Notes on Calculus of Variation with applications in Image Analysis

version 0.2-082015

François Lauze

August 24, 2015

Contents

List of Figures

Chapter 1

Introduction

The purpose of these notes is to introduce, in a somehow informal way, although with a few abstract concepts, some elements of the Calculus of Variations. I will in particular avoid functional analysis aspects, although I will mention a few Hilbert spaces. I will however assume that the reader is familiar with some formulas from Integral Calculus, although I will recall them when needed.

A main point of these notes is the link between the notion of differential of a function and the notion of Gradient. Gradients depend on a Euclidean (Hilbertian) inner product on the space where one works, while differential does not. The adjoint of a linear operator is then naturally introduced. Gradient descent is explained via teh Caucht-Schwarz inequality. I discuss equality constraints and Lagrange multiplier methods. Inequality constraints are skipped in the current version of the document, but they might appear one day. I (intend to) discuss briefly second variation and Hessian, they are often forgotten in Image Processing papers, but should not.

Then I start playing with variational problems defined on some Hilbert space of functions on an open domain of $\mathbb{R}^n$. For that purpose I recall a couple of classical results from integral calculus, they are used to compute a series of first variations and gradients of functionals that are often encountered in Image Analysis.

The two next chapters deal with planar curves and their use in a series of functionals. They come both from classical examples in mathematics as well as segmentation problems in Image Analysis. They include Euclidean Shortening Flows, Snakes, Geodesic Active Contours, Chan Vese models, and should grow when I have time.

This document is targeted toward not too scared electrical engineers and computer scientists, but not hard core mathematicians, I have a reputation to defend.

Chapter 2

Finite Dimension

In this chapter, we are interested in optimization of real functions defined on an open subset of an Euclidean space, generally $\mathbb{R}^n$. I will introduce in a bit formal way the notion of Euclidean (and Hilbertian) structure, adjunction and gradients. I will discuss smooth constrained optimization and Lagrange multipliers.

2.1 Objective functions and optimizers

Given a function $f : U \subset \mathbb{R}^n \to \mathbb{R}$, defined on an open subset U of $\mathbb{R}^n$, a standard question is to find, if it exists, an extremum of f, minimum or maximum, and its corresponding location $\mathbf{x}^* \in \bar{U}$ such that

$$\forall \mathbf{x} \in U, \quad f(\mathbf{x}^*) \leq f(\mathbf{x}) \text{ (resp. } f(\mathbf{x}^*) \geq f(\mathbf{x}) \text{).}$$

When $\mathbf{x}^*$ corresponds to a local (resp. global) minimum $f(\mathbf{x}^*)$, $\mathbf{x}^*$ is called a local (resp. global) *minimizer* of f. When $\mathbf{x}^*$ corresponds to a local (resp. global) maximum $f(\mathbf{x}^*)$, $\mathbf{x}^*$ is called a local (resp. global) *maximizer* of f. In many problems, this function is built as a function that measures "how well" a certain number of criteria are enforced at a given point $\mathbf{x} \in U$. Such a function is called an *objective function* and an extremum $\mathbf{x}_0$ represents the location where an optimal trade-off is achieved between the different criteria.

When the optimization is a minimization, f is generally called a *cost function* or *energy function*, $f(\mathbf{x})$ measures the "cost" of choosing $\mathbf{x}$, or the "energy used" for that choice, and one searches for a point of minimal cost/energy.

When f is differentiable, and we will always assume it from now, a necessary condition is that the differential $df_{\mathbf{x}}$ of f in $\mathbf{x}$ vanishes. Instead of using this differential, one often considers the gradient $\nabla_{\mathbf{x}} f$ of f in $\mathbf{x}$, which is a vector in $\mathbb{R}^n$ and we are naturally led to solve in $\mathbf{x}$ the vector equation $\nabla_{\mathbf{x}} = 0$. This can be done by different methods, more or less "direct", as the equation to solve may be rather complex. Among these methods, the classical *gradient descent*,

when seeking a minimizer $\mathbf{x}^*$ of f, one attempts to obtain it as a solution of the differential equation:

$$\mathbf{x}_t = -\nabla f_{\mathbf{x}}.$$

The underlying idea is that, starting from a point $(\mathbf{x}_0, f(\mathbf{x}_0))$ on the graph G_f of f, one goes down the graph along the curve of steepest slope, and this is encoded by $-\nabla_x f$, and one will hopefully "hit" an equilibrium point, i.e.$\mathbf{x}_t = 0$, which means that $-\nabla_{\mathbf{x}} f = 0$. Under some assumptions on the shape of G_f, this should provide a local minimizer.

2.2 Differentials

It thus becomes necessary to precise the connection between the differential and the gradient of a function. To do so, I will start recalling some (should be) well known notions of differential calculus and Euclidean geometry.

Recall that the differential of f at a point $\mathbf{x} = (x_1, \ldots, x_n)$ is defined as the unique linear form (if it exists) $L_{\mathbf{x}} : \mathbb{R}^n \to \mathbb{R}$ satisfying

$$f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + L_{\mathbf{x}}(\mathbf{h}) + o(\mathbf{h})$$

where $\mathbf{h}$ is "sufficiently" small.

The directional derivative of f at $\mathbf{x}$ in the direction $\mathbf{v}$ is defined as the limit. if it exists, of

$$d_{\mathbf{x}} f\mathbf{v} := \lim_{t \to 0} \frac{f(\mathbf{x} + t\mathbf{v}) - f(\mathbf{x})}{t}$$

A clearly equivalent definition of $d_{\mathbf{x}} f\mathbf{v}$ is the following (I will also use the term "Gâteaux Derivative", especially in the infinite dimensional case). Let $\ell : \mathbb{R} \to \mathbb{R}$ be defined as $\ell(t) = f(\mathbf{x} + t\mathbf{v})$. Then

$$d_{\mathbf{x}} f\mathbf{v} = \ell'(0).$$

In the calculus of Variations, the linear functional $d_{\mathbf{x}} f$ is often called the *first variation of* f at $\mathbf{x}$. When f est differentiable, on has in fact the simple relation (simple but extremely important in the rest of these notes)

$$d_{\mathbf{x}} f\mathbf{v} = L_{\mathbf{x}}(\mathbf{v}).$$

and I will denote $d_{\mathbf{x}} f$ the differential of f at $\mathbf{x}$. Until now, I have not talked of partial derivatives. The partial derivative of f wrt[1] the i-th variable x_i at $\mathbf{x}$ is by definition:

$$\frac{\partial f}{\partial x_i}(\mathbf{x}) := d_{\mathbf{x}} f e_i$$

where e_i is the i-th element of the natural basis of $\mathbb{R}^n$ This notion of partial derivative is thus basically the same as the one of directional derivative, but

[1] wrt = with respect to

the important part is with respect to the choice of a basis of $\mathbb{R}^n$ as follows: If $\mathbf{v} = \sum_{i=1}^n v_i e_i$, one has the relation

$$d_{\mathbf{x}} f \mathbf{v} = \sum_{i=1}^{n} v_i \frac{\partial f}{\partial x_i}(\mathbf{x}).$$

This relation is in fact an *inner product*:

$$\sum_{i=1}^{n} v_i \frac{\partial f}{\partial x_i}(\mathbf{x}) = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \cdot \begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix} \tag{2.1}$$

(I threw the $\mathbf{x}$ away to keep the formulas legible). The vector

$$\begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

is the gradient of f, denoted ∇f. It seems that this gradient depends a priori on the choice of a basis, but in fact *the dependency is "weaker"*: it the standard basis is replaced by an other one, also orthonormal, the the value of expression (2.1) *does not change*. We will use this property to define the gradient in an abstract way.

2.3 Geometry and Gradients

I will first recall a few concepts of linear algebra dealing with the notion of inner product and orthogonality.

Definition 1 Euclidean Spaces, (real) Hilbert Spaces *A* prehilbertian *real vector space E (i.e. a scalars are real) is a vector space endowed with an* inner product *or* inner product*, i.e. a bilinear form $(\mathbf{x}, \mathbf{y}) \mapsto \pi(\mathbf{x}, \mathbf{y}) \in \mathbb{R}$ with the following properties:*

1. *$\forall \mathbf{x} \in E$, $\pi(\mathbf{x}, \mathbf{x}) \geq 0$ (positive)*
2. *$\pi(\mathbf{x}, \mathbf{x}) = 0 \Longleftrightarrow \mathbf{x} = 0$ (definite)*
3. *$\pi(\mathbf{x}, \mathbf{y}) = \pi(\mathbf{y}, \mathbf{x})$ (symmetric).*

The bilinear form π is sometimes called a prehilbertian metric. Two vectors $\mathbf{x}$ et $\mathbf{y}$ are orthogonal if $\pi(\mathbf{x}, \mathbf{y}) = 0$. the prehilbertian norm of a vector $\mathbf{x}$ is $\|\mathbf{x}\|_\pi = \sqrt{\pi(\mathbf{x}, \mathbf{x})}$, it depends of course on the inner product and associated with this norm is the prehilbertian distance $d(\mathbf{x}, \mathbf{y}) := \|\mathbf{x} - \mathbf{y}\|_\pi$. One generally writes $\pi(\mathbf{x}, \mathbf{y}) = \mathbf{x} \cdot \mathbf{y}$ or $\langle \mathbf{x}, \mathbf{y} \rangle$ or $\langle \mathbf{x}, \mathbf{y} \rangle_\pi$ if one wants to emphasize the dependency in π, and his will prove important in several occasions in these notes. When

E is finite-dimensional, one talks of Euclidean *space. A prehilbertian space is called* Hilbertian *or* Hilbert space *if it is complete wrt the above distance, i.e. every Cauchy sequence for that distance converges. An Euclidean space is always Hilbertian, the complications arise in the infinite dimensional case.*

Definition 2 Orthogonality. *Let E be an Euclidean space (resp. Hilbert) with its inner product π. When $\pi(\mathbf{x},\mathbf{y}) = 0$ we say that $\mathbf{x}$ is orthogonal (for π, or π-orthogonal) to $\mathbf{y}$ and we denote it $\mathbf{x}\perp\mathbf{y}$ or $\mathbf{x}\perp_\pi\mathbf{y}$ if we need / wish to make the inner product explicit. The set $\{\mathbf{y}\in E, \pi(\mathbf{x},\mathbf{y}) = 0\}$ is the orthogonal $\mathbf{x}^\perp$ of $\mathbf{x}$. When V is vector subspace of E, then $V^\perp = \{y \in E, \pi(\mathbf{x},\mathbf{y}) = 0, \forall \mathbf{x} \in V\}$ is the orthogonal of V.*

There are plenty of super-classical relations, most of them are assumed to be known. One that we will use explicitly: if V is a subspace of E, $E = V \oplus V^\perp$, i.e any vector $\mathbf{h}$ of E can be written uniquely as a sum $\mathbf{h} = \mathbf{h}_1 + \mathbf{h}_2$ where $\mathbf{h}_1 \in V$ and $\mathbf{h}_2 \in V^\perp$. I may sometimes call h_1 the component of $\mathbf{h}$ *tangent* to V and $\mathbf{h}_2$ as the component of $\mathbf{h}$ *normal* to V.

What's is this stuff useful for? The following result (should be known, at least in the Euclidean case) is paramount for the sequel

Theorem 3 Representation of a linear form by a vector. *Let (E,π) an Hilbert space and $L : E \to \mathbb{R}$ a* linear form. *Then there exists a unique vector $\mathbf{v}$ such that*

$$L(\mathbf{x}) = \langle \mathbf{v}, \mathbf{x}\rangle_\pi.$$

Of course v depends on the inner product π.

In a more concrete way, a linear form on $\mathbb{R}^n$ is classically given by its coefficients $(l_1,\dots,l_n)$ and we clearly have

$$L(\mathbf{x}) = (l_1,\dots,l_n)\begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} = \sum_{i=1}^n l_i x_i$$

which is the usual inner product of $\mathbb{R}^n$! The corresponding vector $\mathbf{v}$ is just

$$\begin{pmatrix} l_1 \\ \vdots \\ l_n \end{pmatrix} = (l_1,\dots,l_n)^T$$

where T denotes transposition. Linked to transposition is the *adjunction* operation:

Definition 4 Adjunction, simple one, to start with. *Let E be an Euclidean space (or Hilbert) with inner product $\langle \mathbf{x},\mathbf{y}\rangle$ and $A : E \to E$ a linear map. The adjoint de A, denoted A^* is the unique linear map $E \to E$ which satisfies the following relation (called* adjunction relation*):*

$$\langle A\mathbf{x},\mathbf{y}\rangle = \langle \mathbf{x}, A^*\mathbf{y}\rangle.$$

If $A = A^*$ one says that A is self-adjoint.

In $\mathbb{R}^n$ with canonical basis and inner product,linear maps $\mathbb{R}^n \to \mathbb{R}^n$ are identified to square matrices. The adjoint A^* of A is just its *transposed* A^T! Thus a matrix is self-adjoint if and only if it is *symmetric*.

Definition 5 Adjunction, a bit more general. *Let E and F two Hilbert spaces (or just Euclidean), $\langle -, - \rangle_E$ and $\langle -, - \rangle_F$ their respective inner products,and $A : E \to F$ a linear map. The adjoint A^* of A is the unique linear map $F \to E$ (if it exists, always the case in finite dimension, but for the infinite dimensional case, I'm not sure) which satisfies, for all $\mathbf{x} \in E$, $\mathbf{y} \in F$,*

$$\langle A\mathbf{x}, \mathbf{y} \rangle_F = \langle \mathbf{x}, A^*\mathbf{y} \rangle_E.$$

Once again, in the finite dimensional situation, if $E = \mathbb{R}^n$, $F = \mathbb{R}^n$ with standard bases and inner products, A is a $n \times m$ matrix and A^* is just its transposed A^T! In fact, transposition and adjunction are identical in the standard Euclidean setting!

It's now time to come back to our main subject of interest: gradients.

Definition 6 *Let g a differentiable function from an (open of an) Euclidean space E (inner product $\langle -, - \rangle_E$) with values in $\mathbb{R}$ and let $g'(x) = d_{\mathbf{x}}g$ be its differential at point $\mathbf{x}$. It's a linear map from E to $\mathbb{R}$ and by the representation theorem above, there exists a unique vector $\mathbf{v} \in E$ such that*

$$\forall \mathbf{h} \in E, \quad d_{\mathbf{x}}g(\mathbf{h}) = \langle \mathbf{v}, \mathbf{h} \rangle_E$$

This vector $\mathbf{v}$ is, by definition, the gradient *of g in $\mathbf{x}$ and is denoted $\nabla_{\mathbf{x}}g$ (or ∇g when there is no ambiguity).*

The important remark here is that the notion of gradient is inner product dependent, while differential is not

Important idea. If, for a "generic direction" $\mathbf{h}$, one can write $d_{\mathbf{x}}g\mathbf{h}$ as $\langle \mathbf{v}, \mathbf{h} \rangle$, that is, such that the second factor is $\mathbf{h}$ itself, then $\mathbf{v}$ must be the gradient of g at $\mathbf{x}$ (for the given inner product). This is this kind of computation one does in order to obtain a more or less analytical form for the gradient.
And what about adjunction here? In fact, in a typical gradient computation, one obtains in general an expression of the form

$$d_{\mathbf{x}}g\mathbf{h} = \langle \mathbf{w}, L\mathbf{h} \rangle$$

where L is some linear map. By adjunction, one gets $\langle \mathbf{w}, L\mathbf{h} \rangle = \langle L^*\mathbf{w}, \mathbf{h} \rangle$, which thus means that the sought gradient is $L^*\mathbf{w}$.

There is in fact an other definition of gradient, which, for an Eucliean/Hilbert space provides an equivalent characterization, but that only requires a Banach space E, we won't use it but I add for the shake of completeness (for more see [**?**]).

Recall (perhaps) that a vector space E with a norm $\| - \|_E$ is a Banach space if it is complete for that norm. A linear form F on E (i.e. a linear map $F : E \to \mathbb{R}$) is

continuous if there exists a constant $c > 0$ such that for each $h \in E$, $|F(h)| \leq c\|h\|_E$ ($|F(h)|$ is just the absolute value of $F(h) \in \mathbb{R}$). The infemum of such constant c's is called the operator (or dual here) norm of F,$\|F\|_{E^*}$. It is also characterized as

$$\|F\|_{E^*} = \sup_{h \in E, h \neq 0} \frac{|F(h)|}{\|h\|_E}.$$

Let g a differentiable function from (an open set of) E to $\mathbb{R}$, $x \in E$ and let $g'(x) = d_x g$ the differential of g at x. This is a linear form on E, that we will assume continuous.

Definition 7 *Among all $h \in E$ such that $\|h\|_E = \|d_x g\|_{E^*}$ (they form a sphere of radius $\|d_x g\|_{E^*}$), there exists a unique one $\bar{h}$ which maximizes $d_x g(h) \in \mathbb{R}$. This is the* gradient $\nabla_x g$ *of g at x.*

The uniqueness of $\bar{h}$ needs to be proved, we will not do it here. This definition is interesting/importannt as it frees us from the Euclidean/Hilbertian restriction allowing optimization via geometric approaches in more general spaces. In particular for ℓ^p and L^p spaces, the operator/dual norm is well known and this lead to interesting gradient computations.

2.4 An Important Gradient Computation

Two differentiable functions $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^m \to \mathbb{R}^n$ are given, with usual bases and inner products. Let us set $k(\mathbf{x}) := f(g(\mathbf{x}))$. We want to compute ∇k. To that extend, we compute the derivative of the function $\ell : t \mapsto \ell(t) = k(\mathbf{x} + t\mathbf{h}) = f(g(\mathbf{x} + t\mathbf{h}))$ at $t = 0$:

$$\ell'(t) = df_{g(\mathbf{x}+t\mathbf{h})} dg_{\mathbf{x}+t\mathbf{h}} \mathbf{h}$$

by the chain rule. The map $dg_{\mathbf{x}}$ from $\mathbb{R}^m \to \mathbb{R}^n$ is linear. At $t = 0$ we then get

$$\begin{aligned} dk_{\mathbf{x}}\mathbf{h} &= df_{g(\mathbf{x})} dg_{\mathbf{x}} \mathbf{h} \\ &= \langle \nabla f_{g(\mathbf{x})}, dg_{\mathbf{x}} \mathbf{h} \rangle \quad \text{definition du gradient de } f \\ &= \langle dg^*_{\mathbf{x}} \nabla f_{g(\mathbf{x})}, \mathbf{h} \rangle \quad \text{adjunction} \end{aligned}$$

We succeeded in writing the directional derivative, for a generic direction $\mathbf{h}$ as an inner product $\langle \mathbf{v}, \mathbf{h} \rangle$ with $\mathbf{v} = dg^*_{\mathbf{x}} \nabla f_{g(\mathbf{x})}$. $\mathbf{v}$ is consequently the sought gradient. In matrix notations, dg is the Jacobian matrix

$$Jg = \begin{pmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_m} \\ \vdots & \vdots & \vdots \\ \frac{\partial g_n}{\partial x_1} & \cdots & \frac{\partial g_n}{\partial x_m} \end{pmatrix}$$

of$g$ at $\mathbf{x}$. We thus have, (with standard bases blablabla...) written the gradient as $\nabla k = Jg^T \nabla f = (df\, Jg)^T$.

This example can appear as a somewhat artificial play with the chain rule, it is in my opinion, one of the most fundamentals in order to understand the

algebra for the Calculus of Variations. It reflects the series of operations that have to be applied in so as to obtain the gradient of a functional and the associated Euler-Lagrange equation. This will be the object of one of the following sections. We will encounter some non linear f, while g will in general be a linear differential operator. It will be necessary to give a meaning to such an operator.

2.5 Change of Metric and Gradient

In $\mathbb{R}^n$, we know that the datum of an inner product is the same as the datum of a symmetric, positive definite matrix A: the form $(\mathbf{x}, \mathbf{y}) \mapsto \mathbf{x}^T A\mathbf{y}$ is symmetric and positive definite. The usual inner product corresponds to $A = I_n$, the identity matrix, I will denote it as $\langle -, - \rangle_{I_n}$, $\langle \mathbf{x}, \mathbf{y} \rangle_{I_n} = \mathbf{x}^T\mathbf{y}$. Given a function $f : (U \subset) \mathbb{R}^n \to \mathbb{R}$, its standard gradient (i.e. gradient for the standard inner product) $\nabla_{I_n} f$ is just

$$\begin{pmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{pmatrix}$$

But what happens if one changes the inner product? We want to compute $\nabla_A f$. This is straightforward, but let us do it step by step. $\nabla_A f$ is defined by the relation

$$\langle \nabla_A f, \mathbf{h} \rangle_A = df\mathbf{h}$$

But

$$\begin{aligned} df\mathbf{h} &= \langle \nabla_A f, \mathbf{h} \rangle_A \\ &= (\nabla_A f)^T A\mathbf{h} \quad \text{definition of inner product} \\ &= \langle \nabla_A f, A\mathbf{h} \rangle_{I_n} \quad \text{definition of inner product} \\ &= \langle A^*\nabla_A f, \mathbf{h} \rangle_{I_n} \quad \text{adjunction} \\ &= \langle A\nabla_A f, \mathbf{h} \rangle_{I_n} \quad \text{symmetric} \iff \text{self-adjoint} \\ &= \langle \nabla_{I_n} f, \mathbf{h} \rangle_{I_n} \quad \text{definition of gradient} \end{aligned}$$

and we have the relation $\nabla_{I_n} f = A\nabla_A f$. In other words,

$$\nabla_A f = A^{-1}\nabla_{I_n} f.$$

In order to make the reader more confused, as well as my explanations, the dimension 1 case is somewhat special. Let us Consider $\mathbb{R}$: every inner product on $\mathbb{R}$ is given by $(x, y) \mapsto \lambda xy$ with $\lambda > 0$ since symmetric positive definite matrices on $\mathbb{R}$ are just strictly positive real numbers. This means, that, up to a strictly positive multiplicative constant, gradient and differential (which is nothing else than the multiplication by the usual derivative) match. We can just suppose that this constant λ is 1, and will no longer distinguish between gradient, differential and derivative.

To add more confusion, in the somewhat opposite situation, in infinite dimension, A will often be represented by a differential operator. To change the

metric and obtain the new gradient from the old one, it will be necessary to solve a linear differential equation. This is what Yezzi *et al.*, in one hand, and Charpiat *et al.* have done, we will see that with the *Sobolev active contours*.

2.6 Cauchy-Scharwz, gradient descent

The Cauchy-Schwarz formula is also a elementary and classical result, which will make precise how the gradient indicates the direction of largest change. It is obviously important for optimization via gradient descent/ascent, and thus in many variational questions.

Theorem 8 Cauchy-Scharwz. *Let* $\mathbf{x}$ *and* $\mathbf{y}$ *two elements from an Hilbert space* E. *Then*

$$|\langle \mathbf{x}, \mathbf{y} \rangle| \leq \|\mathbf{x}\| \|\mathbf{y}\|$$

with equality if and only if $\mathbf{x}$ *and* $\mathbf{y}$ *are colinear.*

The proof of the theorem is elementary. It main consequence, at least for this document is the following. Assume that we have succeeded in writing $df_{\mathbf{x}}\mathbf{h} = \langle \nabla f_{\mathbf{x}}, \mathbf{h} \rangle$. From Cauchy-Schwarz, we thus have

$$|df_{\mathbf{x}}\mathbf{h}| = |\langle \nabla f_{\mathbf{x}}, \mathbf{h} \rangle| \leq \|\nabla_{\mathbf{x}} f\| \|\mathbf{h}\|$$

and the direction $\mathbf{h}$ for which $f(\mathbf{x}+\mathbf{h})$ changes the most is by necessity one for which we have equality in Cauchy-Schwarz, therefore it is the gradient direction and since $f(\mathbf{x}+\mathbf{h}) = f(\mathbf{x}) + \langle \nabla f_{\mathbf{x}}, \mathbf{h} \rangle + o(\mathbf{h})$, it will be $+\nabla f_{\mathbf{x}}$ for a maximal increase, and $-\nabla f_{\mathbf{x}}$ for a maximal decrease, around $\mathbf{x}$.

The following result will conclude this section.

Theorem 9 *Let* $f : \Omega \to \mathbb{R}$ *be a differentiable function, where* Ω *is a bounded open set (a domain), having a minimum at* $\bar{\mathbf{x}} \in \Omega$. *Then* $\nabla f_{\bar{\mathbf{x}}} = 0$. *If this is the only vanishing point for the gradient in* Ω, *then the gradient descent differential equation*

$$\mathbf{x}(0) = \mathbf{x}_0 \in \Omega, \quad \mathbf{x}_t = -\nabla f_{\mathbf{x}}$$

satisfies $\lim_{t\to\infty} \mathbf{x}(t) = \bar{\mathbf{x}}$.

In general the minimum is not attained in finite time. For instance, if $f = x^2$, $\nabla f = 2x$, gradient descent becomes

$$x'(t) = -2x(t)$$

whose general solution has the form $x_0 e^{-2t}$. If $x_0 \neq 0$, the minimum of f won't be reached in finite time.

2.7 Equality Constraints

It is of course not the end of the story! In the situations described aboved, the function f had to be optimized on an open set of an Euclidean space. In many cases, we have to consider situations where f is defined or has to be optimized on a subspace, linear, affine, or more general, of an Euclidean space, and more generally on an open set of such a subspace.

Let us start with a very elementary example, which illustrates the some of the problems one encounters. Let $f : \mathbb{R}^2 \to \mathbb{R}$ be the function $(x, y) \mapsto x^2 + (y-1)^4$. Its "standard" gradient is given by

$$\nabla f_{(x,y)} = \begin{pmatrix} 2x \\ 4(y-1)^3 \end{pmatrix}$$

which vanishes only at point $(0, 1)$. Let us assume that we are interested by the values of f on the line $D : x + y = 0$. Point $(0, 1) \notin D$! It cannot be a solution of the problem

$$\min_D f(x, y).$$

Of course, by a direct substitution $y = -x$, one can compute the minimum of$g(x) = f(x, -x) = x^2 + (-x - 1)^4$. One has $g'(x) = 4x^3 + 12x^2 + 14x + 1$, its unique real root is $\bar{x} \approx -0.0762905$ and the solution of the minimization problem is $(\bar{x}, -\bar{x})$.

Ok, that was easy! But what is D is a more complicated subspace? Let us assume that D is defined implicitely by an equation $\phi(x, y) = 0$ sufficienty complicated so that one can not perform a substitution. Then lets us perform a simple but nevertheless interesting computation.

Let us assume the ambient space is $\mathbb{R}^n$, so I will be able to (re)use some standard notations. We will also assume that ϕ is *differentiable*. Set $D = \{\mathbf{x}, \phi(\mathbf{x}) = 0\}$ and, last but not least, assume that $\nabla\phi$ does not vanish on D. This condition, in the "modern" language of differential geometry, means that D is a *submanifold* of $\mathbb{R}^n$, and in our case, an *hypersurface* of $\mathbb{R}^n$ (pour $n = 2$ this is a curve, for $n = 3$ a surface...). For a given $\mathbf{h}$ small enough, a Taylor expansion gives

$$\phi(\mathbf{x} + \mathbf{h}) = \phi(\mathbf{x}) + \nabla\phi_{\mathbf{x}} \cdot \mathbf{h} + o(\mathbf{h}).$$

We deduce from it that the directions $\mathbf{h}$ for which ϕ changes the least in a neighborhood of $\mathbf{x}$, and thus that $\phi(\mathbf{x} + \mathbf{h}) \approx 0$, i.e. $\mathbf{x} + \mathbf{h}$ remains close to D, must be *orthogonal* to $\nabla\phi_{\mathbf{x}}$, for, in this case, the inner product $\nabla\phi_{\mathbf{x}} \cdot \mathbf{h} = 0$. These directions form the *tangent space* (line, plane..) tangent to D in $\mathbf{x}$ and its usual notation is $T_{\mathbf{x}}D$.

Now Taylor expansion for f gives

$$f(\mathbf{x} + \mathbf{h}) = f(\mathbf{x}) + \nabla f_{\mathbf{x}} \cdot \mathbf{h} + o(\mathbf{h})$$

for $\mathbf{h}$ tangent to D. A few remarks. Let $N_{\mathbf{x}}D$ be the subspace orthogonal to $T_{\mathbf{x}}D$, then

$$\nabla f_{\mathbf{x}} = v_1 + v_2, \quad v_1 \in T_{\mathbf{x}}D,\ v_2 \in N_{\mathbf{x}}V$$

and for $\mathbf{h} \in T_{\mathbf{x}}D$,

$$\nabla f_{\mathbf{x}} \cdot \mathbf{h} = v_1 \cdot \mathbf{h} + v_2 \cdot \mathbf{h} = v_1 \cdot \mathbf{h}$$

because $v_2 \perp \mathbf{h}$ by definition of v_2: in this case, only the component of $\nabla f_{\mathbf{x}}$ tangent to D en $\mathbf{x}$ plays a role. This is the *gradient of f along D in $\mathbf{x}$*. We deduce from it that in order for f to vary minimally in a neighborhood of $\mathbf{x}^*$ *along* D, it is necessary that $\nabla f_{\mathbf{x}^*} \cdot \mathbf{h} = 0$, and thus that $v_1 = 0$,i.e. $\nabla f_{\mathbf{x}^*} \perp \mathbf{h}$. But $\mathbf{h} \perp \nabla \phi_{\mathbf{x}^*}$ and therefore $\nabla f_{\mathbf{x}^*}$ must then be *parallel* to $\nabla \phi_{\mathbf{x}^*}$. In other words: $\nabla f_{\mathbf{x}^*}$ must be orthogonal to the tangen space $T_{\mathbf{x}^*}D$ of D at $\mathbf{x}^*$.

Since we assumed that $\nabla \phi_{\mathbf{x}^*} \neq 0$, this parallelism conditions means that there exists a unique real λ^* such that

$$\nabla f_{\mathbf{x}^*} + \lambda^* \nabla \phi_{\mathbf{x}^*} = 0.$$

This real λ^* is called the *Lagrange multiplier* of f associate to the constraint $\phi = 0$. How to use it? We first define an *augmented objective function* on $\mathbb{R}^{n+1}$:

$$\bar{f} : (\mathbf{x}, \lambda) = f(\mathbf{x}) + \lambda \phi(\mathbf{x})$$

and we compute its differential in $(\mathbf{x}, \lambda)$

$$d\bar{f}_{(\mathbf{x},\lambda)} = (df_{\mathbf{x}} + \lambda d\phi_{\mathbf{x}}, \phi(x)).$$

We need to search for a *stationary point* $(\mathbf{x}^*, \lambda^*)$ at which this differential vanishes, or the associated gradient

$$\nabla \bar{f}_{(\mathbf{x}^*,\lambda^*)} = \begin{pmatrix} \nabla f_{\mathbf{x}^*} + \lambda \nabla \phi_{\mathbf{x}^*} \\ \phi(\mathbf{x}^*) \end{pmatrix} = 0.$$

Thus, at this critical point $(\mathbf{x}^*, \lambda^*)$, the first component $\mathbf{x}^*$ satisfies the constraint. Beware nevertheless that such a critical point in general *is not an extremum* but a *saddle point* and a gradient descent type method will not function as is. Specific resolution strategies may have to be used. One of them will be presented a bit later.

What happens when one has several constraints? Figure 2.1 illustrate the situation where two constraints are given in $\mathbb{R}^3$. The first one is given by $\phi_1(x, y, z) = x^2 + y^2 + z^1 - 1 = 0$, the corresponding surface is the unit sphere $\mathbb{S}^2$. The second is a linear one, $\phi_2(x, y, z) = \alpha x + \beta y + \gamma z = 0$, the corresponding surface is a plane, that I have not drawn, and to satisfy the two constraint simultaneously mean that the (x, y, z) of the problem must lie in the intersection of $\mathbb{S}^2$ the plane, which is a great circle of $\mathbb{S}^2$, denoted C on the figure.

For the minimization problem on C we thus must only look at directions $\mathbf{h}$ that keep us infinitesimally on C. At a given point P, these form the tangent line to C at P, $T_P C$ and this is precisely the ones which are simultaneously orthogonal to $\nabla \phi_{1P}$ and $\nabla \phi_{2P}$, and for a function $f(x, y, z)$ to have a minimum on C at that point, its gradient must be orthogonal to $T_P C$ and thus lie in the subspace generated by the two vectors $\nabla \phi_{1P}$ and $\nabla \phi_{2P}$, denoted $\mathbf{a}$ and $\mathbf{c}$ in the

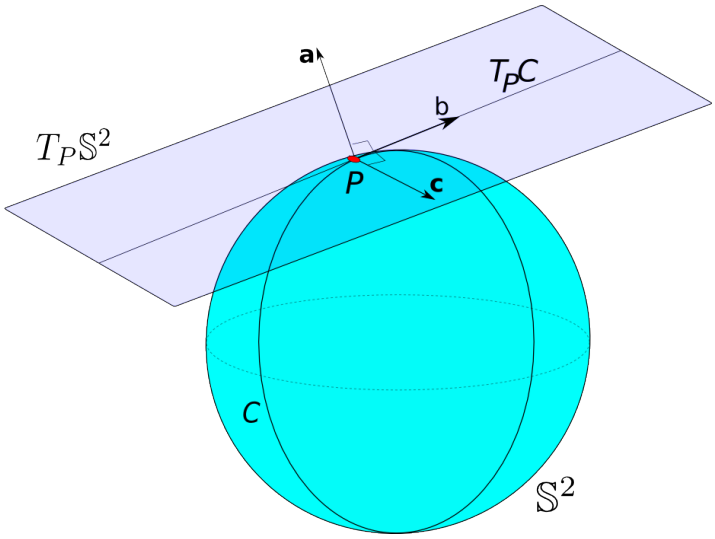

Figure 2.1: Contraintes et espaces tangents

figure. But this means that ∇f_P, $\nabla \phi_{1P}$ et $\nabla \phi_{2P}$ are linearly dependent and there exists two real numbers $\bar{\lambda}_1$ and $\bar{\lambda}_2$ such that

$$\nabla f_P + \bar{\lambda}_1 \nabla \phi_{1P} + \bar{\lambda}_2 \nabla \phi_{2P} = 0.$$

We thus have *two* Lagrange multipliers. The objective function augmentation technique still "works", we will seek a stationary point of

$$\bar{f}(x, y, z, \lambda_1, \lambda_2) = f(x, y, z) + \lambda_1 \phi_1(x, y, z) + \lambda_2 \phi_2(x, y, z).$$

And of course, if one has k constraints of the type $\phi_k(\mathbf{x}) = 0$,we can use k Lagrange multipliers (but I can no longer draw a picture).

Important idea to remember When one constrains an optimization problem to a subspace D, linear or not , one constrains search directions: at each point P of this subspace, search directions must belong to the *tangent space* of D at P, $T_P D$ and the gradient of the objective function at an extremum Q is orthogonal to the tagent space at this point $T_Q D$.
If D is a linear subspece, it is its own tangent space in *each point*. If D is affine, its tangent space at all point is the linear subspace parallel to D, or, what is in fact completely equivalent, one can figure the tangent space of D at P as D itself, but with the origo at P: all the vectors start from P. This can be important so as to understand boundary conditions which act as constraints for problems within the Calculus of Variations.

When the constraint is given as $\phi_1(\mathbf{x}) = 0, \ldots, \phi_k(\mathbf{x}) = 0$, the objective function f to be optimized can augmented by introducing k Lagrange multipliers and seek instead a critical point of

$$\bar{f}(\mathbf{x}, \lambda_1, \ldots, \lambda_k) = f(\mathbf{x}) + \sum_{i=1}^{k} \lambda_i \phi_i(\mathbf{x}).$$

Example. Optimizing a function on the circle $\mathbb{S}^1$. Set $g(x, y) = xy^2 + y^4 - y^2 + x + 1$. We want to find the minimum of g along the unit circle $\mathbb{S}^1 \subset \mathbb{R}^2$. Substitution works here, we can replace y^2 by $1 - x^2$ and solve for

$$\underset{x \in [-1,1]}{\arg.\min} f(x) = x(1 - x^2) + (1 - x^2)^2 - (1 - x^2) + x + 1.$$

The derivative of f has a unique real root, located at $\bar{x} \approx -0.763284$ and it can be check that this is indeed a minimum for f, thus g has two minima, at $(\bar{x}, \pm\sqrt{1 - \bar{x}^2})$.

Now we do it geometrically. If $\mathbf{x} = (x, y) \in \mathbb{S}^1$, we can make an orthonormal basis

$$\vec{\tau} = \begin{pmatrix} y \\ -x \end{pmatrix} \text{ and } \vec{\nu} = \begin{pmatrix} x \\ y \end{pmatrix}$$

with $\vec{\tau}$ tangent to $\mathbb{S}^1$ at $\mathbf{x}$ and $\vec{\nu}$ normal; $\vec{\tau}$ is thus a basis of the tangent space $T_{\mathbf{x}}\mathbb{S}^1$. Our optimality condition is then

$$\nabla_{\mathbf{x}} g \cdot \vec{\tau} = 0.$$

We denote by $\nabla_{\mathbb{S}^1, \mathbf{x}} g$ this quantity, the gradient of g along $\mathbb{S}^1$. A gradient descent approach gives

$$\mathbf{x}_t = -\nabla_{\mathbb{S}^1, \mathbf{x}} g.$$

A first attempt to solve it numerically:

$$\frac{\mathbf{x}^{n+1} - \mathbf{x}^n}{dt} = -\nabla_{\mathbb{S}^1, \mathbf{x}} g \iff \mathbf{x}^{n+1} = \mathbf{x}^n - dt \nabla_{\mathbb{S}^1, \mathbf{x}} g.$$

The problem here is that the new point $\bar{\mathbf{x}}^{n+1} = \mathbf{x}^n - dt - \nabla_{\mathbb{S}^1,\mathbf{x}} g$ will generally not belong to the circle, as illustrated in Figure 2.2, and this because the constraint is non linear, and we take a finite step as opposed to an infinitesimal one. We may project back by renormalizing, this is point $\tilde{\mathbf{x}}^{n+1}$ on the figure, or better, by "rolling" the update vector along the circle such that the length of the arc from $\mathbf{x}^n$ to $\mathbf{x}^{n+1}$ is the same as $\|\bar{\mathbf{x}}^{n+1} - \mathbf{x}^n\|$. Because arc length and angles are

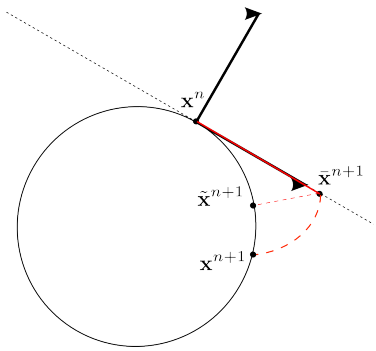

Figure 2.2: Optimization along the circle

the same for the unit circle, what we need to do is to *rotate* $\mathbf{x}^n$ by an angle of size $\|\bar{\mathbf{x}}^{n+1} - \mathbf{x}^n\|$ with the proper sign, this is just $h = -dt\nabla g \cdot \vec{\tau}$ since $\vec{\tau}$ has norm one. This can be written as

$$\mathbf{x}^{n+1} = R_{-h}\mathbf{x} = \begin{pmatrix} -\cos h & \sin h \\ -\sin h & -\cos h \end{pmatrix} \mathbf{x}^n = e^{\begin{pmatrix} 0 & h \\ -h & 0 \end{pmatrix}} \mathbf{x}^n.$$

The matrix R_h is the rotation of angle $-h$. The minus sign is due to the choice of orientation for $T_{\mathbf{x}}\mathbb{S}^1$, a positive h will generate a *clockwise* rotation, while the standard choice is *counterclockwise*. The map that rolls the tangent vector $\bar{\mathbf{x}}^{n+1} - \mathbf{x}^n$ to the circle is called the *Exponential map* and in this case has something to do with matrix exponential!

Finally, a numerical algorithm for our problem can be given as

- Let $\mathbf{x}^0 \in \mathbb{S}^1$ a starting value.
- While not converged up to a given threshold do
  1. Compute $h = -dt\,(\nabla g_{\mathbf{x}^n} \cdot \vec{\tau}_{\mathbf{x}^n})$
  2. Set $\mathbf{x}^{n+1} = R_{-h}\mathbf{x}^n$
- output $\mathbf{x}^\infty$

A numerical implementation in Matlab, gives the following:

Starting point $\mathbf{x}^0$	steady state $\mathbf{x}^\infty$
$\frac{1}{\sqrt{2}}(1,1)$	$(-0.763284, 0.646064)$
$-\frac{1}{\sqrt{2}}(1,1)$	$(-0.763284, -0.646064)$

The two solutions correspond to what we got by substitution. Note that the minimizer is not unique, and that Gradient Descent clearly depends on the starting point.

2.8 Gradient Descent and Lagrange Multipliers

[François dit: Talk about the determination of Lagrange multipliers that are time dependent.]

2.9 Inverse and Implicit Function Theorems

In this section I recall tow classical theorems in $\mathbb{R}^n$ and more generaly for a Banach space (that is: a real, normed vector space, which is complete for the distance induced par norm, Hilbert spaces are Banach spaces), but I will not go beyond $\mathbb{R}^n$ and finite dimensional spaces, although their importance lies also there, where they can provide existence of minimizers.

Theorem 10 Inverse Functions Theorem *Set $f : U \subset \mathbb{R}^n \to \mathbb{R}^n$ with U an open subset of $\mathbb{R}^n$, a continuously differentiable function on U. Assume that there exists a point $\mathbf{x}_0$ of U such that je Jacobian matrix of f at that point $Jf_{\mathbf{x}_0}$ est invertible. Then there exists an open subset $\mathbf{x}_0 \in V \subset U$ and an open subset $f(\mathbf{x}_0) \in W \in \mathbb{R}^n$ such that f is a* diffeomorphism *from V onto W, i.e there exists $g : W \to V$* continuously differentiable *with $g(f(\mathbf{x})) = \mathbf{x}, \forall \mathbf{x} \in V$ and $f(g(\mathbf{y})) = \mathbf{y}, \forall \mathbf{y} \in W$ and $Jg_{f(\mathbf{x})} = (Jf_{\mathbf{x}})^{-1}$.*

The most important consequence of this theorem is the Implicit Functions Theorem (these dtwo theorems are in fact logically equivalent). This theorem precises when an implicit relation between point can be transformed into a differentiable explicit one, and does provide locally the differential: if points $(x_1, \ldots, x_n, y_1, \ldots, y_k)$ satisfy simultaneously k relations of the form

$$f_i(x_1, \ldots, x_n, y_1, \ldots, y_k) = 0, \quad i = 1 \ldots k$$

when can we have

$$y_i = \phi_i(x_1, \ldots, x_n), \quad i = 1 \ldots k.$$

Theorem 11 Implicit Functions Theorem. *Let $f : U \in \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^k$, $f = (f_1, \ldots, f_k)$ be a continuously differentiable map. Assume that $\mathbf{a} = (a_1, \ldots, a_n)$ and $\mathbf{b} = (b_1, \ldots, b_k)$ satisfy $(\mathbf{a}, \mathbf{b}) \in U$ et $f(\mathbf{a}, \mathbf{b}) = 0$. Compute the partial Jacobian matrix with respect to the y_i de f at $(\mathbf{a}, \mathbf{b})$:*

$$J_{\mathbf{y}} f_{(\mathbf{a},\mathbf{b})} = \begin{pmatrix} \frac{\partial f_1}{\partial y_1}(\mathbf{a}, \mathbf{b}) & \ldots & \frac{\partial f_1}{\partial y_k}(\mathbf{a}, \mathbf{b}) \\ \vdots & \vdots & \vdots \\ \frac{\partial f_k}{\partial y_1}(\mathbf{a}, \mathbf{b}) & \ldots & \frac{\partial f_k}{\partial y_k}(\mathbf{a}, \mathbf{b}) \end{pmatrix}$$

this is a $k \times k$ matrix. If $J_{\mathbf{y}} f_{(\mathbf{a},\mathbf{b})}$ is invertible, there exists a neighborhood V of $\mathbf{a} \in \mathbb{R}^n$, a neighborhood W of $\mathbf{b} \in \mathbb{R}^k$ such that $V \times W \subset U$ and a unique continuoulsy differentiable map $g : V \to W$, $g = (g_1, \ldots, g_k)$ such that for all $\mathbf{x} \in V$, one has $f(\mathbf{x}, g(\mathbf{x})) = 0$. Moreover, if $J_{\mathbf{x}} f$ is the partial Jacobian matrix of f with respect to the x_j at such a point $(\mathbf{x}, \mathbf{y} = g(\mathbf{x}))$ and $Jg_{\mathbf{x}}$ is the jacobian matrix of g at $\mathbf{x}$,

$$J_{\mathbf{x}} f_{(\mathbf{x},\mathbf{y})} = \begin{pmatrix} \frac{\partial f_1}{\partial x_1}(\mathbf{x},\mathbf{y}) & \ldots & \frac{\partial f_1}{\partial x_n}(\mathbf{x},\mathbf{y}) \\ \vdots & \vdots & \vdots \\ \frac{\partial f_k}{\partial x_1}(\mathbf{x},\mathbf{y}) & \ldots & \frac{\partial f_k}{\partial x_n}(\mathbf{x},\mathbf{y}) \end{pmatrix}, \quad Jg_{\mathbf{x}} = \begin{pmatrix} \frac{\partial g_1}{\partial x_1}(\mathbf{x}) & \ldots & \frac{\partial g_1}{\partial x_n}(\mathbf{x}) \\ \vdots & \vdots & \vdots \\ \frac{\partial g_k}{\partial x_1}(\mathbf{x}) & \ldots & \frac{\partial g_k}{\partial x_n}(\mathbf{x}) \end{pmatrix}$$

which are $k \times n$ matrices, and one has

$$Jg_{\mathbf{x}} = -\left(J_{\mathbf{y}} f_{(\mathbf{x},\mathbf{y})}\right)^{-1} J_{\mathbf{x}} f_{(\mathbf{x},\mathbf{y})}.$$

A standard example: Set $f(x, y) = x^2 + y^2 - r^2$, the set $f(x, y) = 0$ is the circle of radius r and centre at $(0, 0)$. The partial Jacobian matrix with respect to y is simply is partial derivative

$$f_y = 2y.$$

If (x_0, y_0) is a point of the circle which does not lie on the horizontal plan $y = 0$ we get of course that $f_y(x_0, y_0) = 2y_0 \neq 0$. The implicit functions theorem then says that there exists a unique local parameterization $y = g(x)$ in a neighborhood of (x_0, y_0) with $f(x, g(x)) = 0$ and the derivative of g is given by

$$g'(x) = -\frac{1}{f_y(x,y)} f_x(x,y) = -\frac{1}{2y} 2x = -\frac{x}{y} = -\frac{x}{g(x)}$$

and g is a solution of the differential equation $g'g = -x$. But $g'g = \frac{1}{2}(g^2)'$ and $-x = -\frac{1}{2}(x^2)'$ therefore $g^2(x) = c - x^2$. Since $f(x, g(x)) = x^2 + g^2(x) = r^2$, one must have $c = r^2$ and the two solutions are

$$g(x) = \pm\sqrt{r^2 - x^2}$$

a not entirely unexpected result.

2.10 Second Variation, Hessian, Local Optimizers

Chapter 3

Infinite Dimension and Optimization of Functionals on $\mathbb{R}^n$

I will not try to write rigorous statments in this chapter, only pretend to do so one in a while, and I will allow myself a series of abuses (but still, reasonable abuses). Many, if not all, derivations of formulas will be done purely formally (or somewhat informally I should say).

3.1 Inner Products, Boundary Conditions,...

The general framework is the following: we are given an open set Ω of $\mathbb{R}^n$ (often $n = 2$)and a space of functions, denoted $L^2(\Omega)$, these functions are regular enough in the sense that the following integral always exists (it takes a finite value)

$$f, g \mapsto \langle f, g \rangle = \int_\Omega fg \, dx.$$

This operation defines a bilinear form, symmetric and positive definite on $L^2(\Omega)$, i.e. an inner product and one can show that this space is a Hilbert space. "bingo!", most of what we said on gradients and adjoints will still hold.. We may also restrict this function space to some particular subspaces, generally by imposing the existence of partial derivatives up to a certain order, something one may wish when one is interested in solving Partial Differential Equations. We also may wish to impose special propertiesin the case where Ω has a boundary $\partial\Omega$, along this boundary a vector field oriented towards the exterior of Ω, of norm 1 at each point (for the usual inner product in $\mathbb{R}^n$...) and orthogonal to the boundary (i.e. to the tangent line, plane....) along $\partial\Omega$). We will *almost always* denote this field by $\mathbf{n}$ in the sequel, $\mathbf{n} = (n_1, \ldots, n_n)^T$ and we call it *the*

exterior or outer normal along $\partial\Omega$. Among conditions on $\partial\Omega$, called *boundary conditions* we will be especiallyt interested to

- impose that a function takes predefined values along $\partial\Omega$ (Dirichlet), par exemple 0,
- impose that the directional derivative of a function in the direction de $\mathbf{n}$ along $\partial\Omega$ takes prescribed values (Neumann), often 0.

After I have obfuscated a bit the notion of gradient in the previous section, in the sequel, the gradient of a good old function from $U \subset \mathbb{R}^n \to \mathbb{R}$ will always be the good old one too. However for a functional, the gradient will be defined via the above inner product or a related one, in the case of vector valued functions it is easly extended, but I will at some point be led to consider inner products that include partial derivatives functions, in the framework of *Sobolev spaces*.

In the case of vector-valued functions, if $\mathbf{f}$ et $\mathbf{g}$ are functions with values in $\mathbb{R}^m$one can also define an inner product with the same notations

$$\langle \mathbf{f}, \mathbf{g} \rangle = \int_\Omega \mathbf{f} \cdot \mathbf{g}\, dx.$$

This is the integral of the inner productof the values of these functions,and functions for which this product is always finite will be denoted by $L^2(\Omega)^m$.

3.2 Classical Integration Results

Now we turn to a series of classical results of multivariate integral calculus: integration by parts, Green formulas, divergence theorem, they are essentially three formulations/applications of the Stoke formula.

In all the following results, we assume the functions we manipulate to be differentiable enough so we can forget about complicated functional analysis arguments.

Theorem 12 Integration by parts. *Let u and v be differentiable functions defined on $\Omega \cup \partial\Omega$. On has*

$$\int_\Omega \frac{\partial u}{\partial x_i} v\, dx = -\int_\Omega u \frac{\partial v}{\partial x_i}\, dx + \int_{\partial\Omega} u\, v\, \nu^i\, ds$$

where ds is "boundary integration element" and ν^i the i-th component of the outer normal (I will describe a bit more precisely in the Chen Vese section what this boundary integral is).

In dimension 1, if Ω is the interval (a, b) the outer normal at a is -1 and the outer normal at b is $+1$. In that case, the previous formula becomes

$$\int_a^b u' v\, dx = -\int_a^b uv'\, dx + u(b)v(b) - u(a)v(a)$$

and this is the classical integration par parts formula.

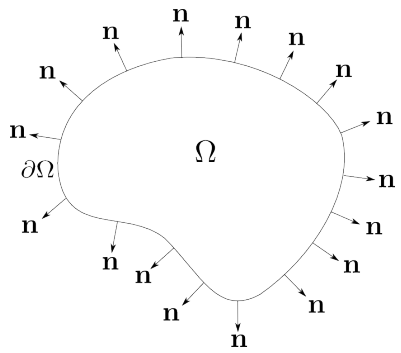

Figure 3.1: Domain Ω, its boundary $\partial\Omega$ and outer normals $\mathbf{n}$

Theorem 13 Green's Formulas. *They are equivalent to the prevous one (after a bity of work)*

1. $$\int_\Omega \Delta u \, dx = \int_{\partial\Omega} \frac{\partial u}{\partial \mathbf{n}} \, ds$$

   where $\frac{\partial u}{\partial \mathbf{n}}$ *is the directional derivative of* u *in the direction of* $\mathbf{n}$ *(it was denoted* $du_{\mathbf{x}}\mathbf{n}$ *or just* $du\mathbf{n}$ *in the previous section).*

2. $$\int_\Omega \nabla u \cdot \nabla v \, dx = -\int_\Omega u \Delta v \, dx + \int_{\partial\Omega} u \frac{\partial v}{\partial \mathbf{n}} \, ds$$

3. $$\int_\Omega u\Delta v - v\Delta u \, dx = \int_{\partial\Omega} u\frac{\partial v}{\partial \mathbf{n}} - v\frac{\partial u}{\partial \mathbf{n}} \, ds.$$

Theorem 14 Divergence Theorem. *Let* $F = (F_1, \ldots, F_n)^T$ *be a vector field defined on* $\Omega \cup \partial\Omega$ *and* $\operatorname{div} F := \sum_{i=1}^n \frac{\partial F_i}{\partial x_i}$ *(remark that* F *has as many components as the dimension of* Ω*)*

$$q \int_{\partial\Omega} F \cdot \mathbf{n} \, ds = \int_\Omega \operatorname{div} F \, dx.$$

In dimension 1, this is the "Fundamental theorem of Calculus"

$$F(b) - F(a) = \int_a^b F'(x) \, dx.$$

3.3 Adjunction

Next step: we start by rewriting the integration by parts formula with the help of our inner product of functions.

$$\langle \frac{\partial u}{\partial x_i}, v \rangle = \langle u, -\frac{\partial v}{\partial x_i} \rangle + \int_{\partial\Omega} uv\mathbf{n}_i \, ds.$$

Assume that we are interested only by functions which vanish along $\partial\Omega$ (they still form an Hilbert space), then the formula becomes

$$\langle \frac{\partial u}{\partial x_i}, v\rangle = \langle u, -\frac{\partial v}{\partial x_i}\rangle$$

and it says that the *adjoint of the linear map* $\frac{\partial}{\partial x_i}$ is $-\frac{\partial}{\partial x_1}$.

Next we look at the Gradient linear map *gradient*, $\nabla : u \mapsto \nabla u$, the good old gradient in $\mathbb{R}^n$. Let $\mathbf{v} = (v^1, \ldots, v^n)$ be a function with values in $\mathbb{R}^n$ (i.e. a function from $L^2(\Omega)^n$). I will from now denote (and it was time to) $\frac{\partial f}{\partial x_i} = f_{x_i}$. Utilising integration by parts for each of the variables x_i, $i = 1 \ldots n$,

$$\begin{aligned}
\langle \nabla u, \mathbf{v}\rangle &= \int_\Omega \nabla u \cdot v \, dx \\
&= \sum_{i=1}^n \int_\Omega u_{x_i} v^i \, dx \\
&= -\sum_{i=1}^n \int_\Omega u v^i_{x_i} \, dx + \sum_{i=1}^n \int_{\partial\Omega} u v^i \mathbf{n}_i \, ds \\
&= -\int_\Omega u \, \mathrm{div}(\mathbf{v}) \, dx + \int_{\partial\Omega} u \mathbf{v} \cdot \mathbf{n} \, ds.
\end{aligned}$$

Now, if u vanishes along $\partial\Omega$ or if $\mathbf{v}$ is orthogonal to $\mathbf{n}$ along $\partial\Omega$, the boundary integral is zero and one gets

$$\langle \nabla u, \mathbf{v}\rangle = \langle u, -\mathrm{div}(\mathbf{v})\rangle$$

i.e. in the "good function spaces" (good depends on boundary conditions) the adjoint of the gradient is $\nabla^* = -\mathrm{div}$ (note the minus sign).

3.4 Change of Variables Theorem

Before ending this chapter, I recall the Change of Variables theorem, in its simplest version to start with.

Theorem 15 *Let $\phi : [e, f] \to [a, b]$ be such that $\forall r \in [e, f]$, $\phi'(r) > 0$. Let $g : [a, b] \to \mathbb{R}$ an integrable function. Then*

$$\int_a^b g(x) \, dx = \int_e^f g(\phi(r))\phi'(r) \, dr.$$

The standard multidimensional version:

Theorem 16 *Let U and V be open subsets of $\mathbb{R}^n$, $\phi : U \to V$ a diffeomorphism. We will denote by $J\phi$ its Jacobian matrix and by $|J\phi|$ its Jacobian, i.e. the determinant of that matrix. Let $g : V \to \mathbb{R}$ an integrable function, then*

$$\int_V g(\mathbf{x}) \, d\mathbf{x} = \int_U g(\phi(\mathbf{y}))|J\phi_{\mathbf{y}}| \, d\mathbf{y}.$$

Chapter 4

Applications to the Optimization of Functionals for Domains of $\mathbb{R}^n$

In this chapter, we collect a series of standard examples of functionals used in Image Analysis and sometimes elsewhere, and the corresponding gradients and Euler-Lagrange equations. Here too the treatment is largely informal, and most questions regarding function spaces and other existence stuffs are completely forgotten.

4.1 Tikhonov Regularization

We will first look at *Tikhonov Regularization* [**?**], or merely a special form of it, used in denoising. It consists in finding the minimizer of the following energy

$$E(v) = \int_\Omega \left((v - u_0)^2 + \lambda |\nabla v|^2\right)\, dx = 0.$$

where $|\nabla v|^2 = \nabla v \cdot \nabla v$. A minimizer must be close enough of u_0 while being smooth. As in the finite dimensional case, a necessary condition for a function u to minimize E,i.e. $E(u)$ is a (local) minimum, is that the gradient of E must vanish at that point: $\nabla E_u = 0$, this is the l'*Euler-Lagrange Equation* associated to E.

We will obtain it with the method of directional derivatives. Set $\ell(t) = E(v + th)$ and compute $\ell'(0)$. First note that if $\mathbf{f}(t) : \mathbb{R} \to \mathbb{R}^2$, then

$$\frac{d}{dt}\mathbf{f} \cdot \mathbf{f} = 2\frac{d\mathbf{f}}{dt} \cdot \mathbf{f}$$

(it is a trivial generalization of the rule $(f^2)' = 2ff'$). A direct computation then gives ($\int$ et $\frac{d}{dt}$ can always be interchanged when the functions under con-

sideration are "nice enough")

$$\begin{aligned}\ell'(t) &= \int_\Omega \frac{d}{dt}\left\{(v+th-u_0)^2+\lambda|\nabla(v+th)|^2\right\}\,dx \\ &= 2\int_\Omega \left\{(v+th-u_o)h+2\lambda\nabla h\cdot\nabla(v+th)\right\}\,dx\end{aligned}$$

and at $t=0$ one gets

$$\ell'(0) = 2\int_\Omega \left\{(v-u_0)h+\lambda\nabla h\cdot\nabla v\right\}\,dx.$$

Recall either Green's second formula or the computation we just did for the adjonction gradient/divergence. Assume for instance that we only are interested to the class of functions vs such that $\nabla v\cdot\mathbf{n}=\frac{\partial v}{\partial\mathbf{n}}=0$, the so called *Neumann boundary conditions*, we then can use the adjonction property so as to get

$$\ell'(0) = 2\int_\Omega \left\{(v-u_0)-\lambda\mathrm{div}(\nabla v)\right\}h\,dx.$$

But $\mathrm{div}(\nabla v)=\Delta v$ and one has

$$\ell'(0) = 2\int_\Omega (v-u_0-\lambda\Delta v)h\,dx = \langle 2(v-u_0-\lambda\Delta v),h\rangle$$

which means precisely that the gradient of E at v is $2(v-u_0-\lambda\Delta v)$ and that the condition for a fonction u that satisfies the boundary condition $\frac{\partial u}{\partial\mathbf{n}}=0$ on $\partial\Omega$ minimizes the Tikhonov regularization functional is

$$u-u_0-\lambda\Delta u=0, \qquad \frac{\partial u}{\partial\mathbf{n}_{|\partial\Omega}}=0.$$

Note that I have included the boundary condition $\frac{\partial u}{\partial\mathbf{n}_{|\partial\Omega}}=0$, it indicates indeed the hypothese we provided in order to have an adjunction. Moreover a boundary condition such as that one is necessay in order to solve the Euler-Lagrange Equation. A carefull look shows that this calulation is, especially in its gradient/divergence part, analogous to the calculation of gradient made in section 2.4: take $\mathbf{f}(\mathbf{y})=\lambda\mathbf{y}\cdot\mathbf{y}$ and $g(v)=\nabla v$. In this example g is linear, therefore equal to its differential. $\mathbf{f}$ is quadratic, its differential is thus linear, the resulting equation is linear. lineaire. In fact, every functional/energy purely quadratic has a linear gradient. It is true in finite dimension, the gradient of an homogeneous polynomial of degree 2 is a linear expression, and it is still true in infinite dimension.

4.2 Rudin-Osher-Fatemi Model, BV Regularization

Tikhonov regularization is generaly not well suited for image denoising as is imposes a too strong constraint on the image gradient ∇u: large gradients

are penalized too much via the quadratic term $|\nabla u|^2$ and the resulting image present generally s too smooth to be acceptable. BV Regularization ("Bounded Variations"), first proposed by Rudin, Osher and Fatemi in [**?**] proposes a weaker constraint by penalizing the gradient norm $|\nabla u|$ instead of its square. The corresponding functional is

$$F(v) = \int_\Omega \left\{(v-u_0)^2 + \lambda|\nabla v|\right\}\, dx.$$

Before starting the actual computation of the gradient, a digression seems necessary in order to explain the difference between Tikhonov regularization and BV regularization.

The total variation of a function is defined by

$$TV(u) = \int_\Omega |\nabla u|\, dx.$$

where $|\nabla u| = \sqrt{\nabla u \cdot \nabla u}$. A function u is sait to have bounded variation, BV if $TV(u) < \infty$. In fact total variation can be defined for functions which have a "serious" differentiability problem via the concept of Radon measure,but, I said I would not dig into Functional Analysis, so I will not continue in that direction. I will propose instead a dimension 1 example that will allow to clarify the concept as well as show the difference with the term

$$\int_\Omega |\nabla u|^2\, dx.$$

We will use a very simple function, a modified Heaviside function defined on $\mathbb{R}$ by

$$H(x) = \begin{cases} 0 & \text{si } x < 0 \\ a & \text{sinon.} \end{cases}$$

with a a real. Which sense can we give to $TV(H)$ since there is a serious discontinuity problem at 0, and thus $H'(0)$ is not defined? Integration removes in some sense the problem! Let us choose $h > 0$, and write (a bit formally) to start with

$$TV(H) = \int_{-\infty}^{\infty} |H'(x)|\, dx = \int_{-\infty}^{-\frac{h}{2}} |H'(x)|\, dx + \int_{-\frac{h}{2}}^{-\frac{h}{2}} |H'(x)|\, dx + \int_{\frac{h}{2}}^{\infty} |H'(x)|\, dx$$

On $]-\infty, -h/2]$ and $[h/2, \infty[$, H is constant, its derivative vanishes and therefore

$$TV(H) = \int_{-\frac{h}{2}}^{-\frac{h}{2}} |H'(x)|\, dx.$$

When $h \to 0$ the remaining term is

$$TV(H) = \lim_{h\to 0} \int_{-\frac{h}{2}}^{-\frac{h}{2}} |H'(x)|\, dx.$$

We replace the pseudo-derivative (it exists in fact as an object called a *distribution*, but not as a function) by the unsurprising approximation

$$\frac{H(h/2) - H(-h/2)}{h}.$$

Indeed, if H was differentiable at 0, this approximation would converge, basically by definition, toward $H'(0)$! With the choosen definition of H, this approximation takes the value a/h. Replacing it into the integral, one gets

$$TV(h) = \lim_{h\to 0} \int_{-\frac{h}{2}}^{-\frac{h}{2}} \frac{|H(h/2) - H(-h/2)|}{h}\, dx = \lim_{h\to 0} \int_{-\frac{h}{2}}^{-\frac{h}{2}} \frac{|a|}{h}\, dx = \lim_{h\to 0} h\frac{|a|}{h} = |a|$$

and the total variation H is the *jump height* at 0. If one would remake the computation for

$$\int_{-\infty}^{\infty} |H'(x)|^2\, dx$$

with the same approximations

$$\int_{-\infty}^{\infty} |H'(x)|^2\, dx = \lim_{h\to 0} h\frac{|a|^2}{h^2} = \lim_{h\to 0} \frac{|a|^2}{h} = \infty.$$

There is even more with the BV model. Instead of the function H above, let us consider the "nicer" edge function

$$H_\epsilon(x) = \begin{cases} 0 & x <= -\epsilon \\ \frac{ax}{2\epsilon} + \frac{a}{2} & -\epsilon < x < \epsilon \\ a & x > \epsilon \end{cases}$$

This is illustrated in figure 4.1.

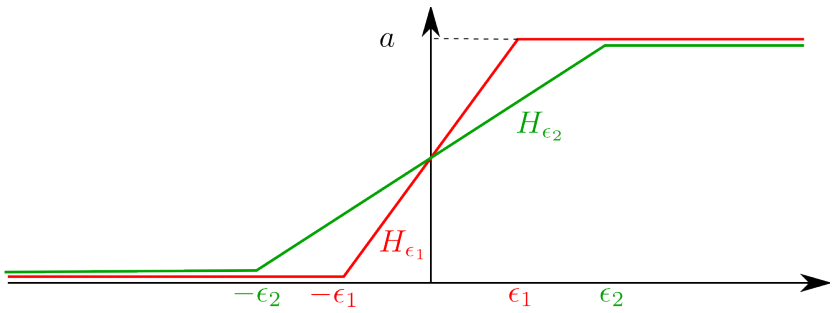

Figure 4.1: Two edge functions H_ϵ for different values of ϵ.

By direct integration (left to the reader) one gets that

$$\int_{\infty}^{\infty} |H'_\epsilon(x)|\, dx = |a|$$

and depends only on the edge height and not on the edge slope $a/2\epsilon$, The same calculation for the L^2 gradient norm gives

$$\int_{\infty}^{\infty} |H'_\epsilon(x)|^2 \, dx = \frac{a^2}{2\epsilon}$$

and decreases with increasing ϵ, i.e. with flatter edges. Tichonov regularization would favor these types of flatter edges, thus blurring contours, while TV regularization disregards slope information.

Conclusion (a bit fast, but it really holds): Solutions of Tikhonov regularization problem cannot have discontinuities, while solutions of BV regularization problems are generally associated to edges/s. Solutions of Tichonov regularization prefer blurred edges to crisper ones. Solutions of the BV problem do not a priori distinguish between them and this can be decided by the data, for instance.

The BV problem is thus much more interesting for contour preservation.

Let us compute now the gradient of $F(v)$. Here too we will use the method of directional derivatives. Define the function $\ell(t) = F(v + th)$. Then one has, using this time the rule

$$(f^{1/2})' = \frac{1}{2} f^{-1/2} f'$$

$$\begin{aligned}
\ell'(t) &= \int_\Omega \frac{d}{dt} \left\{ (v + th - u_0)^2 + \lambda (\nabla(v + th) \cdot \nabla(v + th))^{1/2} \right\} dx \\
&= \int_\Omega \left\{ 2(v + th - u_0)h + \lambda \frac{d}{dt} (\nabla(v + th) \cdot \nabla(v + th))^{1/2} \right\} dx \\
&= \int_\Omega \left\{ 2(v + th - u_0)h + \frac{\lambda}{2|\nabla(v + th)|} \frac{d}{dt} \left(\nabla(v + th) \cdot \nabla(v + th) \right) \right\} dx \\
&= \int_\Omega \left\{ 2(v + th - u_0)h + \lambda \left(\nabla(v + th) \cdot \nabla(v + th) \right)^{-1/2} \nabla h \cdot \nabla(v + th) \right\} dx
\end{aligned}$$

and for $t = 0$, one gets

$$\begin{aligned}
\ell'(0) &= \int_\Omega \left\{ 2(v - u_0)h + \lambda |\nabla v|^{-1} \nabla h \cdot \nabla v \right\} dx \\
&= \int_\Omega \left\{ 2(v - u_0)h + \lambda \nabla h \cdot \frac{\nabla v}{|\nabla v|} \right\} dx.
\end{aligned}$$

We need to tranform the term containing ∇h into a term containing h

$$\int_\Omega \nabla h \cdot \frac{\nabla v}{|\nabla v|} \, dx.$$

In order to get rid of this gradient, we use the l'adjunction gradient / divergence, with boundary conditions $\frac{\partial v}{\partial \mathbf{n}} = 0$ $\partial\Omega$ so as to obtain

$$\int_\Omega \nabla h \cdot \frac{\nabla v}{|\nabla v|} \, dx = - \int_\Omega h \, \mathrm{div} \left(\frac{\nabla v}{|\nabla v|} \right) dx$$

and, gathering the pieces,

$$\ell'(0) = \int_\Omega \left\{ 2(v-u_0) - \lambda \operatorname{div}\left(\frac{\nabla v}{|\nabla v|}\right)\right\} h\, dx = \langle 2(v-u_0) - \lambda \operatorname{div}\left(\frac{\nabla v}{|\nabla v|}\right), h\rangle$$

i.e. the gradient of F at v is $\nabla F_v = 2(v-u_0) - \lambda \operatorname{div}\left(\frac{\nabla v}{|\nabla v|}\right)$ and a minimizer u of F must satisfy the Euler-Lagrange equation, setting $\lambda' = 2\lambda$ and forgetting the "' ":

$$u - u_0 - \lambda \operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right) = 0, \quad \frac{\partial u}{\partial \mathbf{n}}_{|\partial\Omega} = 0.$$

This computation too corresponds to the one performed in section 2.4: one takes $\mathbf{f}(\mathbf{y}) = \lambda\sqrt{\mathbf{y}\cdot\mathbf{y}}$ and $g(v) = \nabla v$. In this example g is linear, therefore equal to its differential. This time $\mathbf{f}$ is more complex than in the Tikhonov regularization case, its differential toand the resulting equation is very non linear.

In reality, it is not only non linear, it is also non smooth, and optimality consition should be defined in term of subdifferential and subgradient. The interested reader may consult [**?**] or [**?**].

4.3 A Somewhat General Lagrangian

In this paragraph we look at the minimization problem

$$\arg\min_v E(v) = \int_\Omega L(x, v, v_{x_1}, \ldots, v_{x_n})\, dx$$

where $L : x \in \mathbb{R}^n, p \in \mathbb{R}, q_1 \in \mathbb{R}, \ldots, q_n \in \mathbb{R} \mapsto L(x, p, q_1, \ldots, q_n)$ is called the *Lagrangian* of the functional E. We will comnpute its gradient via the method of directional derivatives. As we did before, set $\ell(t) = E(v + th)$ and let us calculate $\ell'(0)$. Omitting a few details, one can check that

$$dEv = \ell'(0) = \int_\Omega \left\{ L_p(x, v, v_x, v_y)h + \sum_{i=1}^n L_{q_i}(x, v, v_{x_1}, \ldots, v_{x_n})h_{x_i} \right\} dx. \tag{4.1}$$

This expression is clearly linear in h but with partial derivatives. In order to obtain the gradient, we must get rid of them and we use the integration by part formule, which here reads as $(x, v.v_{x_1} \ldots, v_{x_n})$:

$$\begin{aligned} dEv &= \int_\Omega \left\{ L_p - \sum_{i=1}^n \frac{\partial}{\partial x_i} L_{q_i} \right\} h\, dx + \int_{\partial\Omega} h \sum_{i=1}^n L_{q_i}\mathbf{n}_i\, ds \\ &= \int_\Omega \left\{ L_p - \sum_{i=1}^n \frac{\partial}{\partial x_i} L_{q_i} \right\} h\, dx + \int_{\partial\Omega} h \begin{pmatrix} L_{q_1} \\ \vdots \\ L_{q_n} \end{pmatrix} \cdot \mathbf{n}\, ds \end{aligned}$$

If we are in a space of functions that vanish along $\partial\Omega$ or if $\begin{pmatrix} L_{q_1} \\ \vdots \\ L_{q_n} \end{pmatrix} \cdot \mathbf{n}$ is zero on $\partial\Omega$, the boundary term is 0 and the sought gradient is

$$\begin{aligned} \nabla E_v &= L_p(x, v.v_{x_1} \ldots, v_{x_n}) - \sum_{i=1}^{n} \frac{\partial}{\partial x_i} L_{q_i}(x, v.v_{x_1} \ldots, v_{x_n}) \qquad (4.2) \\ &= L_p(x, v.v_{x_1} \ldots, v_{x_n}) - \\ &\quad \sum_{i=1}^{n}\sum_{j=1}^{n} L_{qq_j}(x, v.v_{x_1} \ldots, v_{x_n}) v_{x_j} \\ &\quad \sum_{i=1}^{n}\sum_{j=1}^{n} L_{q_i q_j}(x, v, v_{x_1} \ldots, v_{x_n}) v_{x_i x_j}. \qquad (4.3) \end{aligned}$$

The Euler-Lagrange equation of a minimizer of u is then

$$L_p(x, u, u_{x_1}, \ldots, u_{x_n}) - \sum_{i=1}^{n} \frac{\partial}{\partial x_i} L_{q_i}(x, u, u_{x_1}, \ldots, u_{x_n}) = 0$$

with corresponding boundary conditions. If one sets

$$F(x, u, \nabla u) = \begin{pmatrix} L_{q_1}(x, u, u_{x_1}, \ldots, u_{x_n}) \\ \vdots \\ L_{q_n}(x, u, u_{x_1}, \ldots, u_{x_n}) \end{pmatrix},$$

the associated Euler-Lagrange equation can be written as

$$L_p(x, u, u_{x_1}, \ldots, u_{x_n}) - \operatorname{div}(F(x, u, \nabla u)) = 0$$

and one says that it is in *conservative form*. The conservative form is interesting in that is indeed expresses certain conservation properties (but I will not say which ones!). Important for us is that the writing with the help of a divergence may have consequences for the numerical schemes developed to solve the Euler-Lagrange equation (see my notes "Sur la réesolution de certaines EDP en Analyse d'Images" which will be translated into English soon (?)).

A remark which has its importance When the differential is expressed via (4.1), v appears in dE with derivatives with the same order as in E, i.e. 1. When the gradient formulation (4.3) is used, v appears with second order derivatives. It really means that solutions of the minimization problems above need not be twice differentiable!

4.4 Heat Equation and Calculus of Variations

In this paragraph, I present two ways of deriving the Heat Equation from variational methods. The first is a gradient descent on an erergy term which is

precisely the regularization term of Tikhonov regularization. The second, more abstract, is also Tikhonov like, with derivatives up to order ∞ in the regularization term and follows the computation of Nielsen *et al.* in [**?**].

In the first derivation, we are interested to the minimization of energy:

$$E(v) = \int_\Omega |\nabla v|^2 \, dx$$

by gradient descent. By redoing verbatim some of the computations performed in the Tikhonov regularization paragraph, the gradient of this energy is, with proper boundary conditions,

$$\nabla E_v = \nabla^* \nabla v = -\mathrm{div}\,(\nabla v) - = \Delta v.$$

and the Euler-Lagrange equation associated to it is $-\Delta u = 0$, this is the *Laplace equation*. Solutions depend deeply on boundary conditions. If one requires that $u_{|\partial\Omega} = 0$, the corresponding solution is $u \equiv 0$. In general, Laplace equation is associated with a Dirichlet boundary condition $u_{|\partial\Omega} = f$ for some prescribed function f defined along $\partial\Omega$. This is one of the most well studied PDE, especially because of its usefulness in a large variety of problems. The solutions the Laplace Equation are called *harmonic functions* and possess many regularity properties, in particular they are *analytic*, i.e. admit and are equal to their Taylor series expansion in a neighborhood of every point of Ω (thus indefinitly differentiable, but harmonicity is more), see [**?**] for instance. They provide an analytic prolongation of f in Ω.

If general Neumann boundary conditions are imposed, instead of Dirichlet, i.e.

$$\left.\frac{\partial u}{\partial \mathbf{n}}\right|_{\partial\Omega} = f$$

there may be no solutions or multiple ones. Using Green's formula (13) (the second one) we find indeed that

$$0 = \int_\Omega \Delta u \, dx = \int_{\partial\Omega} \frac{\partial u}{\partial \mathbf{n}} \, ds = \int_{\partial\Omega} f \, ds.$$

Thus the condition $\int_{\partial\Omega} f \, ds = 0$ is necessary for solutions to exist. When a solution exists, then infinitely many do, by adding any real constant. In fact, at least when Ω has some nice symmetries, then the solutions for $f \equiv 0$ are exactly the constant functions. In our imaging problems, it will always be the case.

We now choose these Neumann conditions and consider the gradient descent approach for the Heat Equation. From an initila value u_0, aradient descent builds a "time dependent" function $u : \Omega \times [0, \infty)$ with $u(\mathbf{x}, 0) = u_0(\mathbf{x})$, which satisfies the following equation

$$u_t = -\Delta u, \quad u(-, 0) = u_0. \tag{4.4}$$

where the Laplacian is computed with respect to spatial variables. This is exactly the Heat equation with initial value $u(-, 0) = u_0$. Its value at $t = \infty$

solves $-\Delta u = 0$, and is therefore a constant from what I said above. This constant is the mean value

$$\frac{1}{|\Omega|}\int_\Omega u_0\,dx$$

where $|\Omega|$ denotes the area (length, volume...) of Ω. This comes from the following: integrate (4.4) and use the first Green's formula (Theorem 13-2).

$$\frac{1}{|\Omega|}\int_\Omega u_t\,dx = -\frac{1}{|\Omega|}\int_\Omega \Delta u\,dx = -\frac{1}{|\Omega|}\int_{\partial\Omega}\frac{\partial u}{\partial \mathbf{n}}\,ds = 0$$

because of our boundary conditions. But

$$0 = \frac{1}{|\Omega|}\int_\Omega u_t\,dx = \frac{d}{dt}\left(\frac{1}{|\Omega|}\int_\Omega u\,dx\right)$$

which means that the mean value remains constant.

We consider now the second approach. This was proposed in dimension 1 by Nielsen *et al.* in [**?**] and a straightforward generalization to dimension 2 was given in [**?**]. I present that one in details. The domain Ω here is $\mathbb{R}^2$ and we work with $\mathcal{C}^\infty$ functions that are regular in the sense that the following energy is always *finite*

$$E(u) = \frac{1}{2}\left(\iint (u-u_0)^2\,dxdy + \sum_{k=1}^{\infty}\frac{\lambda^k}{k!}\iint\sum_{\ell=0}^{k}\left(\binom{k}{\ell}\frac{\partial^k u}{\partial x^\ell \partial y^{k-\ell}}\right)^2 dxdy\right) \tag{4.5}$$

where $\binom{k}{\ell}$ are the binomial coefficients (the space of such function is non-empty, it is in fact infinite-dimensional as it contains, for instance, all the smooth functions with compact support).

In computing its gradient, the problem lies of course in the regularizer part. Let us set

$$\mathcal{L}_k(u) = \frac{\lambda^k}{k!}\iint\sum_{\ell=0}^{k}\left(\binom{k}{\ell}\frac{\partial^k u}{\partial x^\ell \partial y^{k-\ell}}\right)^2 dxdy$$

and compute

$$\frac{d}{dt}\mathcal{L}_k(u+tv)\bigg|_{t=0}$$

i.e. we still use the method of directional derivatives. We use the following fact, whose proof is of course left to the reader: In our function space, we have the following adjunction result: if

$$T^{m,n} := \frac{\partial^{m+n}}{\partial x^m \partial y^n} \quad \text{then} \quad T^* = (-1)^{m+n}T^{m,n}.$$

Thanks to this formula, one get

$$\frac{d}{dt}\mathcal{L}_k(u+tv)\bigg|_{t=0} = \frac{(-\lambda^k)}{k!}\iint\left(\sum_{\ell=0}^{k}\binom{k}{\ell}T^{2k,2\ell}u\right)v\,dxdy$$

Then the reader can check the identity

$$\sum_{\ell=0}^{k} \binom{k}{\ell} T^{2k,2\ell} u = \underbrace{\Delta \circ \cdots \circ \Delta}_{k \text{ times}} u := \Delta^{(k)} u$$

($\Delta^{(k)}$ is the k-th iterated Laplacian). Putting the pieces toghether, we get that

$$\begin{aligned} \frac{d}{dt} E(u+tv)\Big|_{t=0} &= \langle u - u_0 + \sum_{k=1}^{\infty} \frac{(-\lambda)^k}{k!} \Delta^{(k)} u, v \rangle \\ &= \langle -u_0 + \sum_{k=0}^{\infty} \frac{(-\lambda)^k}{k!} \Delta^{(k)} u, v \rangle. \end{aligned}$$

We have therefore the sought gradient as the first argument of the above inner product. Recall a few general facts about the exponentional of an element tA in a Banach algebra (like square matrices for instance, where product and composition are the same), t being a real number,

1. Power series: $e^{tA} = \sum_{k=0}^{\infty} \frac{t^k A^k}{k!}$
2. Inverse: $\left(e^{tA}\right)^{-1} = e^{-tA}$
3. Derivative: $d/dt e^{tA} = A e^{tA}$.

With the first formula, we can rewrite, at least formally

$$\nabla_u E = e^{-\lambda \Delta} u - u_0.$$

For the gradient to be null, we deduce, from the second formula that the solution u, which thus depends on λ must satifsy

$$u := u(\lambda) = e^{\lambda \Delta} u_0$$

from which we get $u(0) = u_0$, and computing the derivative with respect to λ, we get

$$\frac{\partial u}{\partial \lambda} = \Delta e^{\lambda \Delta} u_0 = \Delta u$$

and we find once again the heat equation! Althout apparently purely formal, this derivation is more meaningful that it may seem. It is known indeed that the solution u can be written as a convolution with a Gaussian kernel:

$$u = G_\lambda * u_0, \quad G_\lambda(x,y) = \frac{1}{2\pi\lambda} e^{-\frac{x^2+y^2}{2\lambda}}.$$

However, when discretizing the problem, the Laplace operator becomes a matrix M, and the exponential of tM provides a good approximation of Gaussian kernel (see my notes "Sur la réesolution de certaines EDP en Analyse d'Images" which will be translated into English soon (?)).

To conclude that paragraph, let us add that the computation performed here in dimension 2 extends to higher dimensions in a straightforward way, we need just to replace binomial coefficients and expansion by more general multinomial coefficients and expansion.

4.5 Optical Flow à la Horn and Schunck

Recovery of apparent motion is a standard problem in image anaysis. In its simplest formulation, two images are given u_1 and u_2 and we want to determine a best way of deforming u_1 into u_2, knowing that these images come from the observation of the same dynamic scene at to differenr time instants.

Intensity preservation during displacement, sometimes referred to as the *Lambertian Assumption*, is the simplest hypothesis We may look for a displacement field d that encodes this preservation:

$$u_1(x) = u_2(x + d(x)).$$

By setting $u_1(x) = u(x,t)$, $u_2(x) = u(x,t+dt)$ we have $u(x+d(x),t+dt) = u(x,t)$. Then for small dt the displacement $d(x)$ can be approximated as $dt\, v(x)$ where $v(x) = v(x,t)$ is the instantaneous velocity of x at time t – the approximation comes from Taylor expansion on as the displacement tends to 0 when $dt \to 0$, and by performing a Taylor expansion, this can be approximated as

$$u(x + v(x), t + dt) - u(x,t) \approx dt(\nabla u \cdot v + u_t) \approx 0$$

where ∇u is the gradient of u w.r.t spatial variables and u_t is the time derivative of u. The equation

$$\nabla u \cdot v + u_t = 0 \tag{4.6}$$

is called the *Optical Flow Constraint Equation* (OFC or OFCE in the sequel). At each pixel/image location, it provides at most one constraint, and is thus insufficient to determine the full velocity/displacement field. This is the so called aperture problem.

A first solution has been proposed by Horn and Schunck in [**?**] and consists in solving the OFCE in a least square sense while imposiing regularity to the velocity field $v = (v^x, v^y)$, v^x and v^y being the x and y components of the velocity vectors. Horn and Schunck is thus

$$E_{HS}(v) = \frac{1}{2}\int_\Omega (\nabla u \cdot v + u_t)^2 \, dx + \frac{\lambda}{2}\int_\Omega \left(|\nabla v^x|^2 + |\nabla v^y|^2\right) dx. \tag{4.7}$$

We compute its gradient via the usual directional derivatives technique, by setting $\ell(s) = E_{HS}(v + sw)$ where w is a *vector field* of deformations of v. The only "difficulty" comes from the fact that now v and w are vector valued. A straightforward computation gives

$$\begin{aligned}
\ell'(0) &= \int_\Omega (\nabla u \cdot v + u_t)\, \nabla u \cdot w \, dx + \lambda \int_\Omega (\nabla v^x \cdot \nabla w^x + \nabla v^y \cdot \nabla w^y)\, dx \\
&= \int_\Omega \left[(\nabla u \cdot v + u_t)\, \nabla u - \lambda \begin{pmatrix} \Delta v^x \\ \Delta v^y \end{pmatrix}\right] \cdot w + \int_{\partial\Omega} w \cdot \begin{pmatrix} \frac{\partial v^x}{\partial \mathbf{n}} \\ \frac{\partial v^y}{\partial \mathbf{n}} \end{pmatrix} ds
\end{aligned}$$

In order to get a gradient, boundary conditions are needed, and choosing variations w that vanish at the boundary means Dirichlet conditions for v, it could

be for instance null motion, if one knows for instance that the camera is fixed, the background is static, and projected moving object do not cross the image boundary. We may instead ask that the normal derivatives of the flow components vanish at the boundary, this means some sort of constant motion, for instance a translational motion of the camera parallel to the imaging plane. None of them is completly natural, but this is it...

Choose one of them in order to eliminate the boundary term and the Euler-Lagrange equation is

$$\nabla_v E_{HS} = (\nabla u \cdot v + u_t) \nabla u - \lambda \begin{pmatrix} \Delta v^x \\ \Delta v^y \end{pmatrix} = 0.$$

This equation is only valid for smooth enough images or small displacements. When considering large displacements, a standard solution is to use multiresolution framework. Although widely criticised as a poor quality motion estimator, interestingly, modern discretizations for the Horn and Schunck approach provide decent solutions for the optical flow problem. **[François dit: I should add a reference / link to the Middlebury Optical Flow Evaluation Database]**.

4.6 Non Linear Flow Recovery

[François dit: I will probably have time to write some thing some day, but not now, sorry.]

Chapter 5

Planar curves

In this chapter, I will recall facts about the geometry of planar curves: length, arc length parametrization, curvature, Frenet formulas and the ideas that allow us to define the notions of inner product and gradient, adapted to the problems of optimization of curves.

A parametrized planar curve is a mapping

$$c : [a, b] \to \mathbb{R}^2, \quad p \mapsto c(p) = (x(p), y(p)).$$

The "geometric curve" is the corresponding subset $c([a, b])$ of the plane, together with an orientation. If $A = c(a) \neq c(b) = B$, then c is an *open* curve with endpoints A and B. if $A = B$, then c is a closed curve. If c is differentiable, then we say that it is regular if its velocity vector c_p never vanishes.

From now on we will, unless otherwise explicitly stated, focus on the case where c is continuously differentiable. This is not strictly necessary, but it makes life easier.

5.1 Moving frames

For such a smooth curve we define a *moving frame*: To each point $P = c(p)$, we attach a planar frame defined by the tangent and normal vector to the curve, such that whenever $c_p \neq 0$, we denote by

$$\vec{\tau}(p) = \frac{c_p}{|c_p|} = \frac{1}{\sqrt{x_p^2 + y_p^2}} \begin{pmatrix} x_p \\ y_p \end{pmatrix}$$

the normalized tangent velocity vector, and we define

$$\vec{\nu}(p) = \frac{1}{\sqrt{x_p^2 + y_p^2}} \begin{pmatrix} -y_p \\ x_p \end{pmatrix}.$$

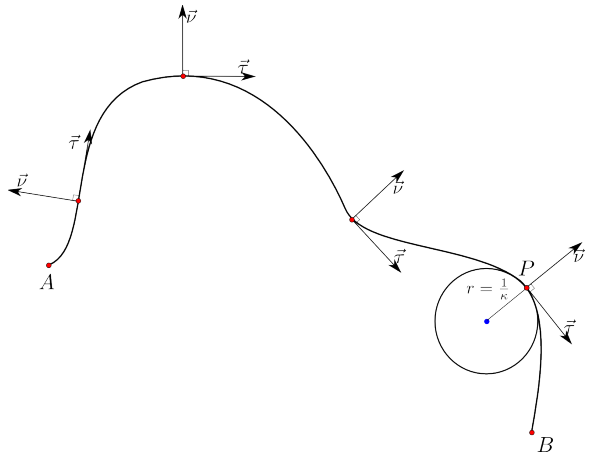

Figure 5.1: A curve and its moving frame, the osculating circle at a point P, and the radius of curvature r.

5.2 Length, arc-length parametrization, curvature

The length of c is given by

$$\ell(c) = \int_a^b |c_p|\, dp.$$

We can, for example, approximate the curve segments as shown in figure 5.2. A curve is an *arc length parametrization*, or for short, an *a-l parametrization*

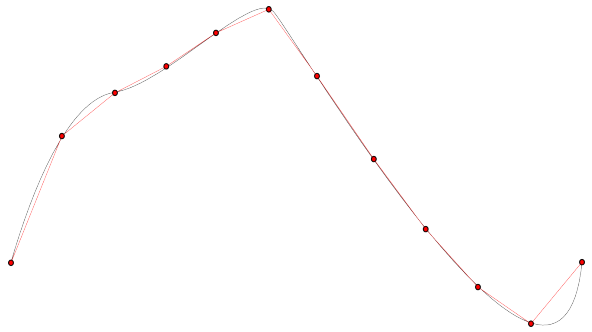

Figure 5.2: Approximation of segments and length.

if $|c_p| \equiv 1$, that is, if $c_p = \vec{\tau}$. If c is regular, it always admits an arc length parametrization. As a consequence, writing L for the length of c, the map

$$\sigma : [a, b] \to [0, L], \quad \sigma(p) = \int_a^p |c_p|\, dp$$

is invertible because $\sigma_p = |c_p| > 0$. Denote by π its inverse mapping; then $\tilde{c} : s \mapsto c(\pi(s))$ is an arc-length parametrization by the differentiation rules for composed maps. In general we forget about the $\tilde{}$ and distinguish between the two parametrizations by their arguments. We use the argument p for arbitrary parametrizations and reserve the argument s for a-l parametrizations, for which $\vec{\tau}(s) = c_s(s)$.

From the last formula, we see that

$$\langle c_s, c_s \rangle \equiv 1 \quad \Longrightarrow \quad \frac{d}{ds}\langle c_s, c_s \rangle = 0 = 2\langle c_{ss}, c_s \rangle$$

and thus c_{ss} is orthogonal to $\vec{\tau}$. Thus there exists a real number κ such that $c_{ss} = \kappa\vec{\nu}$. We call $\kappa = \kappa(s)$ the *algebraic curvature* of s (or rather $c(s)$). Note that its inverse, $1/\kappa(s)$ is the radius of the *osculating circle* of s, that is the largest circle which is tangent to c at s and located in the "concavity" of c at $c(s)$, i.e. it is the best approximation of c by a circle at that point. See figure 5.1.

Since $|\vec{\tau}| = 1$, only the direction of $\vec{\tau}$ is varying, and the curvature indicates the speed of its angular variation. In the case where c is a circle centered at $(0, 0)$ with radius r, c is its own osculating circle of curvature at each point, and we expect the curvature at any point to be $1/r$. We can verify this: The a-l parametrization of c is simply

$$s : [0, 2\pi r] \mapsto (r\cos\frac{s}{r}, r\sin\frac{s}{r}).$$

and

$$c_s = \begin{pmatrix} -\sin\frac{s}{r} \\ \cos\frac{s}{r} \end{pmatrix}, \quad c_{ss} = \frac{1}{r}\begin{pmatrix} -\cos\frac{s}{r} \\ -\sin\frac{s}{r} \end{pmatrix} = \frac{1}{r}\vec{\nu}$$

and the algebraic curvature is exactly $1/r$!

Because $\vec{\nu}$ is obtained by rotating $\vec{\tau}$ by an angle of $\pi/2$, we obtain the following relations, which are known as the Frenet relations:

$$\frac{d\vec{\tau}}{ds}(s) = \kappa(s)\vec{\nu}(s) \tag{5.1}$$

$$\frac{d\vec{\nu}}{ds}(s) = -\kappa(s)\vec{\tau}(s). \tag{5.2}$$

The curvature $\kappa(s)$ is simply given by

$$\kappa(s) = y_{ss}x_s - x_{ss}y_s. \tag{5.3}$$

For a curve which is not an a-l parametrization, we can use the theorem of change of variables to change the original parametrization to an arc-length parametrization, which gives:

$$\frac{1}{|c_p|}\frac{d}{dp}\vec{\tau}(p) = \kappa(p)\vec{\nu}(p) \tag{5.4}$$

$$\frac{1}{|c_p|}\frac{d}{dp}\vec{\nu}(p) = -\kappa(p)\vec{\tau}(p) \tag{5.5}$$

The algebraic curvature is then given by

$$\kappa(p) = \frac{y_{pp}x_p - x_{pp}y_p}{|c_p|^3} = \frac{\det(c_{pp}, c_p)}{|c_p|^3}.$$

5.3 Geometric integrals

In the following paragraphs, we will pay particular attention to integrals along a regular curve $c : [a, b] \to \mathbb{R}^2$:

$$L_c(g) = \int_a^b g(c(p))|c_p(p)|\, dp. \tag{5.6}$$

We have already seen that when $g \equiv 1$, this is the length of c. Our interest in such integrals is due to their invariance under reparametrization: Let $\tilde{c} : [e, f] \to \mathbb{R}^2, q \mapsto \tilde{c}(q)$ be another parametrization, and let $\phi : [a, b] \to [e, f]$ be a diffeomorphic change of parametrization such that $\tilde{c}(\phi(p)) = c(p)$, with inverse diffeomorphism ψ. By the chain rule, we get $c_p(p) = \phi_p(p)\tilde{c}_q(\phi(p))$ and by theorem 15 we get

$$\begin{aligned} L_{\tilde{c}}(g) &= \int_e^f g(\tilde{c}(q))|\tilde{c}_q(q)|\, dq \\ &= \int_a^b g\left(\tilde{c}(\phi(p))\right)|\tilde{c}_q(\phi(p))|\phi_p(p)\, dp \\ &= \int_a^b g(c(p))|c_p(p)|\, dp \\ &= L_c(g). \end{aligned}$$

When s is an a-l parametrization of c, we then have

$$L_c(g) = \int_0^L g(c(s))\, ds$$

because $|c_s| = 1$. We then write

$$L_c(g) = \int_c g(s)\, ds = \oint_c g(s)\, ds$$

and we call this the *curvilinear integral* of g along c. It depends only on the geometric curve c and not on its parametrization, except this: if we change the direction along c, then we must change the sign of the integral.

5.4 Implicit representation of curves

In this paragraph I will present a series of standard results concerning the implicit "level set" representation of curves as constant value level sets of functions. The formulas for length and curvature calculations are particularly important.

The idea is to implicitly represent c as the curve of level 0 of a given function: $c = \{\mathbf{x}, \phi(\mathbf{x}) = 0\}$ or, in other words: $\phi(c(p)) = 0$ for all $p \in [a, b]$. By differentiating this identity, we obtain

$$\nabla\phi \cdot c' = 0 \Leftrightarrow \nabla\phi \perp c'.$$

This means that the outward normal to c is

$$\vec{\nu} = \pm\frac{\nabla\phi}{|\nabla\phi|}$$

because $|\frac{\nabla\phi}{|\nabla\phi|}| = 1$. The sign depends on the choice of ϕ; one generally chooses the $+$, but if not, simply replace ϕ by $-\phi$.

Now, if c is a family of curves $c = c(p, t)$, where t is the parameter of evolution, we can search for a family of functions $\phi(\mathbf{x}, t)$ such that $c(-, t)$ is the 0 level set of $\phi(-, t)$: $\phi(c(p, t), t) = 0$. From now on, by the gradient of ϕ we will always mean its *spatial* gradient, its partial derivative with respect to time will be denoted ϕ_t, the derivative of c with respect to p will be written c_p and that with respect to t will be denote c_t. (??)

Differentiating the identity with respect to t, we obtain

$$\nabla\phi \cdot c_t + \phi_t = 0.$$

If $c(-, t)$ verifies the PDE $c_t = F\vec{\nu}$ or F is a scalar function, we get by substitution:

$$\begin{aligned} \nabla\phi \cdot F\vec{\nu} + \phi_t &= F\nabla\phi \cdot \frac{\nabla\phi}{|\nabla\phi|} + \phi_t \\ &= F|\nabla\phi| + \phi_t = 0. \end{aligned}$$

if we can extend F at least in a neighborhood of the 0th level set of ϕ. Now we are almost done (????) The problem of Chan-Vese (and many others, even most) is to express the curvature of c as a function of ϕ and its derivatives. This is the following theorem:

Theorem 17 *Let $\phi : \mathbb{R}^2 \to \mathbb{R}$ be a differential map, and let $\mathbf{x}_0$ be a point in $\mathbb{R}^2$ for which $\nabla\phi_{\mathbf{x}_0}$ is nonzero. If $\phi(\mathbf{x}_0) = \lambda \in \mathbb{R}$, there exists a regular curve $c :]a, b[\to \mathbb{R}^2, p \mapsto c(p)$ containing $\mathbf{x}_0$ such that $\phi(c(p)) = \lambda$, $\forall p \in]a, b[$. Moreover, the algebraic curvature of c at p is given by*

$$\kappa(p) = \operatorname{div}\left(\frac{\nabla\phi}{|\nabla\phi|}\right)(c(p)).$$

The first part of the theorem can be proved using the implicit function theorem. The second part, which is also not very complicated, needs some calculations using the Frenet formulas.

We shall assume that c is a-l parametrized with $\mathbf{x}_0 = c(s_0)$ so that $\vec{\tau} = c_s$, $\kappa\vec{\nu} = c_{ss}$. In addition, we assume that $\vec{\nu} = \frac{\nabla\phi}{|\nabla\phi|}$ and hence since $\vec{\tau}$ is obtained by rotating $\vec{\nu}$ by an angle of $-\pi/2$, we have

$$\vec{\tau} = -\frac{1}{|\nabla\phi|}\begin{pmatrix} u_y \\ -u_x \end{pmatrix}.$$

From the identity $\phi(c(s)) = 0$ for $s \in]s_0 - \epsilon, s_0 + \epsilon[$ we obtain $\nabla\phi \cdot \vec{\tau} = 0$ by deriving with respect to s. Deriving once more, we get

$$(\mathrm{Hess}\phi\,\vec{\tau}) \cdot \vec{\tau} + \nabla\phi \cdot \frac{d\vec{\tau}}{ds}$$

where $\mathrm{Hess}\phi$ is the hessian of ϕ.

By the results from the first section, $(\mathrm{Hess}\phi\,\vec{\tau}) \cdot \vec{\tau} = \vec{\tau}^t \mathrm{Hess}\phi\,\vec{\tau}$ and we have

$$\kappa = -\frac{\vec{\tau}^t \mathrm{Hess}\phi\,\vec{\tau}}{\nabla\phi \cdot \vec{\nu}}.$$

We replace $\vec{\tau}$ and $\vec{\nu}$ by their expressions in terms of ϕ, and obtain

$$\kappa = -\frac{\begin{pmatrix}\phi_y & -\phi_x\end{pmatrix}\begin{pmatrix}\phi_{xx} & \phi_{xy}\\ \phi_{xy} & \phi_{yy}\end{pmatrix}\begin{pmatrix}\phi_y\\ -\phi_x\end{pmatrix}}{|\nabla\phi|^3} = \frac{\phi_{xx}\phi_y^2 - 2\phi_{xy}\phi_x\phi_y + \phi_{yy}\phi_x^2}{\left(\sqrt{\phi_x^2 + \phi_y^2}\right)^3}.$$

Now this develops as

$$\mathrm{div}\left(\frac{\nabla\phi}{|\nabla\phi|}\right) = \frac{\partial}{\partial x}\left(\frac{\phi_x}{\sqrt{\phi_x^2 + \phi_y^2}}\right) + \frac{\partial}{\partial y}\left(\frac{\phi_y}{\sqrt{\phi_x^2 + \phi_y^2}}\right)$$

and we can see that we obtain the same expression as above!

5.5 Representation by distance functions

Among functions representing implicitly curves, signed distance functions are very attractive, both because of their mathematical properties and their computational ones. To start with, a definition and a theorem

Definition 18 *A continuous curve $c : [0,1] \to \mathbb{R}^2$ is called a* Jordan curve *if it is closed and has no self intersection: $c(0) = c(1)$, and if $p \neq p' \in [0,1)$, then $c(p) \neq c(p')$.*

Theorem 19 The Jordan curve theorem. *Let c a Jordan curve of the plane $\mathbb{R}^2$. Then the image of c separates the plane into two regions, a bounded one, the interior of c and an unbounded one, the exterior of c.*

Interestingly, this very intuitive theorem is quite difficult to prove. Figure 5.3 illustrates the situation.

Given a bounded domain U and a point x in $\mathbb{R}^2$, one can define the function $d_U(x)$, the distance of x to U as

$$d_U(x) = \inf_{y \in U} d(x, y)$$

where d is the usual Euclidean distance of the plane.

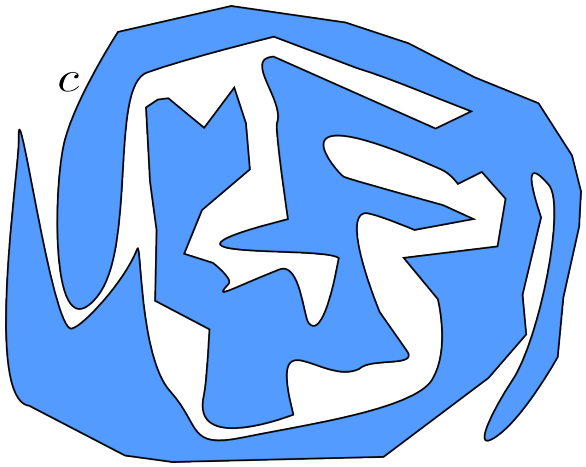

Figure 5.3: A Jordan curve, the interior, bounded region is blue

If now c is a Jordan curve, then call B_c its bounded interior, U_c its unbounded exterior and define

$$d_C(x) = d_{B_c}(x) - d_{U_c}(x).$$

Then it is easy to show that

1. if $x \in U_c$, $d_c(x) > 0$ is the distance from x to C
2. if $x \in B_C$, $d_c(x) < 0$ and $-d_c(x)$ is the distance from x to C
3. if $x \in c$ then $d_c(x) = 0$.

This is the signed distance function associated to c (or more precisely its image in $\mathbb{R}^2$) is the zero-level set of d_C. A very important property of d_C is that for most points x, one has $\|\nabla_x d_c\| = 1$. The points where it fails or where the gradient is not defined is of huge interest, especially in the study of shapes, this is made by the *skeleton* of c and the *cracks*. I won't however say much more about them, see [**?**, **?**].

When c is "sufficiently straight", then one can find a neighborhood along c (a *tubular* neighborhood T) such that is $x \in T$ there is a unique point y of c such that $d_c(x) = d(x, y)$. Within this neighborhood, one has $\|\nabla d_c\| = 1$.

Many formulas from the previous section, including the implicit normal, curvature, etc... are highly simplified in that case.

5.6 A very short introduction to variations of curves

This section introduces basic ideas on how to vary curves, compute differentials of curve functionals and gradients. As usual, mathematical rigor is on vacations.

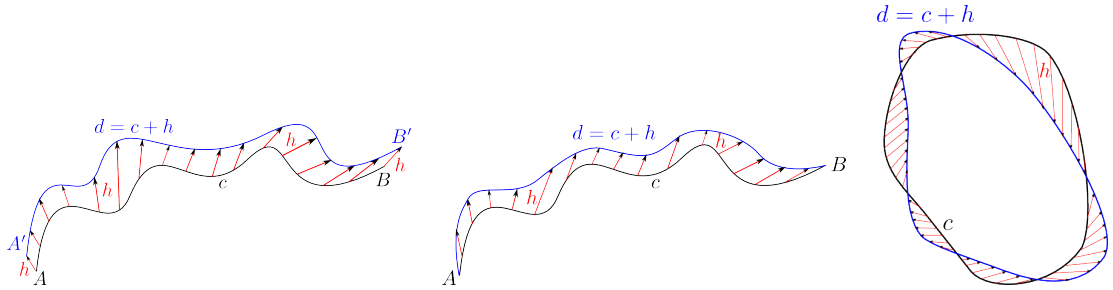

Figure 5.4: Deformations of curves. a) open curve, deformation changing endpoints. b) open curve, deformation fixing endpoints. c) deformation of a closed curve.

We fix a smooth curve $c : p \in [0,1] \mapsto c(p) \in \mathbb{R}^2$. The simplest way curve deformation can be defined is by moving the points of c: a deformation d of c is given as $p \mapsto c(p) + h(p)$ where h is vector field along c in $\mathbb{R}^2$. Figure 5.4 illustrates three different situations: first the case of an open curve whose endpoints move when the curve deforms. The second is the case of an open curve whose deformation does not change the endpoints. The third case is the one of a deforming closed curve. Other situations can occur, they are normally dictated by problems at hand.

To give a deformation of c is thus to give the vector field h. We can use it to develop differential calculus on curves: let $\mathcal{F}$ a functional defined on a space of planar curves; we want to compute its differential and to start with, its directional derivative in a given vector field h direction. We will assume that $\mathcal{F}$ has the following form

$$\mathcal{F}(c) = \int_0^1 F(c, c_p)\, dp. \tag{5.7}$$

Its directional derivative in the direction of h, $d_c\mathcal{F}h$ defined as $\ell'(0)$ where

$$\ell(t) = \mathcal{F}(c + th).$$

Assuming that the Lagrangian $(a, b) \mapsto F(a, b)$ of (5.7) is regular enough, one gets

$$\begin{aligned} \ell'(0) &= \left(\frac{d}{dt} \int_0^1 F(c + t\,h, c_p + t\,h_p)\, dp \right)_{|_{t=0}} \\ &= \int_0^1 \left(F_a(c, c_p) \cdot h + F_b(c, c_p) \cdot h_p \right)\, dp \end{aligned}$$

where F_a and F_b are respectively the derivatives of the Lagrangian F w.r.t its first and second argument ($a, b \in \mathbb{R}^2$). As usual, then, by integration by part,

one gets

$$d_c\mathcal{F}h = \int_0^1 \left(F_a(c, c_p) - \frac{dF_b}{dp}(c, c_p) \right) \cdot h \, dp + F_b(c, c_p) \cdot h|_0^1$$

and in the case of fixed endpoints or close curves, the boundary term above vanishes (for free endpoint problems conditions can be given on F_b in order to make this term vanish if desired). This then suggests a gradient:

$$\nabla_c\mathcal{F} = F_a(c, c_p) - \frac{d}{dp}F_b(c, c_p). \tag{5.8}$$

This is turn means that we have used a very simple inner product:

$$h_1, h_2 \mapsto \langle h_1, h_2 \rangle = \int_0^1 h_1 \cdot h_2 \, dp \tag{5.9}$$

Thus the situation looks completely similar to what has been used in sections 3 and 4. This type of approach has been used in a series of work, the most known being the *snakes* of Kass *et al.* [**?**].

However, with the inner product above, no5.3. geometry the curve c itself appears. Curves are not really used, only their parametrizations are. In order to be geometric, the deformation field should be considered acting not the curve parameter, but on the corresponding point of the curve.

To make it geometric in the sens of paragraph 5.3, it can be replaced by the following. Given two vector fields h_1 and h_2 defined along the image of the regular curve c, set

$$h_1, h_2 \mapsto \langle h_1, h_2 \rangle_c = \int_0^1 h_1(c(p)) \cdot h_2(c(p)) \, |c_p| \, dp. \tag{5.10}$$

This is geometric: if we change the parametrization of c, (5.10) remains unchanged. Remark now that the inner product is indexed by c, the deformations h_1 and h_2 make only sense for the geometric curve c. On the space of regular curves, one does not have one inner product, but a family of products, one per geometric curve. This forms a *Riemannian metric*.

This is not the only geometric structure of the space of regular curves, and this family of inner products, although widely used in many applications, some will be presented in the next section, has also severe shortcomings that will maybe be discussed.

At the same time, the type of functional we looked at, may fail to be geometric in the sense of paragraph 5.3. We look at specific ones, guaranteed to be geometric:

$$F(a, b) = G(a)|b|, \quad \text{so that} \quad \mathcal{F}(c) = \int_0^1 G(c)|c_p| \, dp \tag{5.11}$$

So, with such a functional and the inner product above, we will compute new gradients. They will illustrate a rather interesting fact. As usual, we first

compute the differential $d_c\mathcal{F}h$, details are left to the reader:

$$\begin{aligned} d_c\mathcal{F}h &= \left(\int_0^1 \frac{d}{dt}\left(G(c+t\,h)|c_p+t\,h_p|\,dp\right)\right)_{|t=0} \\ &= \underbrace{\int_0^1 |c_p|\nabla G(c)\cdot h\,dp}_{I_h} + \underbrace{\int_0^1 \left(G(c)\frac{c_p}{|c_p|}\right)\cdot h_p\,dp}_{I_{h_p}} \end{aligned}$$

Before going further, remember that $c_p/|c_p|$ is the unit tangent $\vec{\tau}$ at $c(p)$ (5.1). In order to write it as an inner product of the form (5.10), one needs first to get rid of the derivative in h_p in the term $I_{h_p}$by integration by part and assuming that the boundary term vanishes, I_{h_p} becomes, developing and using Frenet formula (5.4),

$$\begin{aligned} I_{h_p} &= -\int \frac{d}{dp}\left(G(c)\vec{\tau}\right)\cdot h\,dp \\ &= -\int_0^1 \left(\nabla G(c)\cdot c_p\right)\vec{\tau} + |c_p|G(c)\kappa\,\vec{\nu}\,)\cdot h\,dp \\ &= -\int_0^1 |c_p|\left(\left(\nabla G(c)\cdot\vec{\tau}\right)\vec{\tau} + G(c)\kappa\vec{\nu}\right)\cdot h\,dp \end{aligned}$$

using the fact that $c_p = |c_p|\vec{\tau}$. We thus obtain

$$d_c\mathcal{F}h = \int_0^1 |c_p|\left(\nabla G(c) - \left(\nabla G(c)\cdot\vec{\tau}\right)\vec{\tau} - \kappa G(c)\vec{\nu}\right)\cdot h\,dp$$

But the pair $(\vec{\tau},\vec{\nu})$ is an *orthonormal* moving frame along c, so that

$$\nabla G(c) = \left(\nabla G\cdot\vec{\tau}\right)\vec{\tau} + \left(\nabla G\cdot\vec{\nu}\right)\vec{\nu}$$

from which one gets

$$d_c\mathcal{F}h = \int_0^1 |c_p|\left(\left(\nabla G(c)\cdot\vec{\nu}\right) - \kappa G(c)\right)\vec{\nu}\cdot h\,dp$$

And the corresponding gradient, for inner product (5.10) is

$$\nabla_c\mathcal{F} = \left(\nabla G(c)\cdot\vec{\nu} - \kappa G(c)\right)\vec{\nu} \tag{5.12}$$

is a vector field **normal** to c. In fact when evolving curves geometrically, *only the normal direction "counts"*, the tangential direction only changes the parametrization. There is in fact the following result of Epstein and Gage [**?**, **?**].

Theorem 20 *Let $c:[0,1]\times[0,1]\to\mathbb{R}^2, (p,t)\mapsto c(p,t)$ a time-dependent curve family, i.e, for t fixed, $c^t: p\mapsto c^t(p)=c(p,t)$ is a regular curve in $\mathbb{R}^2$, $\mathbb{C}^2$ in space and C^1 in time, and assume that it satisfies the differential equation*

$$\partial_t c(p,t) = h_1(p,t)\vec{\tau}(c(p,t)) + h_2(p,t)\vec{\nu}(c(p,t))$$

where $(\vec{\tau}(-), \vec{\nu}(-)$ *is the orthonormal moving frame at time* t*. Then there exists a time-dependent reparametrization of* c*, denoted* ϕ *such that* $\tilde{c}(p,t) = c(\phi(p,t),t)$ *satisfies the differential equation*

$$\partial_t \tilde{c} = \tilde{h}_2 \vec{\nu}$$

with $\tilde{h_2}(p,t) = h_2(\phi(p,t),t)$.

Figure 5.5 illustrates this fact. A deformation of a curve from a purely geometric point of view, is a *normal vector field to that curve*. If one is interested in purely geometric aspects, then one can restrict to such vector fields.

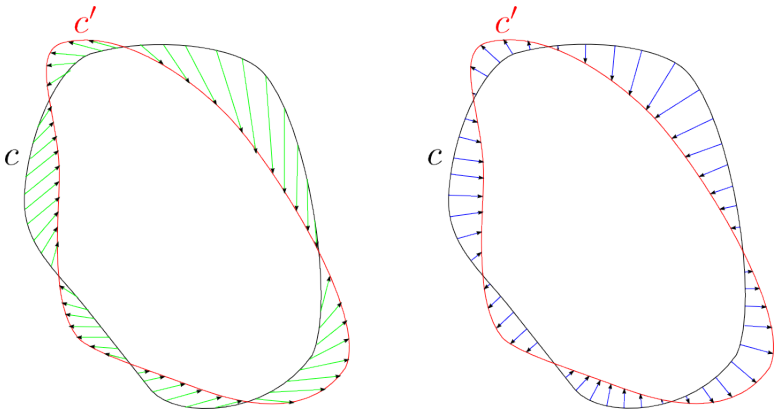

Figure 5.5: Two deformations; that is, two vector fields on a curve c inducing a deformation to the curve c'. The first deformation (in green) has tangential components. The second (in blue) is a purely normal vector field to c.

5.7 An extremely short / ultralight introduction to shape derivatives

A rigorous description of shape calculus is far beyond the purpose of these note, the reader is referred to [**?**]. We discuss it in an informal manner. The regions we consider will be "reasonable ones" in the following sense.

In order to describe how a domain changes, one starts with a time-dependent vector field $v(x,t)$ representing the velocity of a "particle" at position x at time t. At time $t+dt$, this point will move at first order to new position $x+v(x,t)dt$. Now, assume that we are given a domain $\Omega(t)$ and a vector field $v(x,t)$ defined on a domain sufficiently larger than $\Omega(t)$, call it D, the "hold-all" domain. We assume also that at the boundary of D, $v(x,t) = 0$.

One deforms $\Omega(t)$ at time $t+dt$ by moving each of its points as done above and one get the moved/deformed region at time $t+dt$, $\Omega(t+dt)$. What we need is a reasonable, operational, definition of $d\Omega(t)/dt$, the *shape derivative*,

skipping the rather complicated "details": if $\partial\Omega(t)$ is the boundary of $\Omega(t)$, i.e. the exterior delimiting curve of $\Omega(t)$, the instantaneous deformation velocity of $\Omega(t)$ is given by a vector field normal to $\partial\Omega(t)$. Each point x of $\partial\Omega(t)$ moves with instantaneous velocity $v(x,t)$, but one only keeps the normal direction, as the tangent one does not modify $\partial\Omega(t)$ *geometrically*, as I explained in a previous paragraph, see (again) Figure 5.5. Let $\vec{\nu}$ be the exterior normal vector field to $\partial\Omega(t)$. One has the following fundamental result:

Theorem 21 *Let $f(\mathbf{x},t)$ be a differentiable function in $x \in D \subset \mathbb{R}^n$ and $t \in [a,b] \subset \mathbb{R}$ (the "time") and let $g(t)$ be the function defined by*

$$g(t) = \int_{\Omega(t)} f(\mathbf{x},t)\, d\mathbf{x}.$$

Assume that the instantaneous deformation of $\partial\Omega(t)$ is given by the vector field $v(x,t)$. Then

$$g'(t) = \int_{\Omega(t)} \frac{\partial f}{\partial t}(\mathbf{x},t)\, dx + \int_{\partial\Omega(t)} f(\mathbf{x},t)\, v\cdot\vec{\nu}\, ds.$$

For the case where $\partial\Omega(t)$ is a curve, the last integral means is the *geometric* integral along a good parametrization of $\partial\Omega(t)$: if $c : [a,b] \to \mathbb{R}^2, p \mapsto c(p)$ is one of them, oriented such that the orthonormal basis $(\frac{c'}{|c'|}, \vec{\nu})$ is direct, then

$$\int_{\partial\Omega(t)} f(\mathbf{x},t) V\cdot\vec{\nu}\, ds := \int_a^b |\dot{c}p)]| f(c(p),t) v(p)\cdot\vec{\nu}(p)\, dp.$$

This formula is in fact also valid in $\mathbb{R}^n$ with $\partial\Omega(t)$ an hypersurface of $\mathbb{R}^n$, but instead of dealing with the curve velocity $|c'|$, a more complicated expression, for the hypersurface volume element, is necessary.

An alternate way to look at the previous result is as follows. Let χ_t be the characteristic function of $\Omega(t)$,

$$\chi_t(x) = \begin{cases} 1 & \text{if } x \in \Omega(t) \\ 0 & \text{if } x \in D\backslash\Omega(t). \end{cases}$$

Then the above function g can be written as

$$g(t) = \int_D f(x,t)\chi_t(x)\, dx$$

and the domain of integration D is now fixed. In order to compute $g'(t)$ one differentiates the integrand to obtain:

$$g'(t) = \int_D \frac{\partial f}{\partial t}(x,t)\chi_t(x)\, dx + \int_D f(x,t)\frac{\partial \chi_t}{\partial t}(x)\, dx.$$

As χ_t is constant inside Ω_t (value 1) and outside it (value 0), its variation $\frac{\partial\chi_t}{\partial t}$ is concentrated on $\partial\Omega(t)$ and is given by the normal component of $v(x,t)$. This requires of course serious theoretical justifications, they can be found in [**?**].

Chapter 6

Optimization of curves and regions

Thsi chapter collects some classical curve optimization computations, starting with length and the Euclidean shortening flow and then many problems related to image segmentation, starting from Kass, Wiktin and Terzopoulos' Snakes and gradually moving to more complicated formulations.

6.1 Snakes

This is where a large part of the story started in Image Analysis. The idea is to find a curve that delineates an object in an image by following its edges. Because of noise and other problems, the edges might be incomplete and / or noisy. Thus Kass, Witkin and Terzopoloulos proposed to find the segmentation bu minimizing the energy

$$E(c;I) = \int_0^1 \left(\alpha|c_p|^2 + \beta|c_{pp}|^2\right)\, dp + \gamma \int_0^1 g_I(c)\, dp \tag{6.1}$$

where $alpha > 0$, $\beta <= 0$ and $\gamma > 0$, while g should be small at edges of I. This could be

$$g_I(x,y) = -\nabla I_\sigma(x,y), \quad g_I(x,y) = \frac{1}{1+\nabla I_\sigma(x,y)}$$

where I_σ is a (Gaussian) smoothed version of I in order to get rid of some of the noise.

The energy $E(c,I)$ *is not geometric* so we use inner product (5.8) in that

case. Let's do the computation step by step.

$$\begin{aligned} d_cEh &= \frac{d}{dt}\left(\int_0^1 \left(\alpha|c_p + t\,h_p|^2 + \beta|c_{pp} + t\,h_{pp}|^2\right)\,dp + \gamma\int_0^1 g_I(c + t\,h)\,dp\right)_{t=0} \\ &= \int_0^1 \left(\alpha c_p \cdot h_p + \beta c_{pp}\cdot h_{pp}\right)\,dp + \gamma\int_0^1 \nabla_c g_I(c)\cdot h\,dp \\ &= \int_0^1 \left(-\alpha c_{pp}\cdot h + \beta c_{pppp}\cdot h\right)\,dp + \gamma\int_0^1 \nabla_c g_I(c)\cdot h\,dp \\ &\quad \text{(one integration by parts to elimate } h_p,\ \text{two for } h_{pp}) \\ &= \int_0^1 \left(-\alpha c_{pp} + \beta c_{pppp} + \gamma g_I(c)\right)\,h\,dp \end{aligned}$$

and the sought gradient is thus

$$\nabla_c E = -\alpha c_{pp} + \beta c_{pppp} + \gamma g_I(c).$$

6.2 Curves of minimal length

In the second example of Calculus of variation with curves, we wish to characterize the curve of minimal length joining the points A and B in the euclidean plane. It is well known that it is simply the straight line segment joining A to B. But what can we say from the point of view of the calculus of variations?

For a curve $c : [a, b] \to \mathbb{R}^2$, the length functional is given by

$$\ell(c) = \int_a^b |c_p|\,dp$$

which corresponds precisely to the case $F(a, b) = |b|$ and $G = 1$ in the gradient computations performed in §5.6. We do the computations with the two gradients we have discussed. Using inner product (5.9) and corresponding gradient (5.8), one find the following gradient

$$\nabla_c^1\ell = -\frac{d\nabla_b F(c_p)}{dp}p = -\frac{d}{dp}\frac{c_p}{|c_p|} = -\frac{d}{dp}\vec{\tau} = -|c_p|\kappa\vec{\nu}$$

the second equality because $\nabla_b|b| = \frac{b}{|b|}$ and the last by Frenet's formula. Using the geometric product (5.9), we find instead

$$\nabla_c^2\ell = -\kappa\vec{\nu}.$$

Of course, in both case, the optimality condition reduces to the vanishing of the curvature or the curve: $\kappa = 0$. Change to al-parametrization $s \in [0, L] \mapsto c(s)$ where L is the length of c. This can be done w.l.o.g. Then $\kappa\vec{\nu} = c_{ss}$, the second derivative of c. It follows immediatly that c must have the form

$$c(s) = A + (L - s)B,$$

thus a straight-line segment.

6.3 Geodesic active contours

We are given an image $u :]0,1[\times]0,1[\to \mathbb{R}$ containing an object that we would like to segment via its contours, where the segments are given as the interior of the curve given by c.

How can we find or evolve this curve so that it tours the object? I will suppose, at least in this version of these notes, that the reader knows the snake method of Kass, Witkin and Terzopoulos (KWT), because I have good reason to believe that a reader of this document is familiar with these.

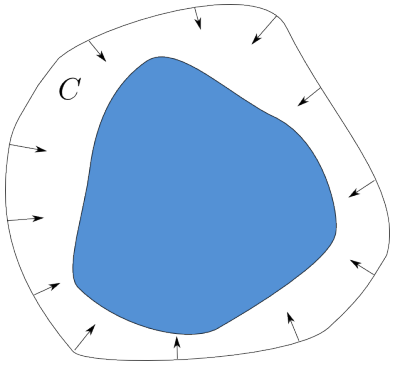

Figure 6.1: Active contours

The idea of geodesic active contours is to remedy the problems with the parametrization by KWT through a geometric formulation: If, in a change of parametrization of the contour, the energy does not change, then the two curves parametrizing the same image in $\mathbb{R}^2$ should have the same energy.

For a curve $c : [a,b] \to \mathbb{R}^2$, where we assume that c is *closed smooth*, that is $c(a) = c(b)$ and $\dot{c}(a) = \dot{c}(b)$, this energy is given by

$$J(c) = \int_a^b g(|\nabla u(c(p))|)|\dot{c}p)|\, dp.$$

and the invariance of energy in this reparametrization comes from the term $|\dot{c}p)|$ in the integrand: let $\bar{c} : [e,f] \to \mathbb{R}^2$ be another parametrization of c, that is, there exists a function $\phi : [e,f] \to [a,b]$, $\phi' > 0$, $\bar{c}(r) = c(\phi(r))$, then we have, since $\bar{c}'(r) = \dot{c}\phi(r))\phi'(r)$,

$$\begin{aligned} J(\bar{c}) &= \int_e^f g(|\nabla u(\bar{c}(r))|)|\bar{c}'(r)|\, dr \\ &= \int_e^f g(|\nabla u(c(\phi(r))|)|\dot{c}\phi(r))|\phi'(r)\, dr \\ &= \int_a^b g(|\nabla u(c(p))|)|\dot{c}p)|\, dp \\ &= J(c) \end{aligned}$$

by the change of variables theorem.

Again and again, we will proceed quickly in calculating the gradient using the method of directional derivatives in a given direction h. The Frenet relations are crucial for this method, and we will generalize our energy (a little) to conceal the dependence on the gradient of the image u in $g : (x, y) \to \mathbb{R}$:

$$J(c) = \int_a^b g(c(p))|\dot{c}p)|\, dp.$$

$$\begin{aligned} dJ_c\, h &= \frac{d}{dt}\int_a^b g(c(p) + th(p))|\dot{c}p) + th'(p)|\, dp\Big|_{t=0} \\ &= \int_a^b \left(\nabla g(c)\cdot h|c'| + g(c)\frac{c'\cdot h'}{|c'|}\right) dp \\ &= \int_a^b \left(\nabla g(c)|c'| - \frac{d}{dp}\frac{c'}{|c'|}\right)\cdot h\, dp \end{aligned}$$

Using the Frenet relations (5.4) given in the previous section,

$$\frac{d}{dp}\frac{c'}{|c'|} = \frac{d\vec{\tau}}{dp} = |c'|\kappa\vec{\nu}$$

we get

$$dJ_c\, h = \int_a^b |c'|\,(\nabla g(c) - \kappa\vec{\nu})\cdot h\, dp = \langle |c'|\,(\nabla g(c) - \kappa\vec{\nu}), h\rangle. \tag{6.2}$$

The factor $|c'|$ appears in the above integral, which means that $dJ_c\, h$ is invariant under reparametrization. We must look for a gradient of the form $|c'|\,(\nabla g(c) - \kappa\vec{\nu})$, yet this is not the one generally presented, which is

$$(\nabla g(c)\cdot\vec{\nu})\vec{\nu} - \kappa\vec{\nu}.$$

In fact, in terms of evolution of geometric objects, both gradients are very close. On one hand, *when applying a tangential deformation to a curve, the geometric curve in $\mathbb{R}^2$ remains unchanged, although the parametrization does change. Purely normal deformations deform the geometric curves.*

With this fact, one has $\nabla g = (\nabla g\cdot\vec{\tau})\vec{\tau} + (\nabla g\cdot\vec{\nu})\vec{\nu}$ and $g(c)\cdot\vec{\nu})\vec{\nu}$ is precisely the normal component of ∇g. From a purely geometric viewpoint these two gradients provide the same solution.

By choosing an arc-length parameterization, $|c'| = 1$, there is no much difference between these gradients. In fact the difference between these two expressions comes from the inner product one works with. The function space inner product does not say much on geometry!

Thus instead, given a closed and smooth curve $c : [a, b] \to \mathbb{R}^2$, then, to two normal vector fields along c, $v_1, v_2 : [a, b] \to \mathbb{R}^2$ (i.e. $\forall p \in [a, b], v_i(P)$ is parallel

to $\vec{\nu}(p)$, $i = 1, 2$) one associates the inner product

$$\langle v_1, v_2 \rangle_c = \int_a^b (v_1 \cdot v_2) \, |c'| \, dp. \tag{6.3}$$

A main difference between this on and the previous one is that *this product is defined w.r.t the curve* c, we are in fact in a situation somewhat analoguous to the case of optimization under smooth equality constraints discussed in section 2.7: at each point $\mathbf{x}$ of the constraint space D one has a tangent space, $T_{\mathbf{x}}D$. Here the tangent space to a curve c is the set of smooth normal vector fields to c (yes!)

With this scala product (6.3), what is teh corresponding gradient? We got

$$dJ_c \, h = \int_a^b |c'| \, (\nabla g(c) - \kappa \vec{\nu}) \cdot h \, dp$$

We shall restrict ourselves to the vector fields h that are normal w.r.t. c and by a straightforward calculation, one gets

$$(\nabla g(c) - \kappa \vec{\nu}) \cdot h = ((\nabla g(c) - \kappa \vec{\nu}) \cdot \vec{\nu}) \, \vec{\nu} \cdot h$$

et

$$dJ_c \, h = \int_a^b |c'| \, ((\nabla g(c) - \kappa \vec{\nu}) \cdot \vec{\nu}) \, \vec{\nu} \cdot h \, dp = \int_a^b |c'| \, ((\nabla g(c) \cdot \vec{\nu}) \vec{\nu} - \kappa \vec{\nu}) \cdot h \, dp$$

from which it follows that, with the inner product (6.3),

$$\nabla J_c = (\nabla g(c) \cdot \vec{\nu}) \vec{\nu} - \kappa \vec{\nu}$$

which is the expected gradient for the geodesic active contours.

When using a steepest gradient descent, in order to solve the minimization, the steepest descent direction h is given via the Cauchy-Schwarz Theorem (Theorem 9), it says that this must be a positive multiple of the opposite of the gradient:

$$h = -\lambda \left((\nabla g(c) \cdot \vec{\nu}) \vec{\nu} - \kappa \vec{\nu} \right).$$

Choosing $\lambda = 1$, one gets

$$c_t = -(\nabla g(c) \cdot \vec{\nu}) \vec{\nu} + \kappa \vec{\nu}$$

another choice for λ will only result in constant *global* multiplication of this evolution speed, and in the implementation, will be "absorbed" in our choice of dt.

6.4 Inner products and distances

6.5 Chan and Vese's Active Contours

In the segmentation problem of Chan and Vese [**?**], variables regions of a given domain, these regions are used as integration domains. We need thus to understand how to manipulate them in order to optimize them w.r.t a given criterion.

Before describing the exact criterion used by Chan and Vese, we describe how the regions can be manipulated.

6.5.1 Chan and Vese cost function

Let $u : D \to \mathbb{R}$ a grayscale image. D is the hold-all domain discussed above, usually a square in $\mathbb{R}^2$. We assume that this image a made of an "inner" region Ω_i and an outer region Ω_e which differ enough by their contents, typically: the mean gray values of each region should sufficiently far away and the gray values well distributed around the means.

The goal of the segmentation is to recover these regions, or, equivalently, to recover their common boundary c. Chan and Vese segmentation has the following variational formulation: find c minimizing the following objective function

$$J(c, a_i, a_e) = \lambda_1 \int_{\Omega_i} (u(x,y) - a_i)^2 \, dxdy + \lambda_2 \int_{\Omega_e} (u(x,y) - a_e)^2 \, dxdy + \mu\ell(c)$$

where c has Ω_i as interior, Ω_e, as exterior a_i et a_e are two real to be determined, while λ_1, λ_2 and μ are three positive constant and $\ell(c)$ is the length of c (defined in one of the previous sections).

So, what we are looking for is a curve of minimal length, separating as best as possible the two regions in the sens that the quantities a_i et a_e associated to each region are as far away as possible from each other (there is, in the original formulation of Chan and Vese [**?**] an area term, which in practice is seldom used, so I have dropped it).

We will look at two formulations of the Chan-Vese problem, the first uses optimization of regions/curves, and the second one optimizes a levelset function. We start with the region approach.

Based on the inner product (6.3), we redo again the calculation of $d_c\ell V$ for a vector field v normal to c, where c is given with a parametrization $c : [a, b] \to \mathbb{R}^2$

$$\begin{aligned}
d_c\ell V &= \int_a^b \frac{c' \cdot v'}{|c'|} \, dp \\
&= -\int_a^b \frac{d}{dp}\left(\frac{c'}{|c'|}\right) \cdot V \, dp \\
&= -\int_a^b |c'| \left(\frac{1}{|c'|}\frac{d}{dp}\left(\frac{c'}{|c'|}\right) \cdot V\right) \, dp \\
&= -\int_a^b |c'|(\kappa\vec{\nu} \cdot V) dp = \langle -\kappa\vec{\nu}, V\rangle_c
\end{aligned}$$

by integration by part and Frenet formula (5.4), and the gradient is thus

$$\nabla\ell_c = -\kappa\vec{\nu}.$$

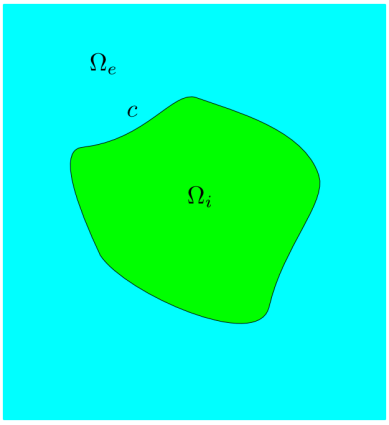

Figure 6.2: Chan-Vese problem.

When c is fixed and thus Ω_i and Ω_e as well, minimization on a_e and a_i is quite simple. For instance, for a_i one gets

$$\frac{\partial J}{\partial a_i} = 2\lambda_i \int_{\Omega_i} (u(x,y) - a_i)\, dxdy$$

and setting $\frac{\partial J}{\partial a_i} = 0$ gives

$$\int_{\Omega_i} u(x,y)\, dxdy = \int_{\Omega_i} a_i\, dxdy = a_i \int_{\Omega_i} dxdy \Rightarrow a_i = \frac{\int_{\Omega_i} u(x,y)\, dxdy}{\int_{\Omega_i} dxdy}$$

i.e. , a_i is the mean value of u in Ω_i. The same result holds of course for a_e. Before we proceed to the gradient computation, let us remark that the outward normal to Ω_i along c is the inward normal to Ω_e along c: Ω_e i the "exterior" of Ω_i and Ω_i is the "exterior" of Ω_e!

We will use Theorem 21 as follows. Suppose that for a given t, $\Omega_i(t)$ and $\Omega_e(t)$ deform via the vector field V, normal to c. Since the values of $a_i(t)$ and $a_e(t)$ are known exactly for t as the mean values of u within the corresponding regions, we fix them and forget their t-dependency [1]. Set $J_i(t) = \lambda_i \int_{\Omega_i(t)} (u - a_i)^2\, dxdy$ and $J_e(t) = \lambda_e \int_{\Omega_i(t)} (u - a_e)^2\, dxdy$. Then

$$J_i'(t) = \lambda_i \int_{\Omega_i(t)} \frac{\partial (u(x,y) - a_i)^2}{\partial t} dxdt + \lambda_i \int_c (u(x,y) - a_i)^2 V \cdot \vec{\nu}\, ds = \lambda_i \int_c (u(x,y) - a_i)^2 V \cdot \vec{\nu}\, ds$$

since the first integral does not depend on t. In the same way, using the remark on the sign of the outward normals,

$$J_e'(t) = -\lambda_i \int_c (u(x,y) - a_e)^2 V \cdot \vec{\nu}\, ds.$$

[1] This may seema bit strange, but one uses implicitly the Lagrange multipliers method: one add extra parameters for a_i et a_e and we impose the mean value constraint **[François dit: need to come with a more decent explanation...]**

So set $\bar{J}_i(c) = \lambda_i \int_{\Omega_i} (u - a_i)^2 \, dxdy$ and $\bar{J}_e(c) = \lambda_i \int_{\Omega_e} (u - a_i)^2 \, dxdy$, and we have in fact obtained

$$\begin{aligned} d\bar{J}_{ic} V &= \lambda_i \int_c (u(x,y) - a_i)^2 V \cdot \vec{\nu} \, ds = \langle \lambda_i (u - a_i)^2 \vec{\nu}, V \rangle_c \\ d\bar{J}_{ec} V &= -\lambda_e \int_c (u(x,y) - a_e)^2 V \cdot \vec{\nu} \, ds = \langle -\lambda_e (u - a_e)^2 \vec{\nu}, V \rangle_c \end{aligned}$$

Reuse the calculation made for the gradient of the length functional $d\ell_c V = \langle -\kappa \vec{\nu}, V \rangle$ to get

$$dJ_c V = \langle \left(\lambda_i (u - a_i)^2 - \lambda_e (u - a_e)^2 \right) \vec{\nu} - \mu \kappa \vec{\nu}, V \rangle_c.$$

If one uses a gradient descent approach, the resulting PDE on c is

$$c_t = -\left(\lambda_i (u - a_i)^2 - \lambda_e (u - a_e)^2 \right) \vec{\nu} + \mu \kappa \vec{\nu}.$$

[François dit: When I have time, I will discuss how to make a Level Set implementation of it and discuss that it will differ from the Chan-Vese direct Level Set approach. Maybe I could say a work about some of the stuffs of Charpiat.]