega_0)$ and we have*

$$\int_{\partial T(\Omega_0)} f ds = \int_{\partial\Omega_0} f\circ T|\det DT|\left|(DT)^{-\top}n\right| ds.$$

PROOF. With the help of local maps and partition of unity we reduce the integral to a finite sum of the form

$$\int_{\partial T(\Omega_0)} f ds = \sum_{i=1}^n \int_{\partial T(\Omega_0)\cap T(\mathcal{O}_i)} f_i ds,$$

where f_i is compactly supported in $T(\mathcal{O}_i.)$ Each of these subsets $\partial T(\Omega_0)\cap T(\mathcal{O}_i)$ admits a $\mathcal{C}^1$ parametric representation of the form $t\in B_i\mapsto\psi(t)\in\partial T(\Omega_0)\cap T(\mathcal{O}_i)$, where B_i is an open subset of $\mathbb{R}^{N-1}$. We will suppose here that $N=3$. The 2D case is left to the reader. A generic proof can be found in [16]. Step 1. We begin with a preliminary algebraic identity. Let $A\in GL_m(\mathbb{R})$, $u, v\in\mathbb{R}^m$. We have for any $w\in\mathbb{R}^m$

$$(Au\wedge Av)\cdot Aw = \det(Au, Av, Aw) = \det A\ \det(u,v,w) = \det A\ (u\wedge v)\cdot w = \det A\ A^{-\top}(u\wedge v)\cdot Aw,$$

which yields

$$Au\wedge Av = \det A\ A^{-\top}(u\wedge v).$$

Step 2. We have by definition of the surface integral

$$\int_{\partial T(\Omega_0)\cap T(\mathcal{O}_i)} f_i ds = \int_{B_i} f_i(\psi(t)) \left|\frac{\partial\psi}{\partial t_1} \wedge \frac{\partial\psi}{\partial t_2}\right| dt_1 dt_2.$$

Now, given a parametric representation $t \in B_i \mapsto \varphi(t)$ of $\partial\Omega_0 \cap \mathcal{O}_i$, a representation of $\partial T(\Omega_0) \cap T(\mathcal{O}_i)$ is achieved by setting $\psi(t) = T(\varphi(t))$. It results in

$$\int_{\partial T(\Omega_0)\cap T(\mathcal{O}_i)} f_i ds = \int_{B_i} f_i \circ T(\varphi(t)) \left|DT(\varphi(t))\frac{\partial\varphi}{\partial t_1} \wedge DT(\varphi(t)\frac{\partial\varphi}{\partial t_2}\right| dt_1 dt_2.$$

By step 1 this rewrites as

$$\int_{\partial T(\Omega_0)\cap T(\mathcal{O}_i)} f_i ds = \int_{B_i} f_i \circ T(\varphi(t))|\det DT(\varphi(t))| \left|DT(\varphi(t))^{-\top}\left(\frac{\partial\varphi}{\partial t_1} \wedge \frac{\partial\varphi}{\partial t_2}\right)\right| dt_1 dt_2.$$

Using the standard expression of the normal

$$n(\varphi(t)) = \frac{\frac{\partial\varphi}{\partial t_1} \wedge \frac{\partial\varphi}{\partial t_2}}{\left|\frac{\partial\varphi}{\partial t_1} \wedge \frac{\partial\varphi}{\partial t_2}\right|}$$

we arrive at

$$\int_{\partial T(\Omega_0)\cap T(\mathcal{O}_i)} f_i ds = \int_{B_i} f_i \circ T(\varphi(t))|\det DT(\varphi(t))| \left|DT(\varphi(t))^{-\top}n(\varphi(t))\right| \left|\frac{\partial\varphi}{\partial t_1} \wedge \frac{\partial\varphi}{\partial t_2}\right| dt_1 dt_2.$$

We recognize the surface integral

$$\int_{\partial T(\Omega_0)\cap T(\mathcal{O}_i)} f_i ds = \int_{\partial\Omega_0\cap\mathcal{O}_i} f_i \circ T|\det DT| \left|DT^{-\top}n\right| ds.$$

The claimed formula is obtained after summation. □

We recall that $\mathcal{C}_b^1(\mathbb{R}^N, \mathbb{R}^N)$ is the set of bounded functions from $\mathcal{C}^1(\mathbb{R}^N, \mathbb{R}^N)$ with bounded first derivatives. It is naturally equipped with the $\mathcal{C}^1$ norm. By Proposition 5.4 and the local inversion theorem, we know that $\mathrm{Id} + \theta$ is a $\mathcal{C}^1$-diffeomorphism of $\mathbb{R}^N$ as soon as $\|\theta\|_{\mathcal{C}^1(\mathbb{R}^N,\mathbb{R}^N)} < 1$.

Theorem 5.23 *Let Ω_0 be a bounded open subset of $\mathbb{R}^N$ of class $\mathcal{C}^1$ and $f \in \mathcal{C}^1(\mathbb{R}^N)$. The map*

$$j : \theta \in \mathcal{C}_b^1(\mathbb{R}^N, \mathbb{R}^N) \mapsto \int_{\partial(\mathrm{Id}+\theta)(\Omega_0)} f ds$$

is differentiable at 0 with

$$dj(0)\tilde{\theta} = \int_{\partial\Omega_0} \left(\nabla f \cdot \tilde{\theta} + f \operatorname{div}\tilde{\theta} - f D\tilde{\theta}n \cdot n\right) ds.$$

If $\partial\Omega_0$ is of class $\mathcal{C}^2$ then we have the alternative expression

$$\boxed{dj(0)\tilde{\theta} = \int_{\partial\Omega_0} \left(\frac{\partial f}{\partial n} + \kappa f\right) \tilde{\theta} \cdot n ds,}$$

where $\kappa = \operatorname{div} n$ is the mean curvature (sum of principal curvatures) of $\partial\Omega_0$.

Proof. Step 1. We have by Theorem 5.22

$$j(\theta) = \int_{\partial\Omega_0} f \circ (\mathrm{Id}+\theta)|\det(I + D\theta)| \left|(I + D\theta)^{-\top}n\right| ds.$$

Similarly to the proof of Theorem 5.21 we show that

$$f \circ (\operatorname{Id} + \theta) = f + \nabla f \cdot \theta + R_1, \qquad \lim_{\theta \to 0} \frac{\|R_1\|_{L^1(\partial\Omega_0)}}{\|\theta\|_{\mathcal{C}^1(\mathbb{R}^N,\mathbb{R}^N)}} = 0,$$

$$\det(I + D\theta)| = 1 + \operatorname{div}\theta + R_2, \qquad \lim_{\theta \to 0} \frac{\|R_2\|_{L^\infty(\partial\Omega_0)}}{\|\theta\|_{\mathcal{C}^1(\mathbb{R}^N,\mathbb{R}^N)}} = 0.$$

Similarly to Lemma 1.9 we have

$$(I + D\theta)^{-\top} n = n - D\theta^\top n + R_3, \qquad \lim_{\theta \to 0} \frac{\|R_3\|_{L^\infty(\partial\Omega_0,\mathbb{R}^N)}}{\|\theta\|_{\mathcal{C}^1(\mathbb{R}^N,\mathbb{R}^N)}} = 0,$$

leading to

$$|(I + D\theta)^{-\top} n| = 1 - D\theta n \cdot n + \tilde{R}_3, \qquad \lim_{\theta \to 0} \frac{\|\tilde{R}_3\|_{L^\infty(\partial\Omega_0)}}{\|\theta\|_{\mathcal{C}^1(\mathbb{R}^N,\mathbb{R}^N)}} = 0.$$

This yields

$$j(\theta) = j(0) + \int_{\partial\Omega_0} (\nabla f \cdot \theta + f \operatorname{div}\theta - f D\theta n \cdot n)\, ds + o(\|\theta\|_{\mathcal{C}^1(\mathbb{R}^N,\mathbb{R}^N)}),$$

from which we infer the first expression of the shape derivative.
Step 2. Choosing $f = 1$ in the previously obtained formula and θ such that $\theta \cdot n = 0$ (first in $\mathcal{C}^2_b(\mathbb{R}^N, \mathbb{R}^N)$ then in $\mathcal{C}^1(\mathbb{R}^N, \mathbb{R}^N)$ by continuity and density) yields by Theorem 5.15

$$\int_{\partial\Omega_0} (\operatorname{div}\theta - D\theta n \cdot n)\, ds = 0.$$

We define the tangential divergence as

$$\operatorname{div}_{\partial\Omega_0} \theta = \operatorname{div}\theta - D\theta n \cdot n.$$

We infer from this definition that for any scalar function $g \in \mathcal{C}^1(\mathbb{R}^N)$

$$\operatorname{div}_{\partial\Omega_0}(gn) = \nabla g \cdot n + g \operatorname{div} n - \nabla g \cdot n - \frac{1}{2} g \nabla(n \cdot n) \cdot n = \kappa g.$$

We conclude that for every $\theta \in \mathcal{C}^1(\mathbb{R}^N, \mathbb{R}^N)$

$$\int_{\partial\Omega_0} \operatorname{div}_{\partial\Omega_0} \theta ds = \int_{\partial\Omega_0} \operatorname{div}_{\partial\Omega_0}(\theta \cdot nn) ds = \int_{\partial\Omega_0} \kappa\theta \cdot n ds. \tag{5.6}$$

A direct proof of this statement can be found in [16].
Step 3. We now reformulate the shape derivative of step 1 as

$$\begin{aligned} dj(0)\theta &= \int_{\partial\Omega_0} (\operatorname{div}(f\theta) - f D\theta n \cdot n)\, ds \\ &= \int_{\partial\Omega_0} \left(\operatorname{div}_{\partial\Omega_0}(f\theta) + \frac{\partial f}{\partial n}\theta \cdot n\right) ds \\ &= \int_{\partial\Omega_0} \left(\kappa f\theta \cdot n + \frac{\partial f}{\partial n}\theta \cdot n\right) ds, \end{aligned}$$

using step 2 for this latter equality. □

As a particular case of Theorem 5.23 we see that under suitable regularity assumptions the shape derivative of the perimeter is the mean curvature. This provides a necessary optimality to the minimal surface problem described in subsection 2.2.4: minimal surfaces have vanishing mean curvature at every point around which they are of class $\mathcal{C}^2$.

5.5 Shape derivative for elliptic boundary value problems

5.5.1 Transport of a boundary value problem: general construction

We consider the following framework.

Let Ω_0 be a bounded open subset of $\mathbb{R}^N$. Given some $\theta \in B_{1,\infty}(\mathbb{R}^N, \mathbb{R}^N)$ we consider a boundary value problem over the domain $(\mathrm{Id} + \theta)(\Omega_0) \in \mathcal{A}(\Omega_0)$ of the form

$$\begin{array}{l} u(\theta) \in H(\theta) \\ a(\theta, u(\theta), \varphi) = l(\theta, \varphi) \qquad \forall \varphi \in H(\theta), \end{array}$$

where $H(\theta)$ is a Hilbert space, $a(\theta, \cdot, \cdot)$ is a continuous coercive bilinear form on $H(\theta)$ and $l(\theta, \cdot)$ is a continuous linear form on $H(\theta)$. We aim at differentiating $u(\theta)$, or a function of $u(\theta)$, with respect to θ by means of the techniques described in chapter 4, but we face the difficulty that the space $H(\theta)$ also depends on θ. To overcome this we use a transport technique, by considering as new state variable the function

$$\bar{u}(\theta) = u(\theta) \circ (\mathrm{Id} + \theta) \tag{5.7}$$

defined in Ω_0. In order to handle this transported state we need to make an assumption on the spaces:

$$\forall \theta \in B_{1,\infty}(\mathbb{R}^N, \mathbb{R}^N), \qquad \left\{ \begin{array}{l} f \in H(\theta) \Leftrightarrow f \circ (\mathrm{Id} + \theta) \in H(0) =: H_0, \\ {[f \mapsto f \circ (\mathrm{Id} + \theta)]} \in \mathrm{isom}(H(\theta), H_0). \end{array} \right. \tag{5.8}$$

Proposition 5.24 *Set*

$$\bar{a}(\theta, \bar{\psi}, \bar{\varphi}) = a(\theta, \bar{\psi} \circ (\mathrm{Id} + \theta)^{-1}, \bar{\varphi} \circ (\mathrm{Id} + \theta)^{-1}) \qquad \forall \bar{\psi}, \bar{\varphi} \in H_0,$$

$$\bar{l}(\theta, \bar{\varphi}) = l(\theta, \bar{\varphi} \circ (\mathrm{Id} + \theta)^{-1}) \qquad \forall \bar{\varphi} \in H_0.$$

Under assumption (5.8)*, $\bar{a}(\theta, \cdot, \cdot)$ is a continuous coercive bilinear form on H_0, $\bar{l}(\theta, \cdot)$ is a continuous linear form on H_0, and the function $\bar{u}(\theta)$ defined in* (5.7) *is the unique solution in H_0 of*

$$\bar{a}(\theta, \bar{u}(\theta), \bar{\varphi}) = \bar{l}(\theta, \bar{\varphi}) \qquad \forall \bar{\varphi} \in H_0. \tag{5.9}$$

Proof. Assumption (5.8) already ensures that $\bar{u}(\theta) \in H_0$. We have by construction

$$a(\theta, \bar{u}(\theta) \circ (\mathrm{Id} + \theta)^{-1}, \varphi) = l(\theta, \varphi) \qquad \forall \varphi \in H(\theta).$$

Choosing $\varphi = \bar{\varphi} \circ (\mathrm{Id} + \theta)^{-1}$ yields (5.9).

Using Assumption (5.8), it is clear that $\bar{a}(\theta, \cdot, \cdot)$ is a continuous bilinear form on H_0, and that $\bar{l}(\theta, \cdot)$ is a continuous linear form on H_0. It is also true that $\bar{a}(\theta, \cdot, \cdot)$ is coercive, since

$$\bar{a}(\theta, \bar{\varphi}, \bar{\varphi}) = a(\theta, \bar{\varphi} \circ (\mathrm{Id} + \theta)^{-1}, \bar{\varphi} \circ (\mathrm{Id} + \theta)^{-1}) \geq c_\theta \|\bar{\varphi} \circ (\mathrm{Id} + \theta)^{-1}\|^2_{H(\theta)} \geq c_\theta c'_\theta \|\bar{\varphi}\|^2_{H_0},$$

where c_θ is the coercivity constant of $a(\theta, \cdot, \cdot)$ and c'_θ is the Lipschitz constant of the map $f \mapsto f \circ (\mathrm{Id} + \theta)$. We infer the uniqueness statement by Lax-Milgram. □

Consider now a cost function of the form

$$j(\theta) = J(\theta, u(\theta)) \in \mathbb{R}$$

defined over the set $\{(\theta, u), \theta \in B_{1,\infty}(\mathbb{R}^N, \mathbb{R}^N), u \in H(\theta)\}$. We transport this cost function as

$$j(\theta) = \bar{J}(\theta, \bar{u}(\theta)) \tag{5.10}$$

with $\bar{J} : B_{1,\infty}(\mathbb{R}^N, \mathbb{R}^N) \times H_0 \to \mathbb{R}$ defined by

$$\bar{J}(\theta, \bar{u}) = J(\theta, \bar{u} \circ (\mathrm{Id} + \theta)^{-1}).$$

Under appropriate differentiability hypothesis, we are in position to apply either the direct method or the adjoint method to find out an expression of the derivative $dj(\theta)\hat{\theta}$ through the problem (5.9) - (5.10). As developed in chapter 4, the direct method involves the derivative $d\bar{u}(\theta)\hat{\theta}$, which is the solution of a specific boundary value problem for each $\hat{\theta}$. We will prefer the adjoint method which permits to bypass this difficulty. Nevertheless the derivative $d\bar{u}(\theta)\hat{\theta}$ may be interesting on its own: it is called material derivative.

To conclude this subsection, we check Assumption (5.8) in the prototype cases.

Proposition 5.25 *Let $T \in \mathcal{T}_N$ and $\Omega_T = T(\Omega_0)$. We have*

$$f \in H^1(\Omega_T) \Leftrightarrow f \circ T \in H^1(\Omega_0),$$

$$[f \mapsto f \circ T] \in \mathrm{isom}(H^1(\Omega_T), H^1(\Omega_0)).$$

The above properties also hold true if H^1 is replaced by H^1_0. Moreover we have for all $f \in H^1(\Omega_T)$

$$\nabla(f \circ T) = DT^\top \nabla f \circ T. \tag{5.11}$$

PROOF. Step 1. By Corollary 5.9 applied to $|f|^2$ we have that

$$f \in L^2(\Omega_T) \Leftrightarrow f \circ T \in L^2(\Omega_0).$$

Step 2. We show that the map

$$f \in L^2(\Omega_T) \mapsto f \circ T \in L^2(\Omega_0)$$

is continuous. Suppose that $f_n \to 0$ in $L^2(\Omega_T)$. Applying Corollary 5.9 to $|f_n|^2$ shows that $\|f_n \circ T\|_{L^2(\Omega_0)} \to 0$. This proves continuity due to linearity.
Step 3. Suppose that $f \in H^1(\Omega_T)$. We show (5.11). Let $\phi \in \mathcal{C}^1_c(\Omega_T)$. Denote $\omega \subset\subset \Omega_T$ such that $\mathrm{supp}\,\phi \subset \omega$, and $\eta \in \mathcal{C}^\infty_c(\Omega_T)$ such that $\eta = 1$ in ω. Since $\eta f \in H^1_0(\Omega_T)$ there exists $f_n \in \mathcal{C}^\infty_c(\Omega_T)$ such that $f_n \to \eta f$ in $H^1(\Omega_T)$. By Lemma 5.10 we have

$$\int_{\Omega_0} f_n(T(x))\,\mathrm{div}\,\phi(x)dx = -\int_{\Omega_0} DT(x)^\top \nabla f_n(T(x))\phi(x)dx.$$

Passing to the limit using step 2 results in

$$\int_{\Omega_0} f(T(x))\,\mathrm{div}\,\phi(x)dx = -\int_{\Omega_0} DT(x)^\top \nabla f(T(x))\phi(x)dx.$$

We infer (5.11).
Step 4. From (5.11) and step 1 we infer that $f \in H^1(\Omega_T) \Rightarrow f \circ T \in H^1(\Omega_0)$.
Step 5. We now show that the map

$$f \in H^1(\Omega_T) \mapsto f \circ T \in H^1(\Omega_0)$$

is continuous. Suppose that $f_n \to 0$ in $H^1(\Omega_T)$. Using (5.11) and step 1 shows that $f_n \circ T \to 0$ in $H^1(\Omega_0)$.
Step 6. Suppose that $f \in H^1_0(\Omega_T)$. By definition there exists $f_n \in \mathcal{C}^\infty_c(\Omega_T)$ such that $f_n \to f$ in $H^1(\Omega_T)$. By step 5, $f_n \circ T \to f \circ T$ in $H^1(\Omega_0)$. Since $f_n \circ T$ is compactly supported in Ω_0 we have that $f_n \circ T \in H^1_0(\Omega_0)$. Therefore $f \circ T \in H^1_0(\Omega_0)$.
Step 7. The map $f \in H^1(\Omega_T) \mapsto f \circ T \in H^1(\Omega_0)$ is an isomorphism because T is bijective from Ω_0 into Ω_T and $T^{-1} \in \mathcal{T}_N$. The same holds for H^1_0. □

5.5.2 Transport of a model bilinear form

We focus our attention on the bilinear form defined for any $\theta \in B_{1,\infty}(\mathbb{R}^N, \mathbb{R}^N)$ by

$$a(\theta, u, v) = \int_{(\mathrm{Id}+\theta)(\Omega_0)} \nabla u \cdot \nabla v dx \qquad \forall u, v \in H^1((\mathrm{Id}+\theta)(\Omega_0)).$$

In view or Proposition 5.24 we set

$$\bar{a}(\theta, \bar{u}, \bar{v}) = a(\theta, \bar{u} \circ (\mathrm{Id}+\theta)^{-1}, \bar{v} \circ (\mathrm{Id}+\theta)^{-1}) \qquad \forall \bar{u}, \bar{v} \in H^1(\Omega_0).$$

Using Proposition 5.25 we obtain

$$\bar{a}(\theta, \bar{u}, \bar{v}) = \int_{(\mathrm{Id}+\theta)(\Omega_0)} (D((\mathrm{Id}+\theta)^{-1}))^\top \nabla \bar{u} \circ (\mathrm{Id}+\theta)^{-1} \cdot (D((\mathrm{Id}+\theta)^{-1}))^\top \nabla \bar{v} \circ (\mathrm{Id}+\theta)^{-1} dx.$$

A change of variables using Corollary 5.9 yields

$$\bar{a}(\theta, \bar{u}, \bar{v}) = \int_{\Omega_0} (D((\mathrm{Id}+\theta)^{-1}))^\top \circ (\mathrm{Id}+\theta) \nabla \bar{u} \cdot (D((\mathrm{Id}+\theta)^{-1}))^\top \circ (\mathrm{Id}+\theta) \nabla \bar{v} |\det D(\mathrm{Id}+\theta)| dx.$$

With the help of Lemma 5.11 we arrive at

$$\bar{a}(\theta, \bar{u}, \bar{v}) = \int_{\Omega_0} (I + D\theta)^{-\top} \nabla \bar{u} \cdot (I + D\theta)^{-\top} \nabla \bar{v} |\det(I + D\theta)| dx.$$

Introducing

$$C(\theta) = |\det(I + D\theta)| (I + D\theta)^{-1} (I + D\theta)^{-\top} \tag{5.12}$$

we have

$$\bar{a}(\theta, \bar{u}, \bar{v}) = \int_{\Omega_0} C(\theta) \nabla \bar{u} \cdot \nabla \bar{v} dx.$$

5.5.3 Transport of model linear forms

1. We consider first the bulk linear form defined for $f \in L^2(\mathbb{R}^N)$ by

$$l_1(\theta, v) = \int_{(\mathrm{Id}+\theta)(\Omega_0)} f v dx \qquad \forall v \in H^1((\mathrm{Id}+\theta)(\Omega_0)).$$

We associate

$$\bar{l}_1(\theta, \bar{v}) = l_1(\theta, \bar{v} \circ (\mathrm{Id}+\theta)^{-1}).$$

A change of variables leads to the expression

$$\bar{l}_1(\theta, \bar{v}) = \int_{\Omega_0} |\det(I + D\theta)| \; f \circ (\mathrm{Id}+\theta) \; \bar{v} dx.$$

2. We consider now the boundary linear form

$$l_2(\theta, v) = \int_{\partial(\mathrm{Id}+\theta)(\Omega_0)} g \gamma_0 v ds \qquad \forall v \in H^1((\mathrm{Id}+\theta)(\Omega_0)).$$

Here we assume that Ω_0 is of class $\mathcal{C}^1$ and that $\theta \in \mathcal{C}_b^1(\mathbb{R}^N, \mathbb{R}^N) \cap B_{1,\infty}(\mathbb{R}^N, \mathbb{R}^N)$ and $g \in \mathcal{C}(\mathbb{R}^N)$. As previously we define

$$\bar{l}_2(\theta, \bar{v}) = l_2(\theta, \bar{v} \circ (\mathrm{Id}+\theta)^{-1}).$$

Using Theorem 5.22 we obtain

$$\bar{l}_2(\theta, \bar{v}) = \int_{\partial\Omega_0} g \circ (\mathrm{Id}+\theta) \gamma_0 \bar{v} |\det(I + D\theta)| \left| (I + D\theta)^{-\top} n \right| ds.$$

5.5.4 Differentiability results

Lemma 5.26 *The map*

$$\Phi : \Theta \in L^\infty(\mathbb{R}^N, GL_n(\mathbb{R})) \mapsto \Theta^{-1} \in L^\infty(\mathbb{R}^N)$$

is differentiable with derivative at I

$$d\Phi(I)\tilde{\Theta} = -\tilde{\Theta}.$$

Proof. It is an adaptation of the proof of Proposition 1.9, where it is easy to see that the remainder of the expansion can be bounded in the L^∞ norm. □

Lemma 5.27 *The map $\theta \in B_{1,\infty}(\mathbb{R}^N, \mathbb{R}^N) \mapsto C(\theta) \in L^\infty(\mathbb{R}^N, \mathcal{M}_n(\mathbb{R}))$ defined by (5.12) is differentiable at 0 with derivative*

$$dC(0)\tilde{\theta} = \operatorname{div}\tilde{\theta} I - D\tilde{\theta} - D\tilde{\theta}^\top.$$

Proof. Lemma 5.17 yields

$$\det(I + D\theta) = 1 + \operatorname{div}\theta + R_1, \qquad \lim_{\theta\to 0} \frac{\|R_1\|_{L^\infty(\mathbb{R}^N)}}{\|\theta\|_{W^{1,\infty}(\mathbb{R}^N,\mathbb{R}^N)}} = 0.$$

Lemma 5.26 yields

$$(I + D\theta)^{-1} = I - D\theta + R_2, \qquad \lim_{\theta\to 0} \frac{\|R_2\|_{L^\infty(\mathbb{R}^N)}}{\|\theta\|_{W^{1,\infty}(\mathbb{R}^N,\mathbb{R}^N)}} = 0.$$

Combining the two above expansions immediately lead to the claim. □

Lemma 5.28 *Let $f \in H^1(\mathbb{R}^N)$. The map $\Phi : \theta \in B_{1,\infty}(\mathbb{R}^N, \mathbb{R}^N) \mapsto |\det(I+D\theta)|\, f\circ(\operatorname{Id}+\theta) \in L^2(\mathbb{R}^N)$ is differentiable at 0 with derivative*

$$d\Phi(0)\tilde{\theta} = \operatorname{div}(f\tilde{\theta}).$$

Proof. This is an adaptation of the proof of Theorem 5.21. It is left to the reader. □

Lemma 5.29 *Let $g \in \mathcal{C}^1(\mathbb{R}^N)$ and suppose that Ω_0 is of class $\mathcal{C}^1$. The map $\Phi : \theta \in \mathcal{C}^1_b(\mathbb{R}^N, \mathbb{R}^N) \cap B_{1,\infty}(\mathbb{R}^N, \mathbb{R}^N) \mapsto g\circ(\operatorname{Id}+\theta)|\det(I+D\theta)|\,\left|(I + D\theta)^{-\top}n\right| \in L^1(\partial\Omega_0)$ is differentiable at 0 with derivative*

$$d\Phi(\theta)\tilde{\theta} = \nabla g \cdot \tilde{\theta} + g \operatorname{div}\tilde{\theta} - gD\tilde{\theta}n \cdot n.$$

Proof. This was done in step 1 of Theorem 5.23. □

5.5.5 Shape derivatives with Dirichlet boundary condition

We address the model problem

$$\begin{cases} -\Delta u = f \text{ in } \Omega \\ u = 0 \text{ on } \partial\Omega, \end{cases} \tag{5.13}$$

with $f \in H^1(\mathbb{R}^N)$. Using the notations of subsection 5.5.1 we have

$$a(\theta, u, v) = \int_{\Omega_\theta} \nabla u \cdot \nabla v dx, \qquad l(\theta, v) = \int_{\Omega_\theta} f v dx,$$

$$\theta \in B_{1,\infty}(\mathbb{R}^N, \mathbb{R}^N), \qquad \Omega_\theta = (\operatorname{Id}+\theta)(\Omega_0), \qquad H(\theta) = H^1_0(\Omega_\theta),$$

and $u(\theta)$ is the solution of (5.13) for $\Omega = \Omega_\theta$. In view of Propositions 5.24 and 5.25 as well as the derivations of subsections 5.5.2 and 5.5.3, the transported state $\bar{u}(\theta) = u(\theta) \circ (\operatorname{Id}+\theta) \in H^1_0(\Omega_0)$ solves

$$\bar{a}(\theta, \bar{u}(\theta), \varphi) = \bar{l}(\theta, \varphi) \qquad \forall\varphi \in H^1_0(\Omega_0)$$

with, incorporating $C(\theta)$ defined by (5.12),

$$\bar{a}(\theta,\bar{u},\bar{v}) = \int_{\Omega_0} C(\theta)\nabla\bar{u}\nabla\bar{v}dx, \qquad \bar{l}(\theta,\bar{v}) = \int_{\Omega_0} |\det(I+D\theta)|\ f\circ(\mathrm{Id}+\theta)\ \bar{v}dx.$$

Combining Proposition 4.3 with Lemmas 5.27 and 5.28, we infer that the map $\theta \in B_{1,\infty}(\mathbb{R}^N,\mathbb{R}^N) \mapsto \bar{u}(\theta)$ is differentiable at 0. We set $u_0 = u(0) = \bar{u}(0)$.

Theorem 5.30 *Let $\bar{J} : B_{1,\infty}(\mathbb{R}^N,\mathbb{R}^N) \times H^1_0(\Omega_0) \to \mathbb{R}$ be differentiable at $(0,u_0)$. The function $\theta \in B_{1,\infty}(\mathbb{R}^N,\mathbb{R}^N) \mapsto j(\theta) = \bar{J}(\theta,\bar{u}(\theta))$ is differentiable at 0 with*

$$dj(0)\tilde{\theta} = d_\theta\bar{J}(0,u_0)\tilde{\theta} + \int_{\Omega_0} (\operatorname{div}\tilde{\theta}I - D\tilde{\theta} - D\tilde{\theta}^\top)\nabla u_0\cdot\nabla v_0 dx - \int_{\Omega_0} \operatorname{div}(f\tilde{\theta})v_0dx,$$

where the adjoint $v_0 \in H^1_0(\Omega_0)$ is the solution of

$$\int_{\Omega_0} \nabla v_0\cdot\nabla\varphi dx = -d_u\bar{J}(0,u_0)\varphi \qquad \forall\varphi \in H^1_0(\Omega_0).$$

If Ω_0 is of class $\mathcal{C}^1$ and $u_0, v_0 \in H^2(\Omega_0)$ then

$$\boxed{dj(0)\tilde{\theta} = -\int_{\partial\Omega_0} \frac{\partial u_0}{\partial n}\frac{\partial v_0}{\partial n}\tilde{\theta}\cdot nds + B(\tilde{\theta})}$$

with

$$B(\tilde{\theta}) = d_\theta\bar{J}(0,u_0)\tilde{\theta} + \int_{\Omega_0} (\nabla u_0\cdot\tilde{\theta})\Delta v_0dx.$$

Proof. Following the approach carried out in Theorem 4.4 we define the Lagrangian

$$\mathcal{L}(\theta,\bar{u},\bar{v}) = \bar{J}(\theta,\bar{u}) + \bar{a}(\theta,\bar{u},\bar{v}) - \bar{l}(\theta,\bar{v}) \qquad \forall(\theta,\bar{u},\bar{v}) \in B_{1,\infty}(\mathbb{R}^N,\mathbb{R}^N)\times H^1_0(\Omega_0)\times H^1_0(\Omega_0).$$

Choosing $\bar{u} = \bar{u}(\theta)$ yields

$$j(\theta) = \mathcal{L}(\theta,\bar{u}(\theta),\bar{v}) \qquad \forall(\theta,\bar{v}) \in B_{1,\infty}(\mathbb{R}^N,\mathbb{R}^N)\times H^1_0(\Omega_0).$$

We differentiate at 0 by the chain rule:

$$dj(0)\tilde{\theta} = d_\theta\mathcal{L}(0,u_0,\bar{v})\tilde{\theta} + d_{\bar{u}}\mathcal{L}(0,u_0,\bar{v})(d\bar{u}(0)\tilde{\theta}).$$

The second derivative is equal to

$$d_{\bar{u}}\mathcal{L}(0,u_0,\bar{v})\tilde{u} = d_u\bar{J}(0,u_0)\tilde{u} + \bar{a}(0,\tilde{u},\bar{v}) = d_{\bar{u}}\bar{J}(0,u_0)\tilde{u} + \int_{\Omega_0}\nabla\tilde{u}\cdot\nabla\bar{v}dx.$$

It vanishes for every $\tilde{u} \in H^1_0(\Omega_0)$ when $\bar{v} = v_0$. We arrive at the classical expression

$$dj(0)\tilde{\theta} = d_\theta\mathcal{L}(0,u_0,v_0)\tilde{\theta}.$$

Let us write this Lagrangian:

$$\mathcal{L}(\theta,u_0,v_0) = \bar{J}(\theta,u_0) + \int_{\Omega_0} C(\theta)\nabla u_0\cdot\nabla v_0dx - \int_{\Omega_0} |\det(I+D\theta)|\ f\circ(\mathrm{Id}+\theta)\ v_0dx.$$

Lemmas 5.27 and 5.28 provide the derivative

$$dj(0)\tilde{\theta} = d_\theta\bar{J}(0,u_0)\tilde{\theta} + \int_{\Omega_0} (\operatorname{div}\tilde{\theta}I - D\tilde{\theta} - D\tilde{\theta}^\top)\nabla u_0\cdot\nabla v_0dx - \int_{\Omega_0}\operatorname{div}(f\tilde{\theta})v_0dx.$$

Suppose now that Ω_0 is of class $\mathcal{C}^1$ and $u_0, v_0 \in H^2(\Omega_0)$. We perform a first integration by parts to obtain

$$dj(0)\tilde{\theta} = d_\theta \bar{J}(0, u_0)\tilde{\theta} + \int_{\partial\Omega_0} \nabla u_0 \cdot \nabla v_0 \tilde{\theta} \cdot n ds - \int_{\Omega_0} \nabla(\nabla u_0 \cdot \nabla v_0) \cdot \tilde{\theta} dx$$
$$- \int_{\Omega_0} (D\tilde{\theta} + D\tilde{\theta}^\top)\nabla u_0 \cdot \nabla v_0 dx - \int_{\Omega_0} \operatorname{div}(f\tilde{\theta}) v_0 dx.$$

We now use

$$\begin{aligned} \nabla(\nabla u_0 \cdot \nabla v_0) \cdot \tilde{\theta} &= (\nabla^2 u_0 \nabla v_0) \cdot \tilde{\theta} + (\nabla^2 v_0 \nabla u_0) \cdot \tilde{\theta} \\ &= \nabla v_0 \cdot (\nabla^2 u_0 \tilde{\theta}) + \nabla u_0 \cdot (\nabla^2 v_0 \tilde{\theta}) \\ &= \nabla(\nabla u_0 \cdot \tilde{\theta}) \cdot \nabla v_0 - D\tilde{\theta} \nabla u_0 \cdot \nabla v_0 + \nabla(\nabla v_0 \cdot \tilde{\theta}) \cdot \nabla u_0 - D\tilde{\theta} \nabla v_0 \cdot \nabla u_0 \end{aligned}$$

to obtain

$$dj(0)\tilde{\theta} = d_\theta \bar{J}(0, u_0)\tilde{\theta} + \int_{\partial\Omega_0} \nabla u_0 \cdot \nabla v_0 \tilde{\theta} \cdot n ds - \int_{\Omega_0} \nabla(\nabla u_0 \cdot \tilde{\theta}) \cdot \nabla v_0 dx$$
$$- \int_{\Omega_0} \nabla(\nabla v_0 \cdot \tilde{\theta}) \cdot \nabla u_0 dx - \int_{\Omega_0} \operatorname{div}(f\tilde{\theta}) v_0 dx.$$

We now use integration by parts for the last three integrals:

$$dj(0)\tilde{\theta} = d_\theta \bar{J}(0, u_0)\tilde{\theta} + \int_{\partial\Omega_0} \nabla u_0 \cdot \nabla v_0 \tilde{\theta} \cdot n ds - \int_{\partial\Omega_0} (\nabla u_0 \cdot \tilde{\theta}) \nabla v_0 \cdot n ds + \int_{\Omega_0} (\nabla u_0 \cdot \tilde{\theta}) \Delta v_0 dx$$
$$- \int_{\partial\Omega_0} (\nabla v_0 \cdot \tilde{\theta}) \nabla u_0 \cdot n ds - \int_{\Omega_0} (\nabla v_0 \cdot \tilde{\theta}) f dx + \int_{\Omega_0} (f\tilde{\theta}) \cdot \nabla v_0 dx.$$

Observing that $\nabla u_0 = (\nabla u_0 \cdot n)n$ on $\partial\Omega_0$ due to the boundary condition, and that the same holds for v_0, this simplifies as

$$dj(0)\tilde{\theta} = d_\theta \bar{J}(0, u_0)\tilde{\theta} - \int_{\partial\Omega_0} \frac{\partial u_0}{\partial n} \frac{\partial v_0}{\partial n} \tilde{\theta} \cdot n ds + \int_{\Omega_0} (\nabla u_0 \cdot \tilde{\theta}) \Delta v_0 dx.$$

□

The term $B(\tilde{\theta})$ appearing in Theorem 5.30 is not straightforward to interpret as it involves the transported cost function $\bar{J}$. Also, it does not exhibit an explicit dependence on the normal displacement $\tilde{\theta} \cdot n$. We examine two particular cases. The first one deals with cost functions defined on a fixed part.

Proposition 5.31 *Let ω be an open, bounded subset of Ω_0 and define the subspace of displacements leaving ω invariant*

$$W^{1,\infty}_\omega(\mathbb{R}^N, \mathbb{R}^N) = \{\theta \in W^{1,\infty}(\mathbb{R}^N, \mathbb{R}^N) : \theta(x) = 0 \ \forall x \in \omega\}.$$

Suppose that

$$\bar{J}(\theta, \bar{u}) = J(\theta, (\bar{u} \circ (\mathrm{Id} + \theta)^{-1})_{|\omega}) \qquad \forall (\theta, \bar{u}) \in (W^{1,\infty}_\omega \cap B_{1,\infty})(\mathbb{R}^N, \mathbb{R}^N) \times H^1_0(\Omega_0)$$

where $J : (\theta, \bar{u}) \in (W^{1,\infty}_\omega \cap B_{1,\infty})(\mathbb{R}^N, \mathbb{R}^N) \times H^1(\omega) \to \mathbb{R}$ is differentiable at $(0, u_{0|\omega})$. Then we have

$$j(\theta) = \bar{J}(\theta, \bar{u}(\theta)) = J(\theta, u(\theta)_{|\omega}) \qquad \forall \theta \in (W^{1,\infty}_\omega \cap B_{1,\infty})(\mathbb{R}^N, \mathbb{R}^N),$$

and Theorem 5.30 applies with

$$d_{\bar{u}} \bar{J}(0, u_0)\varphi = d_u J(0, u_{0|\omega})\varphi_{|\omega} \qquad \forall \varphi \in H^1_0(\Omega_0)$$

$$B(\tilde{\theta}) = d_\theta J(0, u_{0|\omega})\tilde{\theta} \qquad \forall \tilde{\theta} \in W^{1,\infty}_\omega(\mathbb{R}^N, \mathbb{R}^N).$$

Proof. We first note that, by construction, we have

$$\bar{J}(\theta,\bar{u}) = J(\theta,\bar{u}_{|\omega}) \qquad \forall(\theta,\bar{u}) \in (W^{1,\infty}_{\omega} \cap B_{1,\infty})(\mathbb{R}^N,\mathbb{R}^N) \times H^1_0(\Omega_0).$$

This shows the differentiability of $\bar{J}$ at $(0,u_0)$. Of course, $d_\theta \bar{J}(0,u_0) = d_\theta J(0,u_{0|\omega})$. The adjoint state solves

$$\int_{\Omega_0} \nabla v_0 \cdot \nabla\varphi dx = -d_u J(0,u_{0|\omega})\varphi_{|\omega} \qquad \forall \varphi \in H^1_0(\Omega_0),$$

hence choosing $\varphi \in \mathcal{C}^\infty_0(\Omega_0 \setminus \overline{\omega})$ reveals that $\Delta v_0 = 0$ in $\Omega_0 \setminus \overline{\omega}$. Thus $B(\tilde{\theta}) = d_\theta J(0,u_{0|\omega})$ for any $\tilde{\theta} \in W^{1,\infty}_{\omega}(\mathbb{R}^N,\mathbb{R}^N)$. □

For example, for the least square cost

$$J(\theta,u) = \alpha\int_\omega |u-w|^2 dx + \beta\int_\omega |\nabla u - \nabla w|^2 dx, \qquad \alpha,\beta \geq 0, w \in H^1(\omega),$$

we simply have $B(\tilde{\theta}) = 0$.

The second case concerns L^2 type cost functions. When needed, we will implicitly extend by 0 functions in $L^2(\Omega)$, $\Omega \subset \mathbb{R}^N$, and consider them as elements of $L^2(\mathbb{R}^N)$.

Proposition 5.32 *Suppose that*

$$\bar{J}(\theta,\bar{u}) = J(\theta,\bar{u}\circ(\mathrm{Id}+\theta)^{-1}) \qquad \forall(\theta,\bar{u}) \in B_{1,\infty}(\mathbb{R}^N,\mathbb{R}^N) \times H^1_0(\Omega_0)$$

where $J : B_{1,\infty}(\mathbb{R}^N,\mathbb{R}^N) \times \underline{L^2(\mathbb{R}^N)}$ *is differentiable at* $(0,u_0)$*. Then we have*

$$j(\theta) = \bar{J}(\theta,\bar{u}(\theta)) = J(\theta,u(\theta)) \qquad \forall\theta \in B_{1,\infty}(\mathbb{R}^N,\mathbb{R}^N),$$

and Theorem 5.30 applies with

$$d_{\bar{u}}\bar{J}(0,u_0)\varphi = d_u J(0,u_0)\varphi \qquad \forall\varphi \in H^1_0(\Omega_0),$$

$$B(\tilde{\theta}) = d_\theta J(0,u_0)\tilde{\theta} \qquad \forall\tilde{\theta} \in W^{1,\infty}(\mathbb{R}^N,\mathbb{R}^N).$$

Proof. We differentiate $\bar{J}$ by the chain rule with the help of Lemma 5.19:

$$d_{\bar{u}}\bar{J}(0,u_0)\tilde{u} = d_u J(0,u_0)\tilde{u},$$

$$d_\theta\bar{J}(0,u_0)\tilde{\theta} = d_\theta J(0,u_0)\tilde{\theta} - d_u J(0,u_0)(\nabla u_0 \cdot \tilde{\theta}).$$

By the assumption $d_u J(0,u_0)$ identifies with an L^2 function, whereby the adjoint state satisfies

$$-\Delta v_0 = -d_u J(0,u_0) \text{ in } \Omega_0.$$

This leads to

$$B(\tilde{\theta}) = d_\theta J(0,u_0)\tilde{\theta} - d_u J(0,u_0)(\nabla u_0\cdot\tilde{\theta}) + \int_{\Omega_0}(\nabla u_0\cdot\tilde{\theta})\Delta v_0 dx = d_\theta J(0,u_0)\tilde{\theta}.$$

□

As example let us consider the compliance

$$j(\theta) = \int_{\Omega_\theta} f u(\theta) dx = J(\theta,u(\theta))$$

with

$$J(\theta,u) = \int_{\Omega_\theta} f u dx \qquad \forall(\theta,u) \in B_{1,\infty}(\mathbb{R}^N,\mathbb{R}^N)\times L^2(\mathbb{R}^N).$$

We have

$$d_uJ(0,u_0)\tilde{u} = \int_{\Omega_0} f\tilde{u}dx,$$

whereby the adjoint state is identified as $v_0 = -u_0$. We assume that Ω_0 is of class $\mathcal{C}^1$. We have by Theorem 5.21

$$d_\theta J(0,u_0)\tilde{\theta} = \int_{\partial\Omega_0} \gamma_0(fu_0\tilde{\theta})\cdot n = 0.$$

If furthermore Ω_0 is of class $\mathcal{C}^2$ then we have by elliptic regularity that $u_0 \in H^2(\Omega_0)$. Theorem 5.30 and Proposition 5.32 yield

$$dj(0)\tilde{\theta} = \int_{\partial\Omega_0} \left(\frac{\partial u_0}{\partial n}\right)^2 \tilde{\theta}\cdot nds.$$

This has a sign: the compliance increases when the domain is enlarged. This makes sense when interpreting the Dirichlet condition as a clamped condition.

5.5.6 Case of a Neumann boundary condition

We address the mixed problem

$$\begin{cases} -\Delta u = f \text{ in } \Omega \\ u = 0 \text{ on } \Gamma_D \\ \dfrac{\partial u}{\partial n} = g \text{ on } \partial\Omega\setminus\Gamma_D \end{cases} \tag{5.14}$$

with $f \in H^1(\mathbb{R}^N)$, $g \in \mathcal{C}^1(\mathbb{R}^N)$. We only allow variations of the Neumann part of the boundary. Therefore, we consider an open, bounded subset ω of Ω_0 such that $\Gamma_D \subset \overline{\omega}$, and we will consider displacements within

$$\mathcal{C}^1_{b,\omega}(\mathbb{R}^N,\mathbb{R}^N) = \{\theta \in \mathcal{C}^1_b(\mathbb{R}^N,\mathbb{R}^N) : \theta(x) = 0\ \forall x \in \omega\}.$$

The $\mathcal{C}^1$ regularity is needed to transport boundary integrals, but when $g = 0$ we can work with the more general set $W^{1,\infty}_\omega(\mathbb{R}^N,\mathbb{R}^N)$. We also assume that Ω_0 is of class $\mathcal{C}^1$, unless $g = 0$.

Here we have

$$a(\theta,u,v) = \int_{\Omega_\theta} \nabla u\cdot\nabla vdx, \qquad l(\theta,v) = \int_{\Omega_\theta} fvdx + \int_{\partial\Omega_\theta} g\gamma_0 vds,$$

$$\theta \in C^1_{b,\omega}(\mathbb{R}^N,\mathbb{R}^N), \qquad \Omega_\theta = (\mathrm{Id}+\theta)(\Omega_0), \qquad H(\theta) = \{\varphi \in H^1(\Omega_\theta) : \gamma_0\varphi = 0 \text{ on } \partial\Gamma_D\}.$$

The transported state $\bar{u}(\theta) = u(\theta)\circ(\mathrm{Id}+\theta) \in H_0 = H(0)$ solves

$$\bar{a}(\theta,\bar{u},\bar{v}) = \bar{l}(\theta,\bar{v}) \qquad \forall\bar{v} \in H_0$$

with, incorporating $C(\theta)$ defined by (5.12),

$$\bar{a}(\theta,\bar{u},\bar{v}) = \int_{\Omega_0} C(\theta)\nabla\bar{u}\cdot\nabla\bar{v}dx,$$

$$\bar{l}(\theta,\bar{v}) = \int_{\Omega_0} |\det(I+D\theta)|\ f\circ(\mathrm{Id}+\theta)\ \bar{v}dx + \int_{\partial\Omega_0} g\circ(\mathrm{Id}+\theta)\gamma_0\bar{v}|\det(I+D\theta)|\left|(I+D\theta)^{-\top}n\right|ds.$$

Combining Proposition 4.3 with Lemmas 5.27, 5.28 and 5.29, we infer that the map $\theta \in (\mathcal{C}^1_{b,\omega} \cap B_{1,\infty})(\mathbb{R}^N,\mathbb{R}^N) \mapsto \bar{u}(\theta)$ is differentiable at 0. We set $u_0 = u(0) = \bar{u}(0)$.

Theorem 5.33 *Let $\bar{J} : (\mathcal{C}^1_{b,\omega} \cap B_{1,\infty})(\mathbb{R}^N, \mathbb{R}^N) \times H_0 \to \mathbb{R}$ be differentiable at $(0, u_0)$. The function $\theta \in (\mathcal{C}^1_{b,\omega} \cap B_{1,\infty})(\mathbb{R}^N, \mathbb{R}^N) \mapsto j(\theta) = \bar{J}(\theta, \bar{u}(\theta))$ is differentiable at 0 with*

$$dj(0)\tilde{\theta} = d_\theta \bar{J}(0, u_0)\tilde{\theta} + \int_{\Omega_0} (\operatorname{div}\tilde{\theta} I - D\tilde{\theta} - D\tilde{\theta}^\top)\nabla u_0 \cdot \nabla v_0 dx - \int_{\Omega_0} \operatorname{div}(f\tilde{\theta}) v_0 dx \\ - \int_{\partial\Omega_0} \left(\nabla g \cdot \tilde{\theta} + g \operatorname{div}\tilde{\theta} - g D\tilde{\theta} n \cdot n\right) \gamma_0 v_0 ds,$$

where the adjoint $v_0 \in H_0$ is the solution of

$$\int_{\Omega_0} \nabla v_0 \cdot \nabla \varphi dx = -d_u \bar{J}(0, u_0)\varphi \qquad \forall \varphi \in H_0.$$

If Ω_0 is of class $\mathcal{C}^2$ and $u_0, v_0 \in H^2(\Omega_0)$ then

$$\boxed{dj(0)\tilde{\theta} = \int_{\partial\Omega_0} \left(\nabla u_0 \cdot \nabla v_0 - f v_0 - \kappa g v_0 - \frac{\partial(g v_0)}{\partial n}\right) \tilde{\theta} \cdot n ds + B(\tilde{\theta})}$$

with

$$B(\tilde{\theta}) = d_\theta \bar{J}(0, u_0)\tilde{\theta} + d_{\bar{u}} \bar{J}(0, u_0)(\nabla u_0 \cdot \tilde{\theta}).$$

Proof. We define the Lagrangian

$$\mathcal{L}(\theta, \bar{u}, \bar{v}) = \bar{J}(\theta, \bar{u}) + \bar{a}(\theta, \bar{u}, \bar{v}) - \bar{l}(\theta, \bar{v}) \qquad \forall (\theta, \bar{u}, \bar{v}) \in (\mathcal{C}^1_{b,\omega} \cap B_{1,\infty})(\mathbb{R}^N, \mathbb{R}^N) \times H_0 \times H_0.$$

Choosing $\bar{u} = \bar{u}(\theta)$ yields

$$j(\theta) = \mathcal{L}(\theta, \bar{u}(\theta), \bar{v}) \qquad \forall (\theta, \bar{v}) \in (\mathcal{C}^1_{b,\omega} \cap B_{1,\infty})(\mathbb{R}^N, \mathbb{R}^N) \times H_0.$$

We differentiate at 0 by the chain rule:

$$dj(0)\tilde{\theta} = d_\theta \mathcal{L}(0, u_0, \bar{v})\tilde{\theta} + d_{\bar{u}} \mathcal{L}(0, u_0, \bar{v})(d\bar{u}(0)\tilde{\theta}).$$

The second derivative is equal to

$$d_{\bar{u}} \mathcal{L}(0, u_0, \bar{v})\tilde{u} = d_u \bar{J}(0, u_0)\tilde{u} + \bar{a}(0, \tilde{u}, \bar{v}) = d_{\bar{u}} \bar{J}(0, u_0)\tilde{u} + \int_{\Omega_0} \nabla \tilde{u} \cdot \nabla \bar{v} dx,$$

which vanishes for every $\tilde{u} \in H_0$ when $\bar{v} = v_0$. We arrive at

$$dj(0)\tilde{\theta} = d_\theta \mathcal{L}(0, u_0, v_0)\tilde{\theta}.$$

The Lagrangian admits the expression:

$$\mathcal{L}(\theta, u_0, v_0) = \bar{J}(\theta, u_0) + \int_{\Omega_0} C(\theta)\nabla u_0 \cdot \nabla v_0 dx - \int_{\Omega_0} |\det(I + D\theta)| \; f \circ (\operatorname{Id} + \theta) \; v_0 dx \\ - \int_{\partial\Omega_0} g \circ (\operatorname{Id} + \theta)\gamma_0 v_0 |\det(I + D\theta)| \left|(I + D\theta)^{-\top} n\right| ds.$$

Lemmas 5.27, 5.28 and 5.29 provide the derivative

$$dj(0)\tilde{\theta} = d_\theta \bar{J}(0, u_0)\tilde{\theta} + \int_{\Omega_0} (\operatorname{div}\tilde{\theta} I - D\tilde{\theta} - D\tilde{\theta}^\top)\nabla u_0 \cdot \nabla v_0 dx - \int_{\Omega_0} \operatorname{div}(f\tilde{\theta}) v_0 dx \\ - \int_{\partial\Omega_0} \left(\nabla g \cdot \tilde{\theta} + g \operatorname{div}\tilde{\theta} - g D\tilde{\theta} n \cdot n\right) \gamma_0 v_0 ds.$$

Suppose now that Ω_0 is of class $\mathcal{C}^1$ and $u_0, v_0 \in H^2(\Omega_0)$. The same calculation as in Theorem 5.30 yield

$$dj(0)\tilde{\theta} = d_\theta \bar{J}(0,u_0)\tilde{\theta} + \int_{\partial\Omega_0} \nabla u_0 \cdot \nabla v_0 \tilde{\theta}\cdot n ds - \int_{\Omega_0} \nabla(\nabla u_0 \cdot \tilde{\theta})\cdot \nabla v_0 dx - \int_{\Omega_0} \nabla(\nabla v_0 \cdot \tilde{\theta})\cdot \nabla u_0 dx$$
$$- \int_{\Omega_0} \operatorname{div}(f\tilde{\theta}) v_0 dx - \int_{\partial\Omega_0} \left(v_0 \nabla g \cdot \tilde{\theta} + g v_0 \operatorname{div}\tilde{\theta} - g v_0 D\tilde{\theta} n \cdot n \right) ds.$$

The function $\nabla u_0 \cdot \tilde{\theta}$ acts as test function for the adjoint equation, as well as the function $\nabla v_0 \cdot \tilde{\theta}$ acts as test function for the state equation. We obtain

$$dj(0)\tilde{\theta} = d_\theta \bar{J}(0,u_0)\tilde{\theta} + \int_{\partial\Omega_0} \nabla u_0 \cdot \nabla v_0 \tilde{\theta}\cdot n ds + d_{\bar{u}}\bar{J}(0,u_0)(\nabla u_0 \cdot \tilde{\theta})$$
$$- \int_{\Omega_0} f \nabla v_0 \cdot \tilde{\theta} dx - \int_{\partial\Omega_0} g \nabla v_0 \cdot \tilde{\theta} ds - \int_{\Omega_0} \operatorname{div}(f\tilde{\theta}) v_0 dx - \int_{\partial\Omega_0} \left(v_0 \nabla g \cdot \tilde{\theta} + g v_0 \operatorname{div}\tilde{\theta} - g v_0 D\tilde{\theta} n \cdot n \right) ds.$$

This simplifies as

$$dj(0)\tilde{\theta} = d_\theta \bar{J}(0,u_0)\tilde{\theta} + \int_{\partial\Omega_0} \nabla u_0 \cdot \nabla v_0 \tilde{\theta}\cdot n ds + d_{\bar{u}}\bar{J}(0,u_0)(\nabla u_0 \cdot \tilde{\theta})$$
$$- \int_{\partial\Omega_0} f v_0 \tilde{\theta}\cdot n ds - \int_{\partial\Omega_0} \left(\nabla(g v_0) \cdot \tilde{\theta} + g v_0 \operatorname{div}\tilde{\theta} - g v_0 D\tilde{\theta} n \cdot n \right) ds.$$

Using the arguments of steps 2 and 3 of the proof of Theorem 5.23 we can reformulate the last integral to obtain

$$dj(0)\tilde{\theta} = d_\theta \bar{J}(0,u_0)\tilde{\theta} + \int_{\partial\Omega_0} \nabla u_0 \cdot \nabla v_0 \tilde{\theta}\cdot n ds + d_{\bar{u}}\bar{J}(0,u_0)(\nabla u_0 \cdot \tilde{\theta})$$
$$- \int_{\partial\Omega_0} f v_0 \tilde{\theta}\cdot n ds - \int_{\partial\Omega_0} \left(\kappa g v_0 + \frac{\partial(g v_0)}{\partial n} \right) \tilde{\theta}\cdot n ds.$$

□

Propositions 5.31 and 5.32 apply similarly to the above situation.

Proposition 5.34 *Suppose that*

(i) either $\bar{J}(\theta,\bar{u}) = J(\theta, (\bar{u}\circ(\operatorname{Id}+\theta)^{-1})_{|\omega}) \quad \forall(\theta,\bar{u}) \in (\mathcal{C}^1_{b,\omega} \cap B_{1,\infty})(\mathbb{R}^N,\mathbb{R}^N) \times H_0,$ *where* $J : B_{1,\infty}(\mathbb{R}^N,\mathbb{R}^N) \times H^1(\omega)$ *is differentiable at* $(0, u_{0|\omega})$*, meaning that*

$$j(\theta) = \bar{J}(\theta,\bar{u}(\theta)) = J(\theta, u(\theta)_{|\omega}) \qquad \forall\theta \in (\mathcal{C}^1_{b,\omega} \cap B_{1,\infty})(\mathbb{R}^N,\mathbb{R}^N);$$

(ii) or $\bar{J}(\theta,\bar{u}) = J(\theta, \bar{u}\circ(\operatorname{Id}+\theta)^{-1}) \quad \forall(\theta,\bar{u}) \in (\mathcal{C}^1_{b,\omega} \cap B_{1,\infty})(\mathbb{R}^N,\mathbb{R}^N) \times H_0,$ *where* $J : B_{1,\infty}(\mathbb{R}^N,\mathbb{R}^N) \times L^2(\mathbb{R}^N)$ *is differentiable at* $(0, u_0)$*, meaning that*

$$j(\theta) = \bar{J}(\theta,\bar{u}(\theta)) = J(\theta, u(\theta)) \qquad \forall\theta \in (\mathcal{C}^1_{b,\omega} \cap B_{1,\infty})(\mathbb{R}^N,\mathbb{R}^N).$$

Then Theorem 5.33 applies with

$$\text{in case (i)} \qquad \begin{cases} d_{\bar{u}}\bar{J}(0,u_0)\varphi = d_u J(0,u_{0|\omega})\varphi_{|\omega} & \forall\varphi \in H_0, \\ B(\tilde{\theta}) = d_\theta J(0,u_{0|\omega})\tilde{\theta} & \forall\tilde{\theta} \in (\mathcal{C}^1_{b,\omega} \cap B_{1,\infty})(\mathbb{R}^N,\mathbb{R}^N); \end{cases}$$

$$\text{in case (ii)} \qquad \begin{cases} d_{\bar{u}}\bar{J}(0,u_0)\varphi = d_u J(0,u_0)\varphi & \forall\varphi \in H_0, \\ B(\tilde{\theta}) = d_\theta J(0,u_0)\tilde{\theta} & \forall\tilde{\theta} \in (\mathcal{C}^1_{b,\omega} \cap B_{1,\infty})(\mathbb{R}^N,\mathbb{R}^N). \end{cases}$$

Proof. In case (i) we have

$$\bar{J}(\theta, \bar{u}) = J(\theta, \bar{u}_{|\omega}) \qquad \forall (\theta, \bar{u}) \in (\mathcal{C}^1_{b,\omega} \cap B_{1,\infty})(\mathbb{R}^N, \mathbb{R}^N) \times H_0,$$

from which the statements follow straightforwardly.

In case (ii) we differentiate $\bar{J}$ by the chain rule with the help of Lemma 5.19:

$$d_{\bar{u}}\bar{J}(0, u_0)\tilde{u} = d_u J(0, u_0)\tilde{u},$$

$$d_\theta \bar{J}(0, u_0)\tilde{\theta} = d_\theta J(0, u_0)\tilde{\theta} - d_u J(0, u_0)(\nabla u_0 \cdot \tilde{\theta}).$$

This proves the claims. □

Remark 5.35 *In the setting of Proposition 5.34, we can define for all $u \in H^1(\mathbb{R}^N)$ and $\theta \in (\mathcal{C}^1_{b,\omega} \cap B_{1,\infty})(\mathbb{R}^N, \mathbb{R}^N)$*

$$\hat{J}(\theta, u) = \begin{cases} J(\theta, u_{|\omega}) \text{ in case (i)} \\ J(\theta, u_{|\Omega_\theta}) \text{ in case (ii)} \end{cases}.$$

This allows to define a natural (without transport) Lagrangian on $(\mathcal{C}^1_{b,\omega} \cap B_{1,\infty})(\mathbb{R}^N, \mathbb{R}^N) \times H^1(\mathbb{R}^N) \times H^1(\mathbb{R}^N)$

$$L(\theta, u, v) = \hat{J}(\theta, u) + \int_{\Omega_\theta} \nabla u \cdot \nabla v dx - \int_{\Omega_\theta} f v dx - \int_{\partial\Omega_\theta} g \gamma_0 v ds.$$

Then, considering the differentiation rules from Theorems 5.21 and 5.23, Theorem 5.33 in the smooth case can be rephrased as

$$\boxed{dj(0)\tilde{\theta} = d_\theta L(0, u_0, v_0)\tilde{\theta}.}$$

We retrieve the classical formula, as if the tedious transport of the boundary value problem could be bypassed. Actually this direct approach can be partly justified. It is known as Céa's fast derivation method. It can be adapted to the Dirichlet case but it is more tricky: in order to define the Lagrangian in terms of independent variables one treats the Dirichlet condition through the introduction of a Lagrange multiplier. Details can be found in [1].

As example let us consider again the compliance

$$j(\theta) = J(\theta, u(\theta)) = \int_{\Omega_\theta} f u(\theta) dx + \int_{\partial\Omega_0} g \gamma_0 u(\theta) ds.$$

We assume that $f = g = 0$ in a neighborhood of the moving part of the boundary, so that we enter into the framework (i). Proceeding as in the Dirichlet case we obtain that $v_0 = -u_0$. By Theorems 5.21 and 5.23 we have

$$d_\theta J(0, u_0)\tilde{\theta} = \int_{\partial\Omega_0} (f u_0 \tilde{\theta}) \cdot n ds + \int_{\partial\Omega_0} \left(\frac{\partial(g u_0)}{\partial n} + \kappa g u_0 \right) \tilde{\theta} \cdot n ds = 0.$$

We derive from Theorem 5.33 and Proposition 5.32 the shape derivative

$$dj(0)\tilde{\theta} = -\int_{\partial\Omega_0} |\nabla u_0|^2 \tilde{\theta} \cdot n ds,$$

showing that the compliance decreases when the domain is enlarged. It is the opposite to the Dirichlet case, but this is logical.

5.5.7 Extension to the linear elasticity case

Consider the linear elasticity system described in section 1.3.4:

$$\begin{cases} -\operatorname{div}\sigma(u) = f & \text{in } \Omega \\ u = 0 & \text{on } \Gamma_D \\ \sigma(u)n = g & \text{on } \partial\Omega \setminus \Gamma_D. \end{cases} \tag{5.15}$$

We focus on the Neumann case as in subsection 5.5.6, i.e. the Dirichlet part Γ_D is fixed. Here we have

$$a(\theta,u,v) = \int_{\Omega_\theta} A\nabla^s u : \nabla^s v dx, \qquad l(\theta,v) = \int_{\Omega_\theta} f\cdot v dx + \int_{\partial\Omega_\theta} g\cdot\gamma_0 v ds,$$

where the superscript s stands for the symmetric part, namely $\nabla^s u = e(u)$. The transported forms read

$$\bar{a}(\theta,\bar{u},\bar{v}) = \int_{\Omega_0} A\ (\nabla\bar{u}(\mathrm{Id}+D\theta)^{-1})^s : (\nabla\bar{v}(\mathrm{Id}+D\theta)^{-1})^s\ |\det(I+D\theta)|dx,$$

$$\bar{l}(\theta,\bar{v}) = \int_{\Omega_0} |\det(I+D\theta)|\ f\circ(\mathrm{Id}+\theta)\cdot\bar{v}dx + \int_{\partial\Omega_0} g\circ(\mathrm{Id}+\theta)\cdot\gamma_0\bar{v}|\det(I+D\theta)|\left|(I+D\theta)^{-\top}n\right| ds.$$

The derivatives can be computed along the same procedure as in Theorem 5.33. We arrive at the same conclusion as in Remark 5.35. For instance this gives for the compliance and boundary variations in regions where $f = g = 0$

$$dj(0)\tilde{\theta} = -\int_{\partial\Omega_0} A\nabla^s u_0 : \nabla^s u_0\ \tilde{\theta}\cdot n ds = -\int_{\partial\Omega_0} \sigma(u_0) : e(u_0)\ \tilde{\theta}\cdot n ds.$$

5.6 Numerical aspects

5.6.1 General principles

The shape derivative can be used in various ways to design a shape optimization procedure. The basic scheme in order to minimize a shape functional $\mathcal{J}$ is a gradient type algorithm with the following structure.

1. Start with an initial shape Ω_0.
2. Iterate until some stopping criterion is reached:
   (a) Compute the shape derivative for the current shape Ω_k and write it in the form
   $$d_S\mathcal{J}(\Omega_k,\theta) = \int_{\partial\Omega_k} G_k\theta\cdot n ds;$$
   (b) Define Ω_{k+1} by moving the boundary points of Ω_k as $x_{k+1} = x_k - tG_k n$, where t is a stepsize determined by a line search to ensure a decrease of the cost.

In practice, a regularization of the descent direction is often performed to preserve the smoothness of the shape.

Another difficulty encountered in the implementation of this algorithm is the deformation of the mesh, or even full remeshing when the mesh quality is too degraded, that needs to be done not only at each iteration, but also within the line search at each cost function evaluation. To perform mesh deformation one needs to extend the displacement field within Ω_k. There are several techniques for that.

To avoid this, alternative approaches have been developed to work on a fixed mesh. In particular level-set methods consist in defining Ω_k as $\Omega_k = \{x \in D : \psi_k(x) < 0\}$ where D is the hold all and the function ψ_k is now considered as the design variable. Several techniques exist to update ψ_k in relation with G_k.

5.6.2 Example

We consider in Figure 5.1 the classical cantilever problem governed by (5.15). The cost function is the compliance augmented with a volume penalization. The algorithm is the one described above with mesh deformation / remeshing. We clearly observe how the topology of the final shape is imposed by the initialization. We also see that the algorithm is unable to remove the thin band in the last case, although it is useless, because this removal would be a topology change.

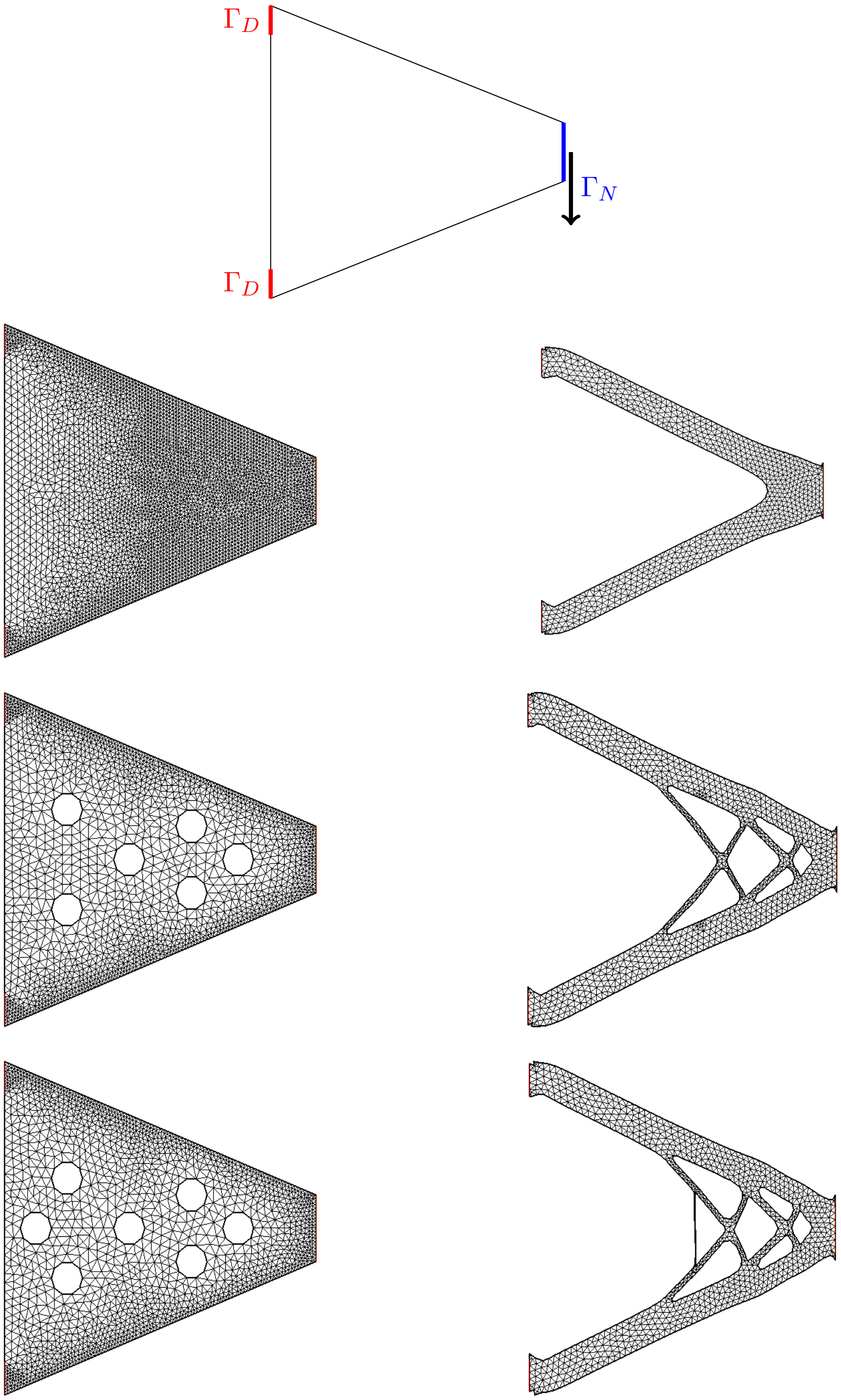


Figure 5.1: Cantilever problem: boundary contidions (top), initialization and optimized design without hole (second row), initialization and optimized design with holes (third and fourth rows).

Chapter 6

Topological derivative

6.1 Introduction

6.1.1 Principle of the topological sensitivity analysis

We have developed in the previous chapter a concept of sensitivity of shape functionals with respect to smooth boundary variations. We now inverstigate the creation of new boundaries by making holes. Such singular perturbations do not allow the use of Fréchet differential calculus, except in the "easy cases" described later in this section. Instead we will directly work with asymptotic expansions.

Consider a shape functional

$$\Omega \in \mathcal{A} \mapsto \mathcal{J}(\Omega) \in \mathbb{R},$$

where $\mathcal{A}$ is the set of all open subsets of an hold all D of $\mathbb{R}^N$. The concept of topological derivative was formally introduced in [12], and the first mathematical justifications appeared in [20, 15].

Definition 6.1 *Let ω be a bounded open subset of $\mathbb{R}^N$. We say that $\mathcal{J}$ admits a topological derivative at $\Omega_0 \in \mathcal{A}$ and at the point $z \in \Omega_0$ with respect to ω if there exists a function $f : \mathbb{R}_+ \to \mathbb{R}_+$ with $\lim_{\varepsilon \searrow 0} f(\varepsilon) = 0$ such that the following limit exists:*

$$d_T\mathcal{J}(\Omega_0, \omega, z) = \lim_{\varepsilon \searrow 0} \frac{\mathcal{J}(\Omega_0 \setminus (\overline{z + \varepsilon\omega})) - \mathcal{J}(\Omega_0)}{f(\varepsilon)}. \tag{6.1}$$

Of course, (6.1) is equivalent to the "topological asymptotic expansion":

$$\mathcal{J}(\Omega_0 \setminus (\overline{z + \varepsilon\omega})) - \mathcal{J}(\Omega_0) = f(\varepsilon) d_T\mathcal{J}(\Omega_0, \omega, z) + o(f(\varepsilon)).$$

The set ω is the hole of reference. It is typically chosen as the unit ball, but other cases are of interest like cracks. In the standard case where the shape functional involves a boundary value problem, the type of boundary condition around the hole plays a crucial role.

6.1.2 Basic purely geometric cases

For the volume functional $\mathcal{J}(\Omega) = |\Omega|$, we obviously have $d_T\mathcal{J}(\Omega, \omega, z) = -1$ for all $z \in \Omega$ with $f(\varepsilon) = |z + \varepsilon\omega| = \varepsilon^N|\omega|$.

For the perimeter functional $\mathcal{J}(\Omega) = \int_{\partial\Omega} ds$ we have $d_T\mathcal{J}(\Omega, \omega, z) = 1$ for all $z \in \Omega$ with $f(\varepsilon) = \varepsilon^{N-1}\int_{\partial\omega} ds$.

6.1.3 Easy PDE cases: perturbation of non principal parts

Let D be a bounded open subset of $\mathbb{R}^N$. Here it is important to assume that $N \in \{1, 2, 3\}$. We recall the Sobolev embedding $H_0^1(D) \hookrightarrow L^p(D)$ with p indicated in Table 6.1. We are given functions

N	p	q	r
1	$\leq +\infty$	> 2	> 2
2	$< +\infty$	> 2	> 2
3	≤ 6	> 6	> 3

Table 6.1: Lebesgue exponents

$h_0, h_1 \in L^q(D)$, $f_0, f_1 \in L^r(D)$ with $h_0, h_1 \geq 0$ and q, r given in Table 6.1. For any measurable $\Omega \subset D$ we consider the boundary value problem

$$\begin{cases} -\Delta u_\Omega + h_\Omega u_\Omega = f_\Omega \text{ in } D \\ u_\Omega = 0 \text{ on } \partial D, \end{cases}$$

where h_Ω and f_Ω are defined by

$$h_\Omega = \chi_\Omega h_1 + (1 - \chi_\Omega) h_0, \qquad f_\Omega = \chi_\Omega f_1 + (1 - \chi_\Omega) f_0.$$

Consider a shape functional of the form $\mathcal{J}(\Omega) = J(u_\Omega)$ where $J : H_0^1(D) \to \mathbb{R}$ is of class $\mathcal{C}^2$.

Proposition 6.2 *The above shape functional admits at a.e. $z \in \Omega$ the topological derivative*

$$d_T \mathcal{J}(\Omega, \omega, z) = \left[(h_0 - h_1) u_\Omega v_\Omega - (f_0 - f_1) v_\Omega\right](z)$$

with $f(z) = \varepsilon^N |\omega|$, where the adjoint state v_Ω is the solution of

$$\int_D (\nabla v_\Omega \cdot \nabla \varphi + h_\Omega v_\Omega \varphi) dx = -dJ(u_\Omega) \varphi dx \qquad \forall \varphi \in H_0^1(D).$$

Proof. Given $p > 2$ according to Table 6.1 we define, for all $(u, v, h, f) \in H_0^1(D) \times H_0^1(D) \times L^{p/(p-2)}(D) \times L^{p/(p-1)}(D)$,

$$a(h, u, v) = \int_D (\nabla u \cdot \nabla v + huv) dx, \qquad l(f, v) = \int_D fv dx.$$

This is well defined since $uv \in L^{p/2}(D)$ by the Cauchy-Schwarz inequality. For any $(h, f) \in L^{p/(p-2)}(D) \times L^{p/(p-1)}(D)$ we define $A(h) \in \mathcal{L}(H_0^1(D), H^{-1}(D))$ and $L(f) \in H^{-1}(D)$ by

$$\langle A(h) u, v \rangle_{H^{-1}(D), H_0^1(D)} = a(h, u, v), \qquad \langle L(f), v \rangle_{H^{-1}(D), H_0^1(D)} = l(f, v).$$

Set

$$\mathcal{U} = \left\{ (h, f) \in L^{p/(p-2)}(D) \times L^{p/(p-1)}(D),\ h \geq 0 \text{ a.e.} \right\}.$$

By the Lax-Milgram theorem and the Poincaré inequality we have $A(h) \in \text{isom}(H_0^1(D), H^{-1}(D))$ for every (h, f) in a neighborhood of any $(f_0, h_0) \in \mathcal{U}$. In addition, due to its affine structure, the map $(h, f) \in L^{p/(p-2)}(D) \times L^{p/(p-1)}(D) \mapsto (A(h), L(f))$ is of class $\mathcal{C}^\infty$. By Proposition 4.3, the map $(h, f) \mapsto u(h, f) := A(h)^{-1} L(f)$ is differentiable in a neighborhood of any element of $\mathcal{U}$. In fact it is not difficult to see that it is of class $\mathcal{C}^2$, even $\mathcal{C}^\infty$, in view of the expression of the derivative of the inverse mapping given in Proposition 1.9. In order to find a convenient expression of the derivative we use the adjoint method. We define the by now standard Lagrangian

$$\mathcal{L}(h, f, u, v) = J(u) + a(h, u, v) - l(f, v).$$

We differentiate $j(h, f) := J(u(h, f)) = \mathcal{L}(h, f, u(h, f), v)$ at (h_Ω, f_Ω):

$$dj(h_\Omega, f_\Omega)(\hat{u}, \hat{f}) = d_h \mathcal{L}(h_\Omega, f_\Omega, u(h_\Omega, f_\Omega), v) \hat{h} + d_f \mathcal{L}(h_\Omega, f_\Omega, u(h_\Omega, f_\Omega), v) \hat{f}$$
$$+ d_u \mathcal{L}(h_\Omega, f_\Omega, u(h_\Omega, f_\Omega), v) du(h_\Omega, f_\Omega)(\hat{h}, \hat{f}).$$

Choosing v as the adjoint state cancels the last term. We arrive at

$$dj(h_\Omega, f_\Omega)(\hat{u}, \hat{f}) = \int_D \hat{h} u_\Omega v_\Omega dx - \int_D \hat{f} v_\Omega dx.$$

By Taylor-Lagrange expansion this implies

$$j(h_\Omega + \hat{h}, f_\Omega + \hat{f}) - j(h_\Omega, f_\Omega) = \int_D \hat{h} u_\Omega v_\Omega dx - \int_D \hat{f} v_\Omega dx + O(\|\hat{h}\|^2_{L^{p/(p-2)}(D)} + \|\hat{f}\|^2_{L^{p/(p-1)}(D)}).$$

Choose $z \in \Omega$ and ε small enough so that

$$\hat{h} := h_{\Omega \setminus (\overline{z+\varepsilon\omega})} - h_\Omega = \chi_{z+\varepsilon\omega}(h_0 - h_1), \qquad \hat{f} := f_{\Omega \setminus (\overline{z+\varepsilon\omega})} - f_\Omega = \chi_{z+\varepsilon\omega}(f_0 - f_1).$$

It is straightforward from Hölder's inequality that

$$\|\hat{h}\|_{L^{p/(p-2)}(D)} \le (\varepsilon^N |\omega|)^{1-\frac{2}{p}-\frac{1}{q}} \|h_0 - h_1\|_{L^q(D)}$$

$$\|\hat{f}\|_{L^{p/(p-1)}(D)} \le (\varepsilon^N |\omega|)^{1-\frac{1}{p}-\frac{1}{r}} \|f_0 - f_1\|_{L^r(D)}.$$

With the assumptions made on q and r we can adjust p in accordance with Table 6.1 in order to have

$$\|\hat{h}\|^2_{L^{p/(p-2)}(D)} + \|\hat{f}\|^2_{L^{p/(p-1)}(D)} = o(\varepsilon^N).$$

So far we have shown that

$$\mathcal{J}(\Omega \setminus (\overline{z+\varepsilon\omega})) - \mathcal{J}(\Omega) = \int_{z+\varepsilon\omega} ((h_0 - h_1) u_\Omega v_\Omega - (f_0 - f_1) v_\Omega)\, dx + o(\varepsilon^N).$$

This can be rewritten as

$$\mathcal{J}(\Omega \setminus (\overline{z+\varepsilon\omega})) - \mathcal{J}(\Omega) = |\omega| \varepsilon^N \left[(h_0 - h_1) u_\Omega v_\Omega - (f_0 - f_1) v_\Omega\right](z)$$
$$+ \int_{z+\varepsilon\omega} (((h_0 - h_1) u_\Omega v_\Omega - (f_0 - f_1) v_\Omega)(x) - ((h_0 - h_1) u_\Omega v_\Omega - (f_0 - f_1) v_\Omega)(z))\, dx + o(\varepsilon^N).$$

By Lebesgue's differentiation theorem this latter integral is a $o(\varepsilon^N)$ for a.e. $z \in \Omega$. □

We observe here that the topological derivative does not depend on the shape of ω. We will see later that this is not a universal property.

6.2 A generalized adjoint method

In the previous section we have been able to use Fréchet differential calculus thanks to Sobolev embeddings and the fact that the characteristic function of the hole was small in an appropriate L^p norm. When we perturb the principal part of the differential operator, differential calculus applies in L^∞, but there is no chance that the L^∞ norm of the characteristic function of the hole be small.

We are going to develop a generalization of the Lagrangian framework. Among other variants, we follow the recent presentation of [14].

Proposition 6.3 *Let H be a Hilbert space and $\varepsilon_0 > 0$. For every $\varepsilon \in [0, \varepsilon_0]$ we consider :*

- *a bilinear form $a_\varepsilon(\cdot,\cdot)$ on H,*
- *a linear form $l_\varepsilon(\cdot)$ on H,*
- *a direct state $u_\varepsilon \in H$ solution of*

$$a_\varepsilon(u_\varepsilon, \varphi) = l_\varepsilon(\varphi) \qquad \forall \varphi \in H,$$

- *a cost function $J_\varepsilon(\cdot)$ continuously Fréchet differentiable on H,*
- *an adjoint state $v_\varepsilon \in H$ solution of*

$$a_\varepsilon(\varphi, v_\varepsilon) = -\int_0^1 dJ_\varepsilon(tu_\varepsilon + (1-t)u_0)\varphi dt \qquad \forall \varphi \in H.$$

Then we have for all $\varepsilon \in [0, \varepsilon_0]$

$$J_\varepsilon(u_\varepsilon) - J_0(u_0) = (\mathcal{L}_\varepsilon - \mathcal{L}_0)(u_0, v_\varepsilon),$$

with the Lagrangian

$$\mathcal{L}_\varepsilon(u, v) = J_\varepsilon(u) + a_\varepsilon(u, v) - l_\varepsilon(v) \qquad \forall(\varepsilon, u, v) \in [0, \varepsilon_0] \times H \times H.$$

Proof. We have the easy equalities:

$$\begin{aligned} J_\varepsilon(u_\varepsilon) - J_0(u_0) &= \mathcal{L}_\varepsilon(u_\varepsilon, v_\varepsilon) - \mathcal{L}_0(u_0, v_\varepsilon) \\ &= \mathcal{L}_\varepsilon(u_\varepsilon, v_\varepsilon) - \mathcal{L}_\varepsilon(u_0, v_\varepsilon) + \mathcal{L}_\varepsilon(u_0, v_\varepsilon) - \mathcal{L}_0(u_0, v_\varepsilon) \\ &= J_\varepsilon(u_\varepsilon) + a_\varepsilon(u_\varepsilon, v_\varepsilon) - l_\varepsilon(v_\varepsilon) - J_\varepsilon(u_0) - a_\varepsilon(u_0, v_\varepsilon) + l_\varepsilon(v_\varepsilon) + \mathcal{L}_\varepsilon(u_0, v_\varepsilon) - \mathcal{L}_0(u_0, v_\varepsilon) \\ &= J_\varepsilon(u_\varepsilon) - J_\varepsilon(u_0) + a_\varepsilon(u_\varepsilon - u_0, v_\varepsilon) + \mathcal{L}_\varepsilon(u_0, v_\varepsilon) - \mathcal{L}_0(u_0, v_\varepsilon) \\ &= J_\varepsilon(u_\varepsilon) - J_\varepsilon(u_0) - \int_0^1 dJ_\varepsilon(tu_\varepsilon + (1-t)u_0)(u_\varepsilon - u_0)dt + \mathcal{L}_\varepsilon(u_0, v_\varepsilon) - \mathcal{L}_0(u_0, v_\varepsilon). \end{aligned}$$

The first three terms cancel out, leading to the claim. □

We stress that the variation of the Lagrangian needs to be evaluated at the variable adjoint state v_ε. We will see that for the analysis of topology perturbations, approximating v_ε by v_0 may lead to an error of dominant order (see subsection 6.3.4).

6.3 Inclusion and Neumann cases

6.3.1 Problem formulation

Let Ω be an open and bounded subset of $\mathbb{R}^N$ and ω be a bounded, smooth open subset of $\mathbb{R}^N$. We consider a point $z \in \Omega$ and, for $\varepsilon \geq 0$ small enough, the set

$$\omega_\varepsilon = z + \varepsilon\omega \subset \Omega.$$

We will denote $\Omega_\varepsilon = \Omega \setminus \overline{\omega_\varepsilon}$.

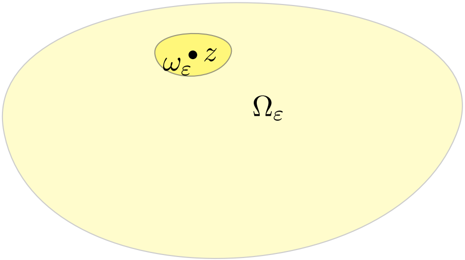


Figure 6.1: Domain perturbation.

We address the two-phase conductivity problem

$$\left\{\begin{array}{l} u_\varepsilon \in H_0^1(\Omega) \\ \displaystyle\int_\Omega \sigma_\varepsilon \nabla u_\varepsilon \cdot \nabla\varphi dx = \int_\Omega f\varphi dx \qquad \forall\varphi \in H_0^1(\Omega) \end{array}\right. \tag{6.2}$$

with $f \in L^2(\Omega)$ and the piecewise constant conductivity

$$\sigma_\varepsilon = \chi_{\Omega_\varepsilon}\alpha + \chi_{\omega_\varepsilon}\beta.$$

We will always assume that $\alpha > 0$ but for β we will investigate two cases:

- $\beta > 0$, called the inclusion case,
- $\beta = 0$, called the Neumann case.

For simplicity and to allow a simultaneous treatment of the two cases we assume that $f = 0$ in a neighborhood z, and that ε is small enough so that $f = 0$ in ω_ε.

The inclusion case obviously admits a unique solution, and it is associated with the strong form

$$\left\{\begin{array}{ll} -\operatorname{div}(\sigma_\varepsilon \nabla u_\varepsilon) = f & \text{in } \Omega \\ u_\varepsilon = 0 & \text{on } \partial\Omega. \end{array}\right. \tag{6.3}$$

The Neumann case can be reformulated as

$$\left\{\begin{array}{l} u_\varepsilon \in H_0^1(\Omega) \\ \displaystyle\int_{\Omega_\varepsilon} \alpha \nabla u_\varepsilon \cdot \nabla\varphi dx = \int_{\Omega_\varepsilon} f\varphi dx \qquad \forall\varphi \in H_0^1(\Omega). \end{array}\right. \tag{6.4}$$

Since every function of $H^1(\Omega_\varepsilon)$ can be extended to a function of $H^1(\Omega)$, we recognize that $u_{\varepsilon|\Omega_\varepsilon}$ is the weak solution of

$$\left\{\begin{array}{ll} -\alpha\Delta u_\varepsilon = f & \text{in } \Omega_\varepsilon \\ \dfrac{\partial u_\varepsilon}{\partial n} = 0 & \text{on } \partial\omega_\varepsilon \\ u_\varepsilon = 0 & \text{on } \partial\Omega. \end{array}\right. \tag{6.5}$$

In this case u_ε is undefined inside ω_ε. The "Neumann" terminology, of course, refers to the boundary condition on the inclusion, which is here a hole. The boundary condition on $\partial\Omega$ plays no role in our study.

For $\varepsilon = 0$, (6.2) gives the unperturbed problem

$$\left\{\begin{array}{ll} -\alpha\Delta u_0 = f & \text{in } \Omega \\ u_0 = 0 & \text{on } \partial\Omega. \end{array}\right. \tag{6.6}$$

To use the notation of Proposition 6.3 we set

$$a_\varepsilon(u, v) = \int_\Omega \sigma_\varepsilon \nabla u \cdot \nabla v dx, \qquad l_\varepsilon(v) = \int_\Omega f v dx \qquad \forall u, v \in H_0^1(\Omega).$$

In the subsequent analysis, for the sake of clarity, we will restrict ourselves to the inclusion case. We will briefly discuss the Neumann case afterwards.

6.3.2 Variation of the direct state

As stated in Proposition 6.3, the crucial point of the topological sensitivity analysis is that the variation of the Lagrangian has to be estimated when it is evaluated at a variable adjoint state v_ε. Moreover, this variable adjoint state is defined by means of the variable direct state u_ε. Therefore, our first step is to analyze the behavior of u_ε. We adapt the approach of [14].

Set $\tilde{u}_\varepsilon = u_\varepsilon - u_0$. Substracting the variational formulations for u_ε and u_0 results in

$$\int_\Omega \sigma_\varepsilon \nabla \tilde{u}_\varepsilon \cdot \nabla \varphi dx = (\alpha - \beta) \int_{\omega_\varepsilon} \nabla u_0 \cdot \nabla \varphi dx \qquad \forall \varphi \in H^1_0(\Omega).$$

We now define

$$U_\varepsilon(y) = \frac{1}{\varepsilon} \tilde{u}_\varepsilon(z + \varepsilon y), \qquad y \in \varepsilon^{-1}(\Omega - z),$$

so that

$$\tilde{u}_\varepsilon(x) = \varepsilon U_\varepsilon(\frac{x-z}{\varepsilon}), \qquad \nabla \tilde{u}_\varepsilon(x) = \nabla U_\varepsilon(\frac{x-z}{\varepsilon}) \qquad \forall x \in \Omega.$$

A straightforward change of variables leads to

$$\int_{\varepsilon^{-1}(\Omega - z)} \sigma_\varepsilon(z+\varepsilon y) \nabla U_\varepsilon(y) \cdot \nabla \varphi(z+\varepsilon y) dy = (\alpha - \beta) \int_\omega \nabla u_0(z+\varepsilon y) \cdot \nabla \varphi(z + \varepsilon y) dy \qquad \forall \varphi \in H^1_0(\Omega).$$

Choosing test functions as

$$\varphi(x) = \varepsilon \Phi(\frac{x-z}{\varepsilon}), \qquad \Phi \in H^1_0(\varepsilon^{-1}(\Omega - z))$$

yields

$$\int_{\varepsilon^{-1}(\Omega - z)} \sigma_\varepsilon(z+\varepsilon y) \nabla U_\varepsilon(y) \cdot \nabla \Phi(y) dy = (\alpha - \beta) \int_\omega \nabla u_0(z+\varepsilon y) \cdot \nabla \Phi(y) dy \qquad \forall \Phi \in H^1_0(\varepsilon^{-1}(\Omega - z)). \tag{6.7}$$

Let us define

$$\sigma(y) = \begin{cases} \beta \text{ if } y \in \omega \\ \alpha \text{ if } y \in \mathbb{R}^N \setminus \omega. \end{cases}$$

We also define the space (from the family of Beppo-Levi spaces)

$$X = \left\{ u \in L^2_{\text{loc}}(\mathbb{R}^N) : \nabla u \in L^2(\mathbb{R}^N) \right\}$$

and the associated quotient space $X/\mathbb{R}$ by the equivalence relation

$$u \sim v \Rightarrow \exists c \in \mathbb{R} \text{ s.t. } u - v = c.$$

Proposition 6.4 *The space $X/\mathbb{R}$ is a Hilbert space when it is equipped with the inner product*

$$\langle u, v \rangle_{X/\mathbb{R}} = \int_{\mathbb{R}^N} \nabla u \cdot \nabla v dx.$$

PROOF. The only point to check is that the space is complete. Let (u_n) be a Cauchy sequence of $X/\mathbb{R}$. Obviously, (∇u_n) is a Cauchy sequence of $L^2(\mathbb{R}^N)$. Hence there exists $g \in L^2(\mathbb{R}^N)$ such that $\nabla u_n \to g$ in $L^2(\mathbb{R}^N)$. Consider now a ball $B_k = B(0,k)$, $k \in \mathbb{N}^*$. We choose representatives such that $\int_{B_1} u_n dx = 0$. We know that $u_n \in H^1(B_k)$. By Theorem 1.36 there exists $c_k > 0$ such that $\|u_n\|_{L^2(B_k)} \leq c_k \|\nabla u_n\|_{L^2(B_k)}$. Hence (u_n) is a Cauchy sequence in $L^2(B_k)$. There exists $v_k \in L^2(B_k)$ such that $u_n \to v_k$ in $L^2(B_k)$. By uniqueness, we have $v_k(x) = v_l(x)$ for all $k \leq l$ and $x \in B_k$. For any $x \in \mathbb{R}^N$ we set $u(x) = v_k(x)$ for some k such that $x \in B_k$. By construction, $u \in L^2_{\text{loc}}(\mathbb{R}^N)$ and $u_n \to u$ in $L^2_{\text{loc}}(\mathbb{R}^N)$. Now, writing for all $\phi \in \mathcal{C}^1_c(\mathbb{R}^N)$

$$-\int_{\mathbb{R}^N} u_n \operatorname{div} \phi dx = \int_{\mathbb{R}^N} \nabla u_n \cdot \phi dx \to \int_{\mathbb{R}^N} g \cdot \phi dx$$

reveals that $g = \nabla u$. □

The function U_ε belongs to $H^1_0(\varepsilon^{-1}(\Omega - z))$. We implicitly consider an extension by 0 over $\mathbb{R}^N$. By the Lax-Milgram theorem there exists a unique $U \in X/\mathbb{R}$ solution of

$$\int_{\mathbb{R}^N} \sigma(y) \nabla U(y) \cdot \nabla \Phi(y) dy = (\alpha - \beta) \int_\omega \nabla u_0(z) \cdot \nabla \Phi(y) dy \qquad \forall \Phi \in X/\mathbb{R}. \tag{6.8}$$

Proposition 6.5 *We have the convergence $\nabla U_\varepsilon \to \nabla U$ in $L^2(\mathbb{R}^N)$ when $\varepsilon \searrow 0$, provided that ∇u_0 be continuous at point z.*

Proof. We implicitly work with a sequence $\varepsilon_n \searrow 0$.
Step 1. By the extension convention, $H^1_0(\varepsilon^{-1}(\Omega - z))/\mathbb{R}$ is a closed linear subspace of $X/\mathbb{R}$. We denote by P_ε the projection of U onto $H^1_0(\varepsilon^{-1}(\Omega - z))/\mathbb{R}$. By a small abuse of notation we will assume that P_ε stands for the representative in $H^1_0(\varepsilon^{-1}(\Omega - z))$. By definition we have

$$P_\varepsilon = \mathrm{argmin}_{\Phi \in H^1_0(\varepsilon^{-1}(\Omega - z))} \|\nabla\Phi - \nabla U\|_{L^2(\mathbb{R}^N)}.$$

Standard properties of the projection onto a linear subspace ensure that $\|\nabla P_\varepsilon\|_{L^2(\varepsilon^{-1}(\Omega-z))} \leq \|\nabla U\|_{L^2(\mathbb{R}^N)}$ and

$$\int_{\mathbb{R}^N} (\nabla P_\varepsilon - \nabla U)\cdot\nabla\Phi dx = 0 \qquad \forall \Phi \in H^1_0(\varepsilon^{-1}(\Omega - z)).$$

The first assertion yields that there exists $Q \in X/\mathbb{R}$ such that $\nabla P_\varepsilon \rightharpoonup \nabla Q$ weakly in $L^2(\mathbb{R}^N)$, up to a subsequence. The second assertion implies that

$$\int_{\mathbb{R}^N} (\nabla Q - \nabla U)\cdot\nabla\Phi dx = 0 \qquad \forall \Phi \in H^1_0(B(0,R)), \forall R > 0.$$

Let $\zeta : \mathbb{R}^N \to [0,1]$ be a smooth function such that $\zeta = 1$ in $B(0,1)$ and $\zeta = 0$ outside $B(0,2)$. Set $\zeta_n(x) = \zeta(x/n)$ and $\Phi_n(x) = (Q - U + \lambda_n)\zeta_n \in H^1_0(B(0,2n))$, with $\lambda_n \in \mathbb{R}$ at the moment arbitrary. This yields

$$\int_{\mathbb{R}^N} |\nabla Q - \nabla U|^2 \zeta_n dx + \int_{\mathbb{R}^N} (Q - U + \lambda_n)(\nabla Q - \nabla U)\cdot\nabla\zeta_n dx = 0. \tag{6.9}$$

A change of variables entails

$$\|(Q - U + \lambda_n)\nabla\zeta_n\|^2_{L^2(\mathbb{R}^N)} = n^{N-2}\int_{\mathbb{R}^N} |(Q - U + \lambda_n)(ny)\nabla\zeta(y)|^2 dy \leq cn^{N-2}\int_{R(0,1,2)} |(Q - U + \lambda_n)(ny)|^2 dy,$$

with $R(0,1,2)$ the ring centered at 0 with radii 1 and 2. We now fix λ_n such that

$$\int_{R(0,1,2)} (Q - U + \lambda_n)(ny) dy = 0.$$

By the Poincaré (Wirtinger) inequality from Theorem 1.36 we infer

$$\|(Q - U + \lambda_n)\nabla\zeta_n\|^2_{L^2(\mathbb{R}^N)} \leq cn^N \int_{R(0,1,2)} |\nabla(Q - U)(ny)|^2 dy \leq c\int_{R(0,n,2n)} |\nabla(Q - U)|^2 dx.$$

Plugging this into (6.9) and using the Cauchy-Schwarz inequality, we arrive at

$$\int_{\mathbb{R}^N} |\nabla Q - \nabla U|^2\zeta_n dx \leq c\|\nabla Q - \nabla U\|_{L^2(\mathbb{R}^N)} \left(\int_{R(0,n,2n)} |\nabla(Q - U)|^2 dx\right)^{1/2}.$$

Letting now n go to $+\infty$ results in $\nabla Q = \nabla U$. Therefore, by uniqueness of the cluster point, the whole sequence ∇P_ε weakly converges to ∇U. In particular we infer that

$$\int_{\mathbb{R}^N} \nabla P_\varepsilon \cdot \nabla U dx \to \int_{\mathbb{R}^N} |\nabla U|^2 dx.$$

From the identity

$$\|\nabla P_\varepsilon - \nabla U\|^2_{L^2(\mathbb{R}^N)} = \|\nabla P_\varepsilon\|^2_{L^2(\mathbb{R}^N)} + \|\nabla U\|^2_{L^2(\mathbb{R}^N)} - 2\int_{\mathbb{R}^N} \nabla P_\varepsilon \cdot \nabla U dx$$

we derive

$$\limsup_{\varepsilon\to 0}\|\nabla P_\varepsilon-\nabla U\|^2_{L^2(\mathbb{R}^N)}\leq\limsup_{\varepsilon\to 0}\|\nabla P_\varepsilon\|^2_{L^2(\mathbb{R}^N)}-\|\nabla U\|^2_{L^2(\mathbb{R}^N)}\leq 0.$$

We have shown that ∇P_ε strongly converges to ∇U in $L^2(\mathbb{R}^N)$.
Step 2. Using (6.7) we obtain

$$\begin{aligned}\int_{\varepsilon^{-1}(\Omega-z)}\sigma(y)\nabla(P_\varepsilon-U_\varepsilon)(y)\cdot\nabla\Phi(y)dy&=\int_{\varepsilon^{-1}(\Omega-z)}\sigma(y)\nabla P_\varepsilon(y)\cdot\nabla\Phi(y)dy\\&-(\alpha-\beta)\int_\omega\nabla u_0(z+\varepsilon y)\cdot\nabla\Phi(y)dy\qquad\forall\Phi\in H^1_0(\varepsilon^{-1}(\Omega-z)).\end{aligned}$$

In view of (6.8) this rewrites as

$$\begin{aligned}\int_{\varepsilon^{-1}(\Omega-z)}\sigma(y)\nabla(P_\varepsilon-U_\varepsilon)(y)\cdot\nabla\Phi(y)dy&=\int_{\varepsilon^{-1}(\Omega-z)}\sigma(y)\nabla(P_\varepsilon-U)(y)\cdot\nabla\Phi(y)dy\\&-(\alpha-\beta)\int_\omega(\nabla u_0(z+\varepsilon y)-\nabla u_0(z))\cdot\nabla\Phi(y)dy\qquad\forall\Phi\in H^1_0(\varepsilon^{-1}(\Omega-z)).\end{aligned}$$

Choose $\Phi=P_\varepsilon-U_\varepsilon$. We obtain for some constant c

$$\begin{aligned}\|\nabla(P_\varepsilon-U_\varepsilon)\|^2_{L^2(\varepsilon^{-1}(\Omega-z))}\leq c\big(&\|\nabla(P_\varepsilon-U)\|_{L^2(\varepsilon^{-1}(\Omega-z))}\|\nabla(P_\varepsilon-U_\varepsilon)\|_{L^2(\varepsilon^{-1}(\Omega-z))}\\&+\|\nabla u_0(z+\varepsilon y)-\nabla u_0(z)\|_{L^2(\omega)}\|\nabla(P_\varepsilon-U_\varepsilon)\|_{L^2(\varepsilon^{-1}(\Omega-z))}\big),\end{aligned}$$

leading to

$$\|\nabla(P_\varepsilon-U_\varepsilon)\|_{L^2(\varepsilon^{-1}(\Omega-z))}\leq c\big(\|\nabla(P_\varepsilon-U)\|_{L^2(\varepsilon^{-1}(\Omega-z))}+\|\nabla u_0(z+\varepsilon y)-\nabla u_0(z)\|_{L^2(\omega)}\big).$$

Using step 1 and the continuity assumption we arrive at $\|\nabla(P_\varepsilon-U_\varepsilon)\|_{L^2(\varepsilon^{-1}(\Omega-z))}\to 0$.
Step 3. The proof is completed by combining step 1 and step 2. □

Corollary 6.6 *Under the assumption of Proposition 6.5 we have*

$$\|\tilde{u}_\varepsilon\|^2_{H^1(\Omega)}=O(\varepsilon^N),$$

and for any $R>0$

$$\|\tilde{u}_\varepsilon\|^2_{H^1(\Omega\backslash B(z,R))}=o(\varepsilon^N).$$

Proof. By change of variables it is straightforward that

$$\|\nabla\tilde{u}_\varepsilon\|^2_{L^2(\Omega)}=\varepsilon^N\|\nabla U_\varepsilon\|^2_{L^2(\varepsilon^{-1}(\Omega-z))}.$$

Proposition 6.5 yields that $\|\nabla U_\varepsilon\|_{L^2(\varepsilon^{-1}(\Omega-z))}=O(1)$, whereby $\|\tilde{u}_\varepsilon\|^2_{H^1(\Omega)}=O(\varepsilon^N)$ by the Poincaré inequality.

Let now $R>0$. The same change of variables provides

$$\|\nabla\tilde{u}_\varepsilon\|^2_{L^2(\Omega\backslash B(z,R))}=\varepsilon^N\|\nabla U_\varepsilon\|^2_{L^2(\varepsilon^{-1}(\Omega-z))\backslash B(0,\varepsilon^{-1}R)}=\varepsilon^N\|\nabla U_\varepsilon\|^2_{L^2(\mathbb{R}^N\backslash B(0,\varepsilon^{-1}R)}.$$

This can be rephrased as

$$\|\nabla\tilde{u}_\varepsilon\|^2_{L^2(\Omega\backslash B(z,R))}=\varepsilon^N\int_{\mathbb{R}^N}(1-\chi_{B(0,\frac{R}{\varepsilon})}(y))|\nabla U_\varepsilon(y)|^2dy.$$

By Proposition 6.5, splitting and the dominated convergence theorem we infer $\|\nabla\tilde{u}_\varepsilon\|^2_{L^2(\Omega\backslash B(z,R))}=o(\varepsilon^N)$. The claim is achieved by the Poincaré inequality. □

6.3.3 Variation of the adjoint state

For simplicity we consider a cost function of the form

$$J_\varepsilon(u) = \hat{J}(u_{|\hat{\Omega}}), \tag{6.10}$$

where $\hat{\Omega}$ is an open subset of Ω excluding a neighborhood of z and $\hat{J} : H^1(\hat{\Omega}) \to \mathbb{R}$ is Fréchet differentiable. We further assume that $d\hat{J}$ is Lipschitz continuous.

In view of Proposition 6.3 we define the adjoint state $v_\varepsilon \in H_0^1(\Omega)$ solution of

$$\int_\Omega \sigma_\varepsilon \nabla v_\varepsilon \cdot \nabla \varphi dx = -\int_0^1 dJ_\varepsilon(tu_\varepsilon + (1-t)u_0)\varphi dt \qquad \forall \varphi \in H_0^1(\Omega).$$

In particular the unperturbed adjoint state v_0 satisfies

$$\int_\Omega \alpha \nabla v_0 \cdot \nabla \varphi dx = -dJ_0(u_0)\varphi \qquad \forall \varphi \in H_0^1(\Omega). \tag{6.11}$$

Set $\tilde{v}_\varepsilon = v_\varepsilon - v_0$. We obtain

$$\int_\Omega \sigma_\varepsilon \nabla \tilde{v}_\varepsilon \cdot \nabla \varphi dx = (\alpha - \beta)\int_{\omega_\varepsilon} \nabla v_0 \cdot \nabla \varphi dx - \int_0^1 dJ_\varepsilon(tu_\varepsilon + (1-t)u_0)\varphi dt + dJ_0(u_0)\varphi \qquad \forall \varphi \in H_0^1(\Omega).$$

By (6.10) this rewrites

$$\int_\Omega \sigma_\varepsilon \nabla \tilde{v}_\varepsilon \cdot \nabla \varphi dx = (\alpha - \beta)\int_{\omega_\varepsilon} \nabla v_0 \cdot \nabla \varphi dx - \int_0^1 \left(d\hat{J}((tu_\varepsilon + (1-t)u_0)_{|\hat{\Omega}}) - d\hat{J}(u_{0|\hat{\Omega}}) \right) \varphi_{|\hat{\Omega}} dt \qquad \forall \varphi \in H_0^1(\Omega).$$

We will later justify that the latter integral can be disregarded, therefore we define $w_\varepsilon \in H_0^1(\Omega)$ solution of

$$\int_\Omega \sigma_\varepsilon \nabla w_\varepsilon \cdot \nabla \varphi dx = (\alpha - \beta)\int_{\omega_\varepsilon} \nabla v_0 \cdot \nabla \varphi dx \qquad \forall \varphi \in H_0^1(\Omega).$$

In order to approximate this w_ε we proceed exactly as for the direct state. We define

$$W_\varepsilon(y) = \frac{1}{\varepsilon} w_\varepsilon(z + \varepsilon y), \qquad y \in \varepsilon^{-1}(\Omega - z),$$

and $W \in X/\mathbb{R}$ solution of

$$\int_{\mathbb{R}^N} \sigma(y) \nabla W(y) \cdot \nabla \Phi(y) dy = (\alpha - \beta)\int_\omega \nabla v_0(z) \cdot \nabla \Phi(y) dy \qquad \forall \Phi \in X/\mathbb{R}. \tag{6.12}$$

We obtain:

Proposition 6.7 *It holds $\nabla W_\varepsilon \to \nabla W$ in $L^2(\mathbb{R}^N)$ when $\varepsilon \searrow 0$, provided that ∇v_0 be continuous at point z.*

We now analyze the approximation of $\tilde{v}_\varepsilon$ by w_ε.

Lemma 6.8 *It holds*

$$\|\tilde{v}_\varepsilon - w_\varepsilon\|^2_{H^1(\Omega)} = o(\varepsilon^N).$$

PROOF. Set $e_\varepsilon = \tilde{v}_\varepsilon - w_\varepsilon$. It solves

$$\int_\Omega \sigma_\varepsilon \nabla e_\varepsilon \cdot \nabla \varphi dx = -\int_0^1 \left(d\hat{J}((tu_\varepsilon + (1-t)u_0)_{|\hat{\Omega}}) - d\hat{J}(u_{0|\hat{\Omega}}) \right) \varphi_{|\hat{\Omega}} dt \qquad \forall \varphi \in H_0^1(\Omega).$$

We choose $\varphi = e_\varepsilon$. Using that $d\hat{J}$ is Lipschitz we obtain

$$\|\nabla e_\varepsilon\|^2_{L^2(\Omega)} \leq c \int_0^1 t \|u_\varepsilon - u_0\|_{H^1(\hat{\Omega})} \|e_\varepsilon\|_{H^1(\hat{\Omega})} dt.$$

Corollary 6.6 and the Poincaré inequality provide the desired estimate. □

6.3.4 Variation of the Lagrangian

According to Proposition 6.3 we define the Lagrangian

$$\mathcal{L}_\varepsilon(u,v) = J_\varepsilon(u) + \int_\Omega \sigma_\varepsilon \nabla u \cdot \nabla v dx - \int_\Omega f v dx. \qquad \forall u, v \in H_0^1(\Omega).$$

We are interested in the variation

$$(\mathcal{L}_\varepsilon - \mathcal{L}_0)(u_0, v_\varepsilon) = (\beta - \alpha) \int_{\omega_\varepsilon} \nabla u_0 \cdot \nabla v_\varepsilon dx.$$

We decompose as

$$(\mathcal{L}_\varepsilon - \mathcal{L}_0)(u_0, v_\varepsilon) = (\beta - \alpha) \int_{\omega_\varepsilon} \nabla u_0 \cdot \nabla v_0 dx + (\beta - \alpha) \int_{\omega_\varepsilon} \nabla u_0 \cdot \nabla \tilde{v}_\varepsilon dx.$$

We now show up w_ε:

$$(\mathcal{L}_\varepsilon - \mathcal{L}_0)(u_0, v_\varepsilon) = (\beta - \alpha) \int_{\omega_\varepsilon} \nabla u_0 \cdot \nabla v_0 dx + (\beta - \alpha) \int_{\omega_\varepsilon} \nabla u_0 \cdot \nabla w_\varepsilon dx + (\beta - \alpha) \int_{\omega_\varepsilon} \nabla u_0 \cdot (\nabla \tilde{v}_\varepsilon - \nabla w_\varepsilon) dx.$$

Lemma 6.9 *If ∇u_0 and ∇v_0 are continuous at z then*

$$(\mathcal{L}_\varepsilon - \mathcal{L}_0)(u_0, v_\varepsilon) = \varepsilon^N (\beta - \alpha) |\omega| \nabla u_0(z) \cdot \nabla v_0(z) + (\beta - \alpha) \int_{\omega_\varepsilon} \nabla u_0 \cdot \nabla w_\varepsilon dx + o(\varepsilon^N).$$

Proof. We first estimate

$$\begin{aligned} \int_{\omega_\varepsilon} \nabla u_0 \cdot \nabla v_0 dx - \varepsilon^N |\omega| \nabla u_0(z) \cdot \nabla v_0(z) &= \int_{\omega_\varepsilon} \left(\nabla u_0 \cdot \nabla v_0 - \nabla u_0(z) \cdot \nabla v_0(z) \right) dx \\ &= \varepsilon^N \int_\omega \left(\nabla u_0(z + \varepsilon y) \cdot \nabla v_0(z + \varepsilon y) - \nabla u_0(z) \cdot \nabla v_0(z) \right) dy \\ &= o(\varepsilon^N). \end{aligned}$$

Secondly, the Cauchy-Schwarz inequality yields

$$\left| \int_{\omega_\varepsilon} \nabla u_0 \cdot (\nabla \tilde{v}_\varepsilon - \nabla w_\varepsilon) dx \right| \leq \|\nabla u_0\|_{L^2(\omega_\varepsilon)} \|\nabla \tilde{v}_\varepsilon - \nabla w_\varepsilon\|_{L^2(\omega_\varepsilon)} = O(\varepsilon^{N/2}) o(\varepsilon^{N/2}),$$

by Lemma 6.8. □

From the expression found in Lemma 6.9 we make a change of variables to obtain

$$(\mathcal{L}_\varepsilon - \mathcal{L}_0)(u_0, v_\varepsilon) = \varepsilon^N (\beta - \alpha) |\omega| \nabla u_0(z) \cdot \nabla v_0(z) + \varepsilon^N (\beta - \alpha) \int_\omega \nabla u_0(z + \varepsilon y) \cdot \nabla W_\varepsilon(y) dy + o(\varepsilon^N).$$

Lemma 6.10 *If ∇u_0 and ∇v_0 are continuous at z then*

$$(\mathcal{L}_\varepsilon - \mathcal{L}_0)(u_0, v_\varepsilon) = \varepsilon^N (\beta - \alpha) |\omega| \nabla u_0(z) \cdot \nabla v_0(z) + \varepsilon^N (\beta - \alpha) \int_\omega \nabla u_0(z) \cdot \nabla W(y) dy + o(\varepsilon^N).$$

Proof. We have to show that

$$\lim_{\varepsilon \to 0} \int_\omega \left(\nabla u_0(z + \varepsilon y) \cdot \nabla W_\varepsilon(y) - \nabla u_0(z) \cdot \nabla W(y) \right) dy \ = 0.$$

It is an immediate consequence of Proposition 6.7, using

$$\begin{aligned} &\int_\omega \left(\nabla u_0(z + \varepsilon y) \cdot \nabla W_\varepsilon(y) - \nabla u_0(z) \cdot \nabla W(y) \right) dy \\ &\qquad = \int_\omega (\nabla u_0(z + \varepsilon y) - \nabla u_0(z)) \cdot \nabla W_\varepsilon(y) dy + \int_\omega \nabla u_0(z) \cdot (\nabla W_\varepsilon(y) - \nabla W(y)) dy \end{aligned}$$

and the Cauchy-Schwarz inequality. □

We are now in position to derive the topological asymptotic expansion from Proposition 6.3:

$$J_\varepsilon(u_\varepsilon) - J_0(u_0) = \varepsilon^N (\beta - \alpha) \left(|\omega| \nabla u_0(z) \cdot \nabla v_0(z) + \int_\omega \nabla u_0(z) \cdot \nabla W(y) dy \right) + o(\varepsilon^N).$$

In order to arrive at a closed formula a last concept is missing: the polarization matrix.

6.3.5 Polarization matrix

The definition (6.12) of W shows that W depends linearly on $\nabla v_0(z)$. More precisely, denote by $(e_1, \cdots, e_N)$ the canonical basis of $\mathbb{R}^N$ and let $\zeta_i \in X/\mathbb{R}$ be the solution of

$$\int_{\mathbb{R}^N} \sigma(y)\nabla\zeta_i(y) \cdot \nabla\Phi(y)dy = (\alpha - \beta)\int_\omega e_i \cdot \nabla\Phi(y)dy \qquad \forall\Phi \in X/\mathbb{R}. \tag{6.13}$$

Then it holds $W(y) = \nabla v_0(z) \cdot \zeta(y)$. It follows that

$$\int_\omega \nabla u_0(z) \cdot \nabla W(y)dy = \nabla u_0(z) \cdot \left(\int_\omega D\zeta(y)^\top dy\right) \nabla v_0(z).$$

Yet, choosing $\Phi = \zeta_j$ in (6.13) yields

$$(\alpha - \beta)\int_\omega e_i \cdot \nabla\zeta_j(y)dy = \int_{\mathbb{R}^N} \sigma(y)\nabla\zeta_i(y) \cdot \nabla\zeta_j(y)dy = (\alpha - \beta)\int_\omega e_j \cdot \nabla\zeta_i(y)dy.$$

This shows that

$$Q := \int_\omega D\zeta(y)dy = Q^\top.$$

We arrive at

$$|\omega|\nabla u_0(z) \cdot \nabla v_0(z) + \int_\omega \nabla u_0(z) \cdot \nabla W(y)dy = \nabla u_0(z) \cdot (|\omega|I + Q)\nabla v_0(z).$$

Definition 6.11 *We call polarization matrix the symmetric matrix*

$$\mathcal{P} = \left(\frac{\beta}{\alpha} - 1\right)(|\omega|I + Q).$$

Note that (6.13) can be equivalently rewritten as

$$\int_{\mathbb{R}^N} \sigma(y)\nabla\zeta_i(y) \cdot \nabla\Phi(y)dy = (\alpha - \beta)\int_{\partial\omega} e_i \cdot n(y)\Phi(y)dy \qquad \forall\Phi \in X/\mathbb{R}. \tag{6.14}$$

Therefore the corresponding strong form is

$$\begin{cases} \Delta\zeta_i = 0 \text{ in } \mathbb{R}^N \setminus \partial\omega \\ \beta\left(\dfrac{\partial\zeta_i}{\partial n}\right)_{\text{int}} - \alpha\left(\dfrac{\partial\zeta_i}{\partial n}\right)_{\text{ext}} = (\alpha - \beta)e_i \cdot n \text{ on } \partial\omega. \end{cases} \tag{6.15}$$

Let us now give an additional property of the polarization matrix.

Proposition 6.12 *The eigenvalues (λ_i) of the polarization matrix satisfy the inequality*

$$\lambda_i \leq \left(\frac{\beta}{\alpha} - 1\right)|\omega|.$$

Moreover the polarization matrix is

- *symmetric positive definite if $\beta > \alpha$,*
- *symmetric negative definite if $\beta < \alpha$.*

Proof. Since $\mathcal{P}$ is symmetric, let us choose an orthogonal basis in which it is diagonal.
Upper bound on eigenvalues. Choosing $\Phi = \zeta_i$ in (6.13) entails

$$(\alpha - \beta)\int_\omega e_i \cdot \nabla\zeta_i(y)dy \geq 0.$$

We infer that

$$\mathcal{P}e_i \cdot e_i = \left(\frac{\beta}{\alpha} - 1\right)\left(|\omega| + \int_\omega e_i \cdot \nabla\zeta_i(y)dy\right) \leq \left(\frac{\beta}{\alpha} - 1\right)|\omega|.$$

Case $\beta < \alpha$*.* The above inequality directly shows that $\mathcal{P}e_i \cdot e_i < 0$, hence $\mathcal{P}$ is symmetric negative definite.
Case $\beta > \alpha$*.* We write

$$\mathcal{P}e_i \cdot e_i = \left(\frac{\beta}{\alpha} - 1\right)\int_\omega (\nabla\zeta_i + e_i) \cdot e_i dy = \left(\frac{\beta}{\alpha} - 1\right)\int_\omega \left(|\nabla\zeta_i + e_i|^2 - (\nabla\zeta_i + e_i) \cdot \nabla\zeta_i\right) dy.$$

Using (6.13) we obtain

$$(\alpha - \beta)\int_\omega (\nabla\zeta_i + e_i) \cdot \nabla\zeta_i dy = (\alpha - \beta)\int_\omega |\nabla\zeta_i|^2 dy + \int_{\mathbb{R}^N} \sigma|\nabla\zeta_i|^2 dy = \alpha\int_{\mathbb{R}^N} |\nabla\zeta_i|^2 dy.$$

This yields

$$\mathcal{P}e_i \cdot e_i = \left(\frac{\beta}{\alpha} - 1\right)\int_\omega |\nabla\zeta_i + e_i|^2 dy + \int_{\mathbb{R}^N} |\nabla\zeta_i|^2 dy > 0,$$

hence $\mathcal{P}$ is symmetric positive definite. □

6.3.6 Expression of the topological asymptotic expansion

Let us summarize.

Theorem 6.13 *Consider a cost function of form* (6.10)*. Let* $v_0 \in H_0^1(\Omega)$ *be the solution of* (6.11)*. Suppose that* ∇u_0 *and* ∇v_0 *are continuous at* z*. Then*

$$\boxed{J_\varepsilon(u_\varepsilon) - J_0(u_0) = \varepsilon^N \alpha\nabla u_0(z) \cdot \mathcal{P}\nabla v_0(z) + o(\varepsilon^N),}$$

where $\mathcal{P}$ *is the polarization matrix.*

Therefore we can set the topological derivative of a shape functional $\mathcal{J}$ such that $\mathcal{J}(\Omega_\varepsilon) = J_\varepsilon(u_\varepsilon)$ as

$$d_T\mathcal{J}(\Omega, \omega, z) = \alpha\nabla u_0(z) \cdot \mathcal{P}\nabla v_0(z).$$

Remark 6.14 *The regularity assumption for* u_0 *and* v_0 *is actually redundant, since with the assumptions made it is ensured by elliptic regularity. For more general problems it is nevertheless a requirement.*

6.3.7 Ball-shaped inclusions

We now aim at computing the polarization matrix in the special case where ω is the unit ball $B(0,1)$. The main step is to solve (6.15). The typical method is to "guess" a candidate solution with some parameters, and to plug this candidate in the system to fix the parameters. The form of the guess is inspired from the fundamental solution of the operator, here the Laplacian, and therefore depends on the dimension.

2D case

The exterior solution is constructed on the basis of the partial derivative $\frac{\partial}{\partial y_i}$ of the fundamental solution $E(y) = \frac{-1}{2\pi}\log|y|$. We find the solution

$$\zeta_i(y) = \frac{\alpha - \beta}{\alpha + \beta} \times \left\{ \begin{array}{l} e_i \cdot y \text{ in } \omega \\ \dfrac{e_i \cdot y}{|y|^2} \text{ in } \mathbb{R}^2 \setminus \overline{\omega}, \end{array} \right.$$

with gradient

$$\nabla\zeta_i(y) = \frac{\alpha-\beta}{\alpha+\beta} \times \begin{cases} e_i \text{ in } \omega \\ \left(\dfrac{e_i}{|y|^2} - 2(e_i \cdot y)\dfrac{y}{|y|^4}\right) \text{ in } \mathbb{R}^2 \setminus \overline{\omega}. \end{cases}$$

This entails

$$Q = \pi\frac{\alpha-\beta}{\alpha+\beta}I, \qquad \boxed{\mathcal{P} = 2\pi\frac{\beta-\alpha}{\beta+\alpha}I.}$$

3D case

The procedure is similar but more tedious. We find

$$\boxed{\mathcal{P} = 4\pi\frac{\beta-\alpha}{\beta+2\alpha}I.}$$

6.3.8 Note on the Neumann case

The Neumann case ($\beta = 0$) can be analyzed along the same lines as the inclusion case. Only minor modifications have to be performed. Typically, volume integrals within the inclusion have to be replaced by boundary integrals on the boundary of the hole with the help of the Green formula. The space X is replaced by

$$X_\omega = \left\{u \in L^2_{\text{loc}}(\mathbb{R}^N \setminus \omega) : \nabla u \in L^2(\mathbb{R}^N \setminus \bar{\omega})\right\}.$$

In order to use the semi-norm as a norm on the quotient space $X_\omega/\mathbb{R}$, it is required that $\mathbb{R}^N \setminus \omega$ be connected. This is in principle not a very strong assumption, except in dimension 1! Actually, it is clear that in dimension 1 the cost function is likely to be discontinuous at $\varepsilon = 0$. This singularity can also be seen a posteriori by computing the polarization matrix (in fact just a number) for an inclusion and let β go to 0: it diverges. This is left as exercise.

Eventually, in dimension $N \geq 2$ with $\mathbb{R}^N \setminus \omega$ connected, we arrive at the same result as in Theorem 6.13 with the polarization matrix computed with $\beta = 0$. This gives for the ball in dimensions 2 and 3

$$\mathcal{P} = -2\pi I.$$

For various extensions of the the topological derivative including the linear elasticity framework and the cration of cracks we refer e.g. to [4, 18, 3]. One of the difficulties is the computation of the polarization matrix (or tensor in the vector cases), which may be fairly technical.

6.3.9 Example

Consider again the compliance, for which $v_0 = -u_0$. In view of Theorem 6.13 and Proposition 6.12, introducing a weak inclusion (or a Neumann hole) increases the compliance. Introducing a strong inclusion decreases the compliance. That was to be expected! The topological derivative gives an information on the best place to make such a perturbation.

6.4 Dirichlet case

6.4.1 Problem formulation

We modify the setting of section 6.3 as follows. Let us first underline that the 2D and 3D cases have to be addressed in significantly different manners. We restrict ourselves here to the 2D case. A brief comment on the 3D case will made in the end.

Let Ω be an open and bounded subset of $\mathbb{R}^2$ and ω be a bounded, smooth open subset of $\mathbb{R}^2$. We consider a point $z \in \Omega$ and, for $\varepsilon \geq 0$ small enough, the "hole"

$$\omega_\varepsilon = z + \varepsilon\omega \subset \Omega.$$

We assume for convenience (but without loss of generality) that $0 \in \omega \subset\subset B(0,1)$. We again denote $\Omega_\varepsilon = \Omega \setminus \overline{\omega_\varepsilon}$.

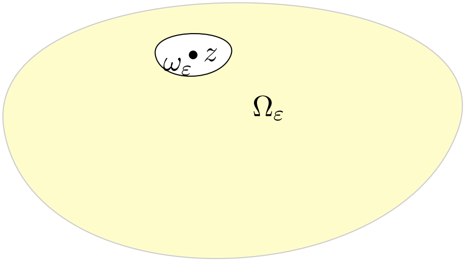


Figure 6.2: Perforated domain.

We address the problem

$$\begin{cases} -\Delta u_\varepsilon = f \text{ in } \Omega_\varepsilon \\ u_\varepsilon = 0 \text{ on } \partial\Omega_\varepsilon. \end{cases} \tag{6.16}$$

It is assumed that $f \in L^2(\Omega)$ with $f = 0$ in a neighborhood of z. We will implicitly suppose that ε is small enough so that $f = 0$ in ω_ε. Again, the important aspect is the boundary condition on $\partial\omega_\varepsilon$. We denote by u_0 the unperturbed state solution of

$$\begin{cases} -\Delta u_0 = f \text{ in } \Omega \\ u_0 = 0 \text{ on } \partial\Omega. \end{cases} \tag{6.17}$$

In order to develop the adjoint method in a fixed space we extend u_ε by 0 inside ω_ε. We set

$$a_\varepsilon(u,v) = \int_\Omega \nabla u \cdot \nabla v dx \qquad \forall u, v \in H_0^1(\Omega),$$

$$l_\varepsilon(v) = \int_\Omega \nabla u_\varepsilon \cdot \nabla v dx = \int_{\Omega_\varepsilon} \nabla u_\varepsilon \cdot \nabla v dx = \int_{\Omega_\varepsilon} f v dx - \int_{\partial\omega_\varepsilon} \frac{\partial u_\varepsilon}{\partial n} v ds \qquad \forall v \in H_0^1(\Omega).$$

By convention the normal to $\partial\omega_\varepsilon$ is chosen outward to ω_ε. This construction ensures that $u_\varepsilon \in H_0^1(\Omega)$ satisfies

$$a_\varepsilon(u_\varepsilon, v) = l_\varepsilon(v) \qquad \forall v \in H_0^1(\Omega).$$

6.4.2 A preliminary estimate

We introduce the weighted Sobolev space

$$W(\mathbb{R}^2) = \left\{ u \in L^2_{\text{loc}}(\mathbb{R}^2) : wu \in L^2(\mathbb{R}^2), \nabla u \in L^2(\mathbb{R}^2) \right\},$$

with the weight function

$$w(x) = \frac{1}{(1+|x|)\log(2+|x|)}.$$

It is easily shown to be a Hilbert space for the inner product

$$\langle u, v \rangle_{W(\mathbb{R}^2)} = \int_{\mathbb{R}^2} (w^2 uv + \nabla u \cdot \nabla v) dx.$$

We also define

$$W_0(\mathbb{R}^2 \setminus \overline{\omega}) = \left\{ u \in W(\mathbb{R}^2 \setminus \overline{\omega}) : \gamma_0 u = 0 \text{ on } \partial\omega \right\}.$$

We have the Poincaré inequality:

Proposition 6.15 *There exists $c > 0$ such that*

$$\|u\|_{W(\mathbb{R}^2\setminus\overline{\omega})} \leq c\|\nabla u\|_{L^2(\mathbb{R}^2\setminus\overline{\omega})} \qquad \forall u \in W_0(\mathbb{R}^2 \setminus \overline{\omega}).$$

Proof. Step 1. Consider first a function $u \in \mathcal{C}_c^\infty(\mathbb{R}^2 \setminus \overline{B}(0,a))$, $a > 1$. For an arbitrary unit vector e we set $f(r) = u(re)$. Integration by parts yields

$$\int_a^{+\infty} \frac{1}{r\log^2 r} f(r)^2 dr = \int_a^{+\infty} \frac{2}{\log r} f(r)f'(r)dr,$$

whereby we obtain by the Cauchy-Schwarz inequality

$$\int_a^{+\infty} \frac{1}{r\log^2 r} f(r)^2 dr \leq 2\left(\int_a^{+\infty} \frac{1}{r\log^2 r} f(r)^2 dr\right)^{1/2} \left(\int_a^{+\infty} rf'(r)^2 dr\right)^{1/2}.$$

This implies

$$\int_a^{+\infty} \frac{1}{r\log^2 r} f(r)^2 dr \leq 4\int_a^{+\infty} rf'(r)^2 dr.$$

With the help of polar coordinates, this shows that

$$\|u\|_{W(\mathbb{R}^2\setminus\overline{B}(0,a))} \leq \sqrt{5}\|\nabla u\|_{L^2(\mathbb{R}^2\setminus\overline{B}(0,a))}.$$

By a density argument, this holds for all $u \in W_0(\mathbb{R}^2 \setminus \overline{B}(0,a))$.
Step 2. Let now $u \in W_0(\mathbb{R}^2 \setminus \overline{\omega})$, and $a > 1$. Let $\theta \in \mathcal{C}_c^\infty(\mathbb{R}^2)$ such that $\theta = 1$ in $B(0,2a)$ and $\theta = 0$ outside $B(0,3a)$. By step 1 we have

$$\|(1-\theta)u\|_{W(\mathbb{R}^2\setminus\overline{\omega})} \leq c\|(1-\theta)\nabla u - u\nabla\theta\|_{L^2(\mathbb{R}^2\setminus\overline{\omega})} \leq c(\|\nabla u\|_{L^2(\mathbb{R}^2\setminus\overline{\omega})} + \|u\|_{L^2(B(0,3a)\setminus\overline{\omega})}).$$

The Poincaré inequality in $\{v \in H^1(B(0,3a) \setminus \overline{\omega})) : \gamma_0 v = 0 \text{ on } \partial\omega\}$ permits to conclude. □

Lemma 6.16 *Let $\psi \in H^{1/2}(\partial\omega)$, $\psi_\varepsilon(x) = \psi((x-z)/\varepsilon)$ and $w_\varepsilon \in H^1(\Omega_\varepsilon)$ be the solution of*

$$\begin{cases} -\Delta w_\varepsilon = 0 \text{ in } \Omega_\varepsilon \\ w_\varepsilon = 0 \text{ on } \partial\Omega \\ w_\varepsilon = \psi_\varepsilon \text{ on } \partial\omega_\varepsilon. \end{cases}$$

Let $R > 0$ such that $B(z,R) \subset \Omega$. There exists a constant $c > 0$ independent of ε and ψ such that, for ε small enough,

$$\|w_\varepsilon\|_{H^1(\Omega_\varepsilon)} \leq c\|\psi\|_{H^{1/2}(\partial\omega)},$$

$$\|w_\varepsilon\|_{H^1(\Omega\setminus B(z,R)} \leq \frac{c}{\sqrt{-\log\varepsilon}}\|\psi\|_{H^{1/2}(\partial\omega)}.$$

Proof. We assume for convenience of notation that $z = 0$.
Step 1. We denote by $\Psi \in H^1(B(0,1) \setminus \bar{\omega})$ a function such that $\gamma_0\Psi = \psi$ on $\partial\omega$ and $\gamma_0\Psi = 0$ on $\partial B(0,1)$, obtained from standard lifting, then extended by 0 outside $B(0,1)$. We set $\Psi_\varepsilon(x) = \Psi(x/\varepsilon)$ and $\tilde{w}_\varepsilon = w_\varepsilon - \Psi_\varepsilon$. We have from the weak formulation

$$\int_{\Omega_\varepsilon} \nabla w_\varepsilon \cdot \nabla\tilde{w}_\varepsilon dx = 0,$$

whereby

$$\|\nabla\tilde{w}_\varepsilon\|_{L^2(\Omega_\varepsilon)}^2 = -\int_{\Omega_\varepsilon} \nabla\Psi_\varepsilon \cdot \nabla\tilde{w}_\varepsilon dx.$$

This entails $\|\nabla\tilde{w}_\varepsilon\|_{L^2(\Omega_\varepsilon)} \leq \|\nabla\Psi_\varepsilon\|_{L^2(\Omega_\varepsilon)}$ and subsequently $\|\nabla w_\varepsilon\|_{L^2(\Omega_\varepsilon)} \leq 2\|\nabla\Psi_\varepsilon\|_{L^2(\Omega_\varepsilon)}$. We infer by change of variables

$$\|\nabla w_\varepsilon\|_{L^2(\Omega_\varepsilon)} \leq 2\|\nabla\Psi\|_{L^2(B(0,1)\setminus\omega)} \leq c\|\psi\|_{H^{1/2}(\partial\omega)}.$$

We can also lift ψ inside ω by a function $\tilde{\psi}$, and setting $\tilde{\psi}_\varepsilon(x) = \tilde{\psi}(x/\varepsilon)$ we get

$$\|\nabla\tilde{\psi}_\varepsilon\|_{L^2(\omega_\varepsilon)} = \|\nabla\tilde{\psi}\|_{L^2(\omega)} \leq c\|\psi\|_{H^{1/2}(\partial\omega)}.$$

Extending w_ε by $\tilde{\psi}_\varepsilon$ in ω_ε and applying the Poincaré inequality in $H^1_0(\Omega)$ yields

$$\|w_\varepsilon\|_{H^1(\Omega_\varepsilon)} \leq c\|\psi\|_{H^{1/2}(\partial\omega)}.$$

Step 2. We first focus on the problem

$$\begin{cases} -\Delta w_\varepsilon = 0 \text{ in } \Omega_\varepsilon \\ w_\varepsilon = 0 \text{ on } \partial\Omega \\ w_\varepsilon = \bar{\psi} \text{ on } \partial\omega_\varepsilon, \end{cases}$$

with $\bar{\psi} \in \mathbb{R}$ constant. We use the primal variational principle (see section 1.3.5):

$$\|\nabla w_\varepsilon\|^2_{L^2(\Omega_\varepsilon)} \leq \|\nabla v_\varepsilon\|^2_{L^2(\Omega_\varepsilon)}$$

for any $v_\varepsilon \in H^1(\Omega_\varepsilon)$ such that $\gamma_0 v_\varepsilon = 0$ on $\partial\Omega$ and $\gamma_0 v_\varepsilon = \bar{\psi}$ on $\partial\omega_\varepsilon$. We choose the following one, for some $\rho > 0$ such that $B(0,\rho) \subset \Omega$:

$$v_\varepsilon(x) = \begin{cases} \bar{\psi} & \text{if } |x| \leq \varepsilon \\ \bar{\psi}\dfrac{\log|x| - \log\rho}{\log\varepsilon - \log\rho} & \text{if } \varepsilon \leq |x| \leq \rho \\ 0 & \text{if } |x| \geq \rho. \end{cases}$$

This yields

$$\nabla v_\varepsilon(x) = \begin{cases} 0 & \text{if } |x| \leq \varepsilon \\ \dfrac{\bar{\psi}}{\log\varepsilon - \log\rho}\dfrac{x}{|x|^2} & \text{if } \varepsilon \leq |x| \leq \rho \\ 0 & \text{if } |x| \geq \rho, \end{cases}$$

thus

$$\|\nabla v_\varepsilon\|^2_{L^2(\Omega_\varepsilon)} = \left(\frac{\bar{\psi}}{\log\varepsilon - \log\rho}\right)^2 \int_\varepsilon^\rho \frac{1}{r^2} 2\pi r dr = 2\pi\frac{\overline{\psi}^2}{\log\rho - \log\varepsilon}.$$

It follows that

$$\|\nabla w_\varepsilon\|_{L^2(\Omega_\varepsilon)} \leq \left(\frac{2\pi}{\log\rho - \log\varepsilon}\right)^{1/2} |\bar{\psi}|.$$

The Poincaré inequality yields for ε small enough

$$\|w_\varepsilon\|_{H^1(\Omega_\varepsilon)} \leq \frac{c}{\sqrt{-\log\varepsilon}}|\bar{\psi}|.$$

Step 3. We turn to the general case. By lifting, Proposition 6.15 and the Lax-Milgram theorem, there exists a unique $S \in W(\mathbb{R}^2 \setminus \overline{\omega})$ such that $\gamma_0 S = \psi$ on $\partial\omega$ and

$$\int_{\mathbb{R}^2\setminus\bar{\omega}} \nabla S \cdot \nabla\Phi dx = 0 \qquad \forall\Phi \in W_0(\mathbb{R}^2 \setminus \overline{\omega}). \tag{6.18}$$

Obviously it holds $-\Delta S = 0$ in $\mathbb{R}^2 \setminus \bar{\omega}$ in the sense of distributions. Let ζ be a smooth function equal to 0 in $B(0,1)$ and 1 outside $B(0,2)$. Set $\hat{S} = \zeta S$ and

$$G = -\Delta\hat{S} = -\Delta\zeta S - 2\nabla\zeta \cdot \nabla S. \tag{6.19}$$

By construction G is supported in the ring $R(0,1,2)$, and it is smooth by elliptic regularity for S. Let now $\xi : \mathbb{R}^2 \to \mathbb{R}$ be a smooth function equal to 1 in $B(0,2)$ and 0 outside $B(0,3)$ and set $\xi_\rho = \xi(x/\rho)$, $\rho > 1$. The Green formula yields

$$\int_{\mathbb{R}^2} G dx = \int_{\mathbb{R}^2} G\xi_\rho dx = \int_{\mathbb{R}^2} \nabla\hat{S} \cdot \nabla\xi_\rho dx = \int_{\mathbb{R}^2\setminus\bar{B}(0,2\rho)} \nabla\hat{S} \cdot \nabla\xi_\rho dx.$$

Applying the Cauchy-Schwarz inequality, using $\nabla\hat{S} \in L^2(\mathbb{R}^2)$, a change of variables, and letting ρ go to $+\infty$ results in

$$\int_{\mathbb{R}^2} G dx = 0. \tag{6.20}$$

We have for all $\Phi \in W(\mathbb{R}^2)$, using (6.19)

$$\begin{aligned}
\int_{\mathbb{R}^2} G\Phi dx &= \int_{\mathbb{R}^2} \nabla\zeta \cdot \nabla(S\Phi) dx - 2\int_{\mathbb{R}^2} \nabla\zeta \cdot \nabla S \Phi dx \\
&= \int_{\mathbb{R}^2} S\nabla\zeta \cdot \nabla\Phi dx - \int_{\mathbb{R}^2} \nabla\zeta \cdot \nabla S\Phi dx \\
&= \int_{\mathbb{R}^2} \nabla\hat{S} \cdot \nabla\Phi dx - \int_{\mathbb{R}^2} \zeta\nabla S \cdot \nabla\Phi dx - \int_{\mathbb{R}^2} \nabla\zeta \cdot \nabla S\Phi dx \\
&= \int_{\mathbb{R}^2} \nabla\hat{S} \cdot \nabla\Phi dx - \int_{\mathbb{R}^2} \nabla S \cdot \nabla(\zeta\Phi) dx.
\end{aligned}$$

By (6.18) the latter integral vanishes, resulting in

$$\int_{\mathbb{R}^2} \nabla\hat{S} \cdot \nabla\Phi dx = \int_{\mathbb{R}^2} G\Phi dx \qquad \forall\Phi \in W(\mathbb{R}^2). \tag{6.21}$$

Let

$$E(y) = \frac{-1}{2\pi} \log|y|$$

be the fundamental solution of the Laplacian and $\hat{S}_0 = G * E$. Since G is smooth and compactly supported and $E \in L^1_{\text{loc}}(\mathbb{R}^2)$, it follows that $\hat{S}_0$ is smooth. Using (6.20) we obtain the expressions

$$\hat{S}_0(x) = \int_{\mathbb{R}^2} G(y)\left(E(x-y) - E(x)\right) dy \qquad \forall x \neq 0,$$

$$\nabla\hat{S}_0(x) = \int_{\mathbb{R}^2} G(y)\left(\nabla E(x-y) - \nabla E(x)\right) dy \qquad \forall x \neq 0.$$

From the mean value theorem we infer that $|\hat{S}_0(x)| \leq c/|x|$ and $|\nabla\hat{S}_0(x)| \leq c/|x|^2$, implying that $\hat{S}_0 \in W(\mathbb{R}^2)$. Let $\Phi \in W(\mathbb{R}^2)$. We have

$$\begin{aligned}
\int_{\mathbb{R}^2} \nabla\hat{S}_0 \cdot \nabla\Phi dx = \lim_{\rho\to+\infty} \int_{\mathbb{R}^2} \xi_\rho \nabla\hat{S}_0 \cdot \nabla\Phi dx &= \lim_{\rho\to+\infty} \left(\int_{\mathbb{R}^2} \nabla\hat{S}_0 \cdot \nabla(\xi_\rho\Phi) dx - \int_{\mathbb{R}^2} \nabla\hat{S}_0 \cdot \nabla\xi_\rho \Phi dx \right) \\
&= \int_{\mathbb{R}^2} G\Phi dx - \lim_{\rho\to+\infty} \int_{\mathbb{R}^2} \nabla\hat{S}_0 \cdot \nabla\xi_\rho \Phi dx = \int_{\mathbb{R}^2} G\Phi dx,
\end{aligned}$$

using the decay properties of $\hat{S}_0$ and Φ. Comparing with (6.21), choosing $\Phi = \hat{S} - \hat{S}_0$, we obtain that $\hat{S} = \hat{S}_0 + \lambda$, for some $\lambda \in \mathbb{R}$. In particular we have the expression

$$\lambda = -\hat{S}_0(0),$$

showing that $|\lambda| \leq c\|\psi\|_{H^{1/2}(\partial\omega)}$. Denoting $S_0 = S - \lambda$ we have $S_0 = \hat{S} - \lambda = \hat{S}_0$ in $\mathbb{R}^2 \setminus B(0,2)$, i.e.,

$$S_0(x) = \int_{B(0,2)} G(y)E(x-y) dy \qquad \forall x \in \mathbb{R}^2 \setminus B(0,2).$$

We now set

$$s_\varepsilon(x) = S_0(\frac{x}{\varepsilon}).$$

Using again (6.20) we get

$$s_\varepsilon(x) = \int_{B(0,2)} G(y)\left(E(\frac{x}{\varepsilon} - y) - E(\frac{x}{\varepsilon})\right) dy \qquad \forall x \in \mathbb{R}^2 \setminus B(0,2\varepsilon).$$

The particular form of the fundamental solution leads to

$$s_\varepsilon(x) = \int_{B(0,2)} G(y)\,(E(x-\varepsilon y) - E(x))\,dy \qquad \forall x \in \mathbb{R}^2 \setminus B(0,2\varepsilon).$$

The mean value theorem easily shows that

$$\|s_\varepsilon\|_{H^1(\Omega\setminus B(0,R))} \le c\varepsilon \|G\|_{L^2(\mathbb{R}^2)} \le c\varepsilon \|\psi\|_{H^{1/2}(\partial\omega)}.$$

We note that on $\partial\omega_\varepsilon$ we have $s_\varepsilon(x) = S_0(x/\varepsilon) = S(x/\varepsilon) - \lambda = \psi(x/\varepsilon) - \lambda$. We now define $r_\varepsilon = w_\varepsilon - s_\varepsilon$, solution of

$$\begin{cases} -\Delta r_\varepsilon = 0 \text{ in } \Omega_\varepsilon \\ r_\varepsilon = -s_\varepsilon \text{ on } \partial\Omega \\ r_\varepsilon = \lambda \text{ on } \partial\omega_\varepsilon. \end{cases}$$

We have by step 2 and a standard decomposition

$$\|r_\varepsilon\|_{H^1(\Omega_\varepsilon)} \le c\|s_\varepsilon\|_{H^{1/2}(\partial\Omega)} + c\frac{|\lambda|}{\sqrt{-\log\varepsilon}} \le \frac{c}{\sqrt{-\log\varepsilon}}\|\psi\|_{H^{1/2}(\partial\omega)}.$$

This completes the proof by $w_\varepsilon = s_\varepsilon + r_\varepsilon$. □

6.4.3 Variation of the direct state

Set $\tilde{u}_\varepsilon = u_\varepsilon - u_0$. It solves

$$\begin{cases} -\Delta \tilde{u}_\varepsilon = 0 \text{ in } \Omega_\varepsilon \\ \tilde{u}_\varepsilon = 0 \text{ on } \partial\Omega \\ \tilde{u}_\varepsilon = -u_0 \text{ on } \partial\omega_\varepsilon. \end{cases} \tag{6.22}$$

We define $h_\varepsilon \in \mathcal{C}^\infty(\mathbb{R}^2 \setminus \{z\})$ and $r_\varepsilon \in H^1(\Omega)$ by

$$h_\varepsilon(x) = -\frac{\log|x-z|}{\log\varepsilon} u_0(z), \qquad \begin{cases} -\Delta r_\varepsilon = 0 \text{ in } \Omega \\ r_\varepsilon = -h_\varepsilon \text{ on } \partial\Omega. \end{cases}$$

It is left to the reader to check, using polar coordinates, that $\Delta h_\varepsilon = 0$ in $\mathbb{R}^2 \setminus \{z\}$. We now set $e_\varepsilon = \tilde{u}_\varepsilon - h_\varepsilon - r_\varepsilon$. We have

$$\begin{cases} -\Delta e_\varepsilon = 0 \text{ in } \Omega_\varepsilon \\ e_\varepsilon = 0 \text{ on } \partial\Omega \\ e_\varepsilon = -u_0 - h_\varepsilon - r_\varepsilon \text{ on } \partial\omega_\varepsilon. \end{cases} \tag{6.23}$$

Lemma 6.17 *If u_0 is $\mathcal{C}^1$ is a neighborhood of z then*

$$\|r_\varepsilon\|_{H^1(\Omega)} = O((-\log\varepsilon)^{-1}),$$

$$\|e_\varepsilon\|_{H^1(\Omega\setminus B(z,R))} = O((-\log\varepsilon)^{-3/2}).$$

PROOF. Step 1. The first estimate is obvious since $\|h_\varepsilon\|_{H^{1/2}(\partial\Omega)} = O((-\log\varepsilon)^{-1})$ by construction.
Step 2. Set

$$\psi_\varepsilon(x) = (-u_0 - h_\varepsilon - r_\varepsilon)(z + \varepsilon x).$$

We decompose as

$$\psi_\varepsilon(x) = \underbrace{[u_0(z) - u_0(z+\varepsilon x)]}_{p_\varepsilon(x)} - \underbrace{[u_0(z) + h_\varepsilon(z+\varepsilon x)]}_{q_\varepsilon(x)} - \underbrace{r_\varepsilon(z+\varepsilon x)}_{\hat{r}_\varepsilon(x)}.$$

By regularity of u_0 we have immediately $\|p_\varepsilon\|_{H^{1/2}(\partial\omega)} = O(\varepsilon)$. Next, from

$$q_\varepsilon(x) = u_0(z)\left(1 - \frac{\log|\varepsilon x|}{\log\varepsilon}\right) = -u_0(z)\frac{\log|x|}{\log\varepsilon},$$

we get $\|q_\varepsilon\|_{H^{1/2}(\partial\omega)} = O((-\log\varepsilon)^{-1})$. Lastly, a change of variables yields

$$\|\hat{r}_\varepsilon\|_{H^1(\omega)} \leq \|\nabla r_\varepsilon\|_{L^2(\omega_\varepsilon)} + \varepsilon^{-1}\|r_\varepsilon\|_{L^2(\omega_\varepsilon)} \leq \|\nabla r_\varepsilon\|_{L^2(\omega_\varepsilon)} + c\|r_\varepsilon\|_{L^\infty(\omega_\varepsilon)}.$$

By elliptic regularity we have $\|\nabla r_\varepsilon\|_{L^2(\omega_\varepsilon)}+\|r_\varepsilon\|_{L^\infty(\omega_\varepsilon)} \leq c\|h_\varepsilon\|_{H^{1/2}(\partial\Omega)}$, thus $\|\hat{r}_\varepsilon\|_{H^{1/2}(\partial\omega)} = O((-\log\varepsilon)^{-1})$. We conclude using Lemma 6.16. □

We infer from Lemma 6.17 and the triangle inequality:

Lemma 6.18 *If u_0 is $\mathcal{C}^1$ is a neighborhood of z then*

$$\|\tilde{u}_\varepsilon\|_{H^1(\Omega\setminus B(z,R)} = O((-\log\varepsilon)^{-1}).$$

6.4.4 Variation of the adjoint state

We again consider a cost function of the form

$$J_\varepsilon(u) = \hat{J}(u_{|\hat{\Omega}}), \tag{6.24}$$

where $\hat{\Omega}$ is an open subset of Ω excluding a neighborhood of z and $\hat{J} : H^1(\hat{\Omega}) \to \mathbb{R}$ is Fréchet differentiable. We further assume that $d\hat{J}$ is Lipschitz continuous. In view of Proposition 6.3 we define the adjoint state $v_\varepsilon \in H^1_0(\Omega)$ solution of

$$\int_\Omega \nabla v_\varepsilon \cdot \nabla\varphi dx = -\int_0^1 dJ_\varepsilon(tu_\varepsilon + (1-t)u_0)\varphi dt \qquad \forall\varphi \in H^1_0(\Omega).$$

In particular the unperturbed adjoint state v_0 satisfies

$$\int_\Omega \nabla v_0 \cdot \nabla\varphi dx = -dJ_0(u_0)\varphi \qquad \forall\varphi \in H^1_0(\Omega). \tag{6.25}$$

Lemma 6.19 *If u_0 is $\mathcal{C}^1$ is a neighborhood of z then*

$$\|v_\varepsilon - v_0\|_{H^1(\Omega)} = O((-\log\varepsilon)^{-1}).$$

Proof. Set $\tilde{v}_\varepsilon = v_\varepsilon - v_0$. We have

$$\int_\Omega \nabla\tilde{v}_\varepsilon \cdot \nabla\varphi dx = \int_0^1 (dJ_0(u_0) - dJ_\varepsilon(tu_\varepsilon + (1-t)u_0))\varphi dt \qquad \forall\varphi \in H^1_0(\Omega),$$

leading to

$$\int_\Omega \nabla\tilde{v}_\varepsilon \cdot \nabla\varphi dx = \int_0^1 (d\hat{J}(u_{0|\hat{\Omega}}) - d\hat{J}((tu_\varepsilon + (1-t)u_0)_{|\hat{\Omega}})\varphi_{|\hat{\Omega}} dt \qquad \forall\varphi \in H^1_0(\Omega).$$

Choosing $\varphi = \tilde{v}_\varepsilon$ and using that $d\hat{J}$ is Lipschitz yields

$$\|\tilde{v}_\varepsilon\|_{H^1(\Omega)} \leq c\|\tilde{u}_\varepsilon\|_{H^1(\hat{\Omega})}.$$

The conclusion follows from Lemma 6.18. □

6.4.5 Variation of the Lagrangian

We define the standard Lagrangian

$$\mathcal{L}_\varepsilon(u,v) = J_\varepsilon(u) + \int_\Omega \nabla u \cdot \nabla v dx - \int_\Omega fv + \int_{\partial\omega_\varepsilon} \frac{\partial u_\varepsilon}{\partial n} v ds.$$

This provides the variation

$$(\mathcal{L}_\varepsilon - \mathcal{L}_0)(u_0, v_\varepsilon) = \int_{\partial\omega_\varepsilon} \frac{\partial u_\varepsilon}{\partial n} v_\varepsilon ds.$$

Lemma 6.20 *If u_0, v_0 are of class $\mathcal{C}^1$ in a neighborhood of z then*

$$(\mathcal{L}_\varepsilon - \mathcal{L}_0)(u_0, v_\varepsilon) = \int_{\partial\omega_\varepsilon} \frac{\partial \tilde{u}_\varepsilon}{\partial n} v_0 ds + O\left((-\log\varepsilon)^{-2}\right).$$

Proof. We work with the decomposition

$$\begin{aligned}(\mathcal{L}_\varepsilon - \mathcal{L}_0)(u_0, v_\varepsilon) - \int_{\partial\omega_\varepsilon} \frac{\partial \tilde{u}_\varepsilon}{\partial n} v_0 ds &= \int_{\partial\omega_\varepsilon} \frac{\partial u_0}{\partial n} v_\varepsilon ds + \int_{\partial\omega_\varepsilon} \frac{\partial \tilde{u}_\varepsilon}{\partial n} (v_\varepsilon - v_0) ds \\ &= \int_{\omega_\varepsilon} \nabla u_0 \cdot \nabla v_\varepsilon dx - \int_{\Omega_\varepsilon} \nabla \tilde{u}_\varepsilon \cdot \nabla (v_\varepsilon - v_0) dx.\end{aligned}$$

Extending $\tilde{u}_\varepsilon$ by $-u_0$ in ω_ε allows to write

$$\begin{aligned}(\mathcal{L}_\varepsilon - \mathcal{L}_0)(u_0, v_\varepsilon) - \int_{\partial\omega_\varepsilon} \frac{\partial \tilde{u}_\varepsilon}{\partial n} v_0 ds &= -\int_{\Omega} \nabla \tilde{u}_\varepsilon \cdot \nabla (v_\varepsilon - v_0) dx + \int_{\omega_\varepsilon} \nabla u_0 \cdot \nabla v_0 dx \\ &= \int_0^1 (d\hat{J}((tu_\varepsilon + (1-t)u_0)_{|\hat{\Omega}}) - d\hat{J}(u_{0|\hat{\Omega}}))(\tilde{u}_\varepsilon)_{|\hat{\Omega}} dt \\ &\quad + \int_{\omega_\varepsilon} \nabla u_0 \cdot \nabla v_0 dx,\end{aligned}$$

where the last equality is obtained as in Lemma 6.19. We conclude using Lemma 6.18. □

Lemma 6.21 *If u_0, v_0 are of class $\mathcal{C}^1$ in a neighborhood of z then*

$$(\mathcal{L}_\varepsilon - \mathcal{L}_0)(u_0, v_\varepsilon) = \frac{2\pi}{-\log\varepsilon} u_0(z) v_0(z) + O\left((-\log\varepsilon)^{-3/2}\right).$$

Proof. We decompose the expression found in Lemma 6.20 as

$$\begin{aligned}(\mathcal{L}_\varepsilon - \mathcal{L}_0)(u_0, v_\varepsilon) &= \int_{\partial\omega_\varepsilon} \frac{\partial h_\varepsilon}{\partial n} v_0 ds + \int_{\partial\omega_\varepsilon} \frac{\partial r_\varepsilon}{\partial n} v_0 ds + \int_{\partial\omega_\varepsilon} \frac{\partial e_\varepsilon}{\partial n} v_0 ds + O\left((-\log\varepsilon)^{-2}\right) \\ &= \int_{\partial\omega_\varepsilon} \frac{\partial h_\varepsilon}{\partial n} v_0(z) ds + \int_{\partial\omega_\varepsilon} \frac{\partial h_\varepsilon}{\partial n} (v_0 - v_0(z)) ds \\ &\quad + \int_{\omega_\varepsilon} \nabla r_\varepsilon \cdot \nabla v_0 dx - \int_{\Omega_\varepsilon} \nabla e_\varepsilon \cdot \nabla v_0 dx + O\left((-\log\varepsilon)^{-2}\right).\end{aligned}$$

Let $\rho > 0$ such that $B(0,\rho) \subset \omega$. We define

$$\tilde{h}_\varepsilon(x) = \begin{cases} h_\varepsilon(x) & \text{if } |x - z| \geq \rho\varepsilon \\ -\left(\dfrac{\log\rho}{\log\varepsilon} + 1\right) u_0(z) & \text{if } |x - z| \leq \rho\varepsilon. \end{cases}$$

This truncation ensures that $\tilde{h}_\varepsilon \in H^1(\Omega)$. We extend e_ε by $-u_0 - \tilde{h}_\varepsilon - r_\varepsilon$ in ω_ε to write

$$\begin{aligned}(\mathcal{L}_\varepsilon - \mathcal{L}_0)(u_0, v_\varepsilon) &= \int_{\partial\omega_\varepsilon} \frac{\partial h_\varepsilon}{\partial n} v_0(z) ds + \int_{\partial\omega_\varepsilon} \frac{\partial h_\varepsilon}{\partial n} (v_0 - v_0(z)) ds \\ &\quad - \int_{\Omega} \nabla e_\varepsilon \cdot \nabla v_0 dx - \int_{\omega_\varepsilon} \nabla(\tilde{h}_\varepsilon + u_0) \cdot \nabla v_0 dx + O\left((-\log\varepsilon)^{-2}\right) \\ &= \int_{\partial\omega_\varepsilon} \frac{\partial h_\varepsilon}{\partial n} v_0(z) ds + \int_{\partial\omega_\varepsilon} \frac{\partial h_\varepsilon}{\partial n} (v_0 - v_0(z)) ds \\ &\quad + d\hat{J}(u_{0|\hat{\Omega}}) e_{\varepsilon|\hat{\Omega}} - \int_{\omega_\varepsilon} \nabla(\tilde{h}_\varepsilon + u_0) \cdot \nabla v_0 dx + O\left((-\log\varepsilon)^{-2}\right).\end{aligned}$$

By the definition of h_ε, Lemma 6.17 and the smoothness assumptions we arrive at

$$(\mathcal{L}_\varepsilon - \mathcal{L}_0)(u_0, v_\varepsilon) = \int_{\partial\omega_\varepsilon} \frac{\partial h_\varepsilon}{\partial n} v_0(z) ds + O\left((-\log\varepsilon)^{-3/2}\right).$$

The Green formula yields

$$\int_{\partial\omega_\varepsilon} \frac{\partial h_\varepsilon}{\partial n} ds = \int_{\partial B(z,\varepsilon)} \frac{\partial h_\varepsilon}{\partial n} ds = \frac{-2\pi}{\log \varepsilon} u_0(z),$$

which completes the proof. □

6.4.6 Expression of the topological asymptotic expansion

Theorem 6.22 *Consider a cost function of form* (6.24)*. Let $v_0 \in H^1_0(\Omega)$ be the solution of* (6.25)*. Suppose that u_0 and v_0 are of class $\mathcal{C}^1$ in a neighborhood of z. Then*

$$\boxed{J_\varepsilon(u_\varepsilon) - J_0(u_0) = \frac{2\pi}{-\log \varepsilon} u_0(z) v_0(z) + o\left(\frac{1}{-\log \varepsilon}\right).}$$

Remark 6.23 *As for the inclusion case the regularity assumption for u_0 and v_0 is actually redundant, due to elliptic regularity, but has to be kept in mind for more general situations.*

We observe that this expression does not depend on the shape of the hole. This is typical of the 2D Dirichlet problem and related to the notion of capacity. We infer the topological derivative

$$d_T \mathcal{J}(\Omega, \omega, z) = u_0(z) v_0(z).$$

Another important observation is about the speed of convergence: it is very slow (see Fig. 6.3). This is due to the fact that making a small Dirichlet hole has a drastic effect on the solution.

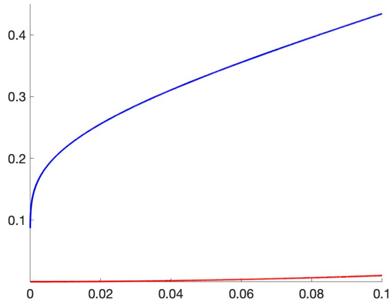


Figure 6.3: Comparison of the functions $f_1(\varepsilon) = \frac{1}{-\log \varepsilon}$ (in blue) and $f_2(\varepsilon) = \varepsilon^2$ (in red).

Remark 6.24 *In 3D the capacity of ω depends on the shape of ω. The dominant term of the topological asymptotic expansion has the form $\varepsilon C(\omega) u_0(z) v_0(z)$. In case of vector problems, like in elasticity, we have a capacity matrix.*

6.4.7 Example

For the compliance the topological derivative is always negative. Interpreting the Dirichlet condition as a clamped condition, this is logical.

6.5 Application

The topological derivative can be used in several ways to address topology optimization problems. Roughly speaking, we can distinguish three types of approaches.

6.5.1 One-shot approaches

The simplest way to use the topological derivative is just to compute it and represent its map. This provides a decision helping tool to topology perturbation.

An example is shown in figure 6.4. The context is that of linear elasticity for compliance minimization. We consider the creation of a Dirichlet hole, which can be interpreted as a bolt. The initial configuration has already two bolts and the question is: where to put a third one? It can be shown [15, 4] that the topological asymptotic expansion takes the form

$$J_\varepsilon(u_\varepsilon) - J_0(u_0) = \frac{k}{-\log\varepsilon} u_0(z)\cdot v_0(z) + o\left(\frac{1}{-\log\varepsilon}\right)$$

for some $k > 0$ depending on material parameters.

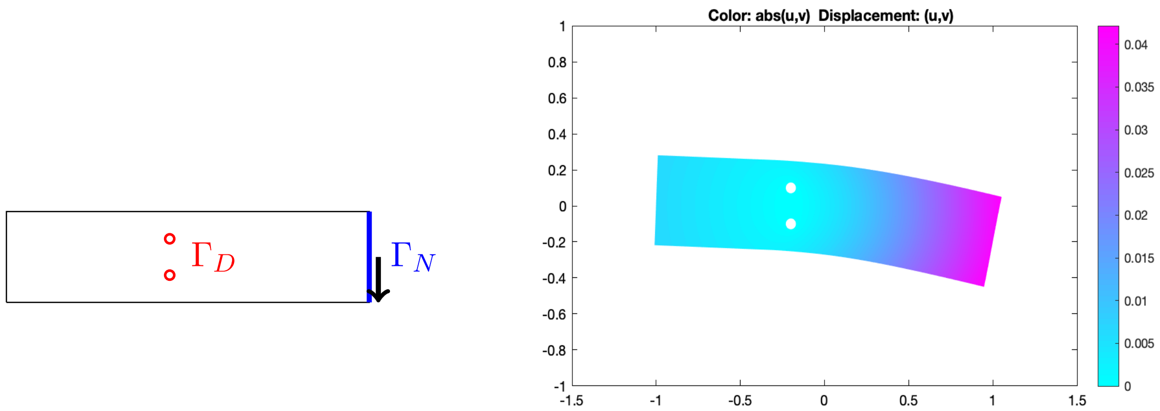


Figure 6.4: Optimal positioning of a bolt in linear elasticity: computational domain (left) and norm of the displacement field u_0 on the deformed configuration (right). The topological derivative is proportional to $-|u_0|^2$, hence the best locations are of course on the right side, if possible, but the left side is also relevant.

This kind of approach can be applied to detection problems. In such cases it is classical to consider a least square cost function of form $\|u - u_m\|^2_{L^2(\Omega_m)}$, where u_m is the measurement and Ω_m is the measurement location (it can also be at the boundary). An example is shown in figure 6.5. Here the defects are Neumann holes (air bubbles) in an elastic domain (elastodynamics). Since the shape of the defects is unknown the topological derivative is computed for balls. The sensors are located in the bottom edge. Here the measurements are synthetic (i.e. simulated).

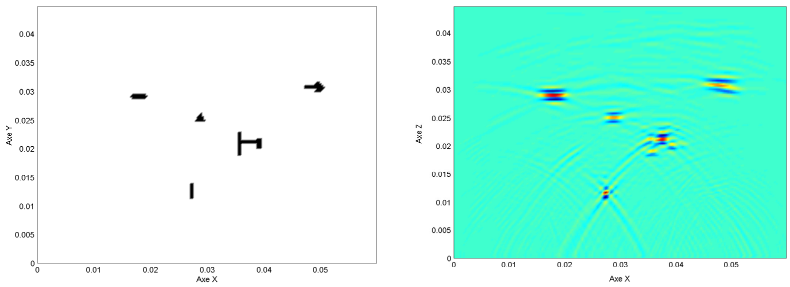
6.5.2 Iterative topology modifications

A very natural approach is to iteratively create small holes at points where the topological derivative is the most negative. Within an integrated shape and topology optimization procedure, this can be easily combined with geometry optimization based on the shape derivative, in an alternating fashion.

6.5.3 Solving optimality conditions

In the absence of constraint, an obvious necessary condition of optimality for Ω is

$$d_T\mathcal{J}(\Omega,\omega,z) \geq 0 \qquad \forall z\in\Omega,\ \forall\omega.$$



(a) (b)

Figure 6.5: Detection of defects. (a) Actual defects, (b) Map of the topological derivative.

Since we most often content ourselves with local minimizers, it is reasonable to fix ω, typically the ball, and define optimality with respect to such perturbations.

An elementary thresholding algorithm

A first algorithm, sometimes called "hard-kill", can be formulated as the iteration

$$\Omega_{k+1} = \{x \in \Omega_k : d_T\mathcal{J}(\Omega, \omega, x) \geq -\lambda_k\},$$

where (λ_k) is a sequence of positive numbers going to 0 (ideally...). This thresholding has to be thoroughly monitored, for instance by choosing the λ_k's in order to remove a given or adaptatively chosen volume fraction at each iteration. This kind of algorithm has shown its efficiency (see e.g. [15]), but it suffers from some drawbacks:

- the iterations consist in material removal only;
- there is no guarranty of descent or of convergence, hence rough stopping criteria are mostly used;
- geometrical optimality conditions are not taken into account.

Therefore, the obtained shapes are likely to be sub-optimal.

Interpolation methods

Another approach is to use the topological derivative in order to design interpolation methods. Consider the two-phase conductivity problem

$$\begin{cases} -\operatorname{div}(\sigma_\Omega \nabla u_\Omega) = f & \text{in } D \\ u_\Omega = 0 & \text{on } \partial D, \end{cases} \tag{6.26}$$

$$\sigma_\Omega = \chi_\Omega \sigma_1 + (1-\chi_\Omega)\sigma_0, \qquad \sigma_0, \sigma_1 > 0,$$

for the penalized compliance (for instance)

$$\mathcal{J}(\Omega) = \int_D f u_\Omega dx + \ell|\Omega|.$$

We place ourselves in two space dimensions to fix ideas. Recall the topological derivative for a circular inclusion in Ω (we choose here $f(\varepsilon) = \pi\varepsilon^2 = |\omega_\varepsilon|$)

$$d_T\mathcal{J}(\Omega, \omega, z) = -2\sigma_1 \frac{\sigma_0 - \sigma_1}{\sigma_0 + \sigma_1} |\nabla u_\Omega(z)|^2 - \ell, \qquad z \in \Omega.$$

It is also relevant to create an inclusion in $D \setminus \bar{\Omega}$, for which the topological derivative is

$$d_T\mathcal{J}(\Omega, \omega, z) = -2\sigma_0 \frac{\sigma_1 - \sigma_0}{\sigma_1 + \sigma_0} |\nabla u_\Omega(z)|^2 + \ell, \qquad z \in D \setminus \bar{\Omega}.$$

We intend to solve the optimality conditions

$$d_T\mathcal{J}(\Omega,\omega,z) \geq 0 \qquad \forall z \in \Omega \cup D \setminus \bar{\Omega}. \tag{6.27}$$

We now associate with (6.26) the interpolated problem

$$\begin{cases} -\operatorname{div}(\bar{\sigma}(\rho)\nabla \bar{u}_\rho) = f & \text{in } D \\ \bar{u}_\rho = 0 & \text{on } \partial D, \end{cases} \tag{6.28}$$

$$\bar{\mathcal{J}}(\rho) = \int_D f\bar{u}_\rho dx + \ell \int_D \rho dx.$$

Here, the pseudo-density ρ is sought within the convex set $L^\infty(D,[0,1])$. The interpolation profile $\bar{\sigma} : [0,1] \to \mathbb{R}_+^*$ is supposed to satisfy $\bar{\sigma}(0) = \sigma_0$ and $\bar{\sigma}(1) = \sigma_1$. Therefore we have

$$\rho = \chi_\Omega \Rightarrow \bar{\mathcal{J}}(\rho) = \mathcal{J}(\Omega).$$

We further assume that $\bar{\sigma}$ is differentiable. Then (see chapter 4) we have the Fréchet derivative

$$d\bar{\mathcal{J}}(\rho)\hat{\rho} = -\int_D \bar{\sigma}'(\rho)\hat{\rho}|\nabla \bar{u}_\rho|^2 dx + \ell \int_D \hat{\rho} dx.$$

The necessary optimality condition $d\bar{\mathcal{J}}(\rho)(\tilde{\rho}-\rho) \geq 0 \;\forall \tilde{\rho} \in L^\infty(D,[0,1])$ is equivalent to

$$\bar{g}_\rho := -\bar{\sigma}'(\rho)|\nabla \bar{u}_\rho|^2 + \ell \begin{cases} \geq 0 \text{ a.e. on } & \{\rho = 0\} \\ = 0 \text{ a.e. on } & \{0 < \rho < 1\} \\ \leq 0 \text{ a.e. on } & \{\rho = 1\}. \end{cases} \tag{6.29}$$

Indeed, for any $m \in \mathbb{N}^*$, choosing $\tilde{\rho} = \rho + \varphi\chi_{\{\rho \leq 1-1/m\}}$ with an arbitrary $\varphi \in L^\infty(D,[0,1-1/m])$ shows that $\bar{g}_\rho\chi_{\{\rho\leq 1-1/m\}} \geq 0$, whereby $\bar{g}_\rho \geq 0$ a.e. on $\cup_{m\in\mathbb{N}^*}\{\rho \leq 1-1/m\} = \{\rho < 1\}$. Likewise we show that $\bar{g}_\rho \leq 0$ a.e. on $\{\rho > 0\}$.

We now require that (6.27) and (6.29) be equivalent when $\rho = \chi_\Omega$. For this it is sufficient that

$$\bar{\sigma}'(1) = -2\sigma_1\frac{\sigma_0-\sigma_1}{\sigma_0+\sigma_1}, \qquad \bar{\sigma}'(0) = 2\sigma_0\frac{\sigma_1-\sigma_0}{\sigma_1+\sigma_0}.$$

When $\sigma_0 \approx 0$ and $\sigma_1 = 1$ this gives $\bar{\sigma}'(1) = 2$ and $\bar{\sigma}'(1) = 0$. Therefore a suitable interpolation profile is $\boxed{\bar{\sigma}(t) = t^2}$.

Remark 6.25 *In 3 dimensions, the same arguments yield as unique third degree polynomial profile $\bar{\sigma}(t) = -\frac{1}{2}t^3 + \frac{3}{2}t^2$. In planar elasticity with Poisson ratio $\nu = 1/3$ we find $\bar{\sigma}(t) = t^3$. This cubic profile is very popular in structural optimization: it is known as the* **SIMP method** *(Solid Isotropic Material with Penalization).*

In order to optimize the interpolated problem, standard continuous optimization methods can be used. The simplest one is the gradient method with projection onto the constraint $0 \leq \rho \leq 1$, namely:

$$\rho_{k+1} = \min(1, \max(0, \rho_k - t_k\bar{g}_{\rho_k})).$$

Of course, there is no guarranty that the obtained density field will be a characteristic function. However, convex profiles tend to penalize intermediate densities, since they yield in such cases a weak material property. Two examples of compliance minimization in elasticity are shown in figures 6.6 and 6.7. The initialization is full ($\rho = 1$). Here the final densities do not exhibit regions of intermediate densities, however it is not always the case.

Remark 6.26 *Interpolation methods often incorporate filtering techniques. For appropriate filters, it can be shown that the interpolated formulation is also consistent with the geometric optimality condition $d_S\mathcal{J}(\Omega,\theta) = 0$ for any displacement θ of $\partial\Omega$, see [5].*

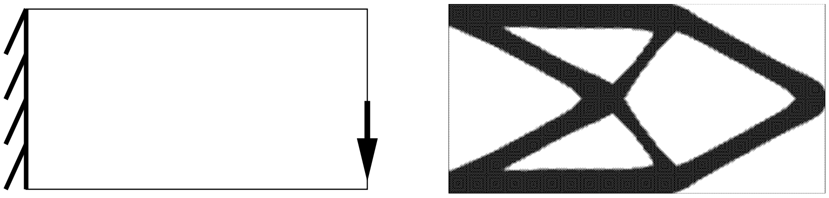


Figure 6.6: Cantilever: boundary conditions (left) and optimized density (right)

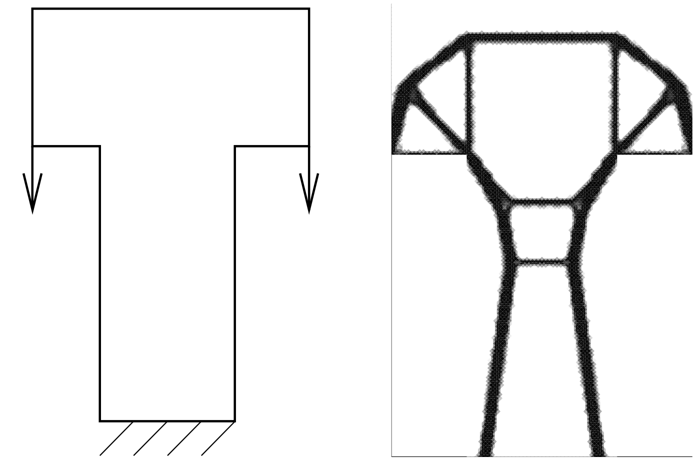


Figure 6.7: Mast: boundary conditions (left) and optimized density (right)

Notation

$\mathcal{L}(X, Y)$: set of continuous linear maps from X to Y

$\mathrm{isom}(X, Y)$: set of continuous isomorphisms from X to Y

$X' = \mathcal{L}(X, \mathbb{R})$: continuous dual space of X

$\mathcal{C}_c(X)$: set of continuous functions on X with compact support

$\mathcal{C}_b(X)$: set of bounded continuous functions on X

$\mathcal{C}_b^1(X)$: set of bounded $\mathcal{C}^1$ functions on X with bounded first order partial derivatives

$\mathcal{M}_n(\mathbb{R})$: set of $N \times N$ real matrices

$\mathcal{S}_N(\mathbb{R})$: set of symmetric $N \times N$ real matrices

$GL_n(\mathbb{R})$: set of $N \times N$ invertible real matrices

$\mathrm{Hom}(A, B)$: set of homeomorphisms from A into B.

Index

Bibliography

[1] G. Allaire. Conception optimale de structures. Springer, 2007.

[2] G. Allaire. Shape optimization by the homogenization method. Springer, 2002.

[3] H. Ammari, H. Kang. Polarization and moment tensors. Springer, 2007.

[4] S. Amstutz. Analyse de sensibilité topologique et applications en optimisation de formes. Habilitation thesis, 2011.

[5] S. Amstutz, C. Dapogny, A. Ferrer. A consistent relaxation of optimal design problems for coupling shape and topological derivatives. Numerische Mathematik 140(1), pp. 35-94, 2018.

[6] H. Attouch, G. Buttazzo, G. Michaille. Variational analysis in Sobolev and BV spaces. SIAM MPS, 2006.

[7] S. Benzoni-Gavage. Calcul différentiel et équations différentielles. Dunod, 2010.

[8] H. Brézis. Analyse fonctionnelle. Dunod, 1999.

[9] D. Bucur, G. Buttazzo. Variational methods in shape optimization problems. Birkhäuser, 2005.

[10] P.G. Ciarlet. Linear and nonlinear functional analysis with applications. SIAM, 2013.

[11] F. Demengel, G. Demengel. Functional spaces for the theory of elliptic partial differential equations. Springer, 2012.

[12] H. Eschenauer, V.V. Kobelev, A. Schumacher. Bubble method for topology and shape optimization of structures. Structural optimization 8:42-51, 1994.

[13] L.C. Evans, R.F. Gariepy. Measure theory and fine properties of functions. CRC Press, 1992.

[14] P. Gangl and K. Sturm. A simplified derivation technique of topological derivatives for quasi-linear transmission problems. ESAIM:COCV 26(106), 2020.

[15] S. Garreau, Ph. Guillaume, M. Masmoudi. The topological asymptotic for PDE systems: the elasticity case. SIAM J. Control Optim. 39(6):1756:1778, 2001.

[16] A. Henrot, M. Pierre. Variation et optimisation de formes. Springer, 2005.

[17] F. Murat, J. Simon. Sur le contrôle par un domaine géométrique. Internal report 76 015, L. D'Analyse Numérique Univ. Paris 6, 1976.

[18] A.A. Novotny, J. Sokolowski. Topological derivatives in shape optimization. Springer, 2013.

[19] W. Rudin. Analyse réelle et complexe. Dunod, 1998.

[20] J. Sokolowski, A. Zochowski. On the topological derivative in shape optimization. SIAM J. Control Optim. 37(4): 1251-1272, 1999.

[21] J. Sokolowski, J.-P. Zolesio. Introduction to shape optimization. Springer, 1992.