Variational Methods for Image Processing

MSIAM 2^{nd} **year and** 3^{rd} **year**

S. Meignen

2022/2023

Contents

Chapter 1

Differentiability on Normed Vector Spaces

In all this chapter, we will consider $\mathbb{R}$-vector space.

1.1 Fréchet Differentiability

Definition 1 *Let E and F be two normed vector spaces, and U be a non empty open set of E and $x \in U$. Let $f : U \longrightarrow F$. f is said to be Fréchet-differentiable at x if there exists $L \in \mathcal{L}(E, F)$ such that:*

$$f(x+h) = f(x) + L(h) + \underbrace{\|h\|\varepsilon(h)}_{o(\|h\|)}$$

where $\varepsilon : V \longrightarrow F$ is an application defined on an open neighborhood V of 0_E such that $\forall\ h \in V$, $x + h \in U$, and $\lim\limits_{\|h\|\to 0} \varepsilon(h) = 0$.
L is called differential of f at x and is denoted either by $f'(x)$, df_x or $Df(x)$. In what follows, we will use the term differentiable for Fréchet-differentiable.

Some properties

1) L is unique.

2) f is differentiable on U if f is differentiable at every point in U. In this case, one has $f' : U \longrightarrow \mathcal{L}(E, F)$. f is said to be C^1 if f is differentiable on U and f' is continuous (f is C^2 if f' is C^1, ...).

3) The set of the applications differentiable at $x \in U$ (resp. on U) is a vector space.

4) If f is differentiable at $x \in U$ then f is continuous at x.

proof

- Point 1 : If L_1 and L_2 equal $Df(x)$ then: Let $h \in E \setminus 0$ and $t \in]0;\varepsilon'[$ (ε' sufficiently small so that $x + t\,h \in V$).

$$f(x+t\,h) = f(x) + L_1(t\,h) + o(\|t\,h\|) = f(x) + L_2(t\,h) + o(\|t\,h\|)$$

Thus $\|L_1(h) - L_2(h)\| = \left\| \dfrac{L_1(t\,h) - L_2(t\,h)}{t} \right\| = \dfrac{o(t\,\|h\|)}{t} \xrightarrow[t\to 0]{} 0$

- Point 4 : $\|f(x+h) - f(x)\| \leq \|L(h)\| + \|h\|\,\|\varepsilon(h)\| \xrightarrow[\|h\|\to 0]{} 0$

Remark 1 *In the case, $E = F = \mathbb{R}$, we know that if f is differentiable at x then $\lim\limits_{h\to 0, h\neq 0} \dfrac{f(x+h)-f(x)}{h} = f'(x)$, which can be rewritten as $f(x+h) = f(x) + f'(x)(h) + o(|h|)$, and which corresponds to Definition 1.*

Examples 1

1) Let $f : U \longrightarrow F$ and $c \in F$ such that $\forall\, x \in U$, $f(x) = c$. Then $\forall\, x \in U$, $f'(x) = 0$. So $f' : U \longrightarrow \mathcal{L}(E,F)$ is null. f is C^∞. $f'' \in \mathcal{L}(E, \mathcal{L}(E,F))$.

2) $U = E$ and $f \in \mathcal{L}(E,F)$. Then $\forall\, x \in E$, $f'(x) = f$ (because $f(x+h) - f(x) = f(h) = f'(x)(h)$). Thus f' is constant, f is C^∞ and higher order differentials are null.

3) Let E_1, E_2 and F be normed vector spaces, and $B : E_1 \times E_2 \to F$ a continuous bilinear application. Then B is differentiable and $B'(x,y)(h,k) = B(x,k) + B(h,y)$.

proof of 3). Let us first show that

$$\exists\; M > 0 \;s.t.\; \|B(x,y)\| \leq M\; \|x\|\; \|y\|$$

As B is continuous at 0:

$$\exists\; \delta > 0 \;s.t.\; \|(x,y)\|_\infty \leq \delta \Longrightarrow \|B(x,y)\| \leq 1$$

with $\|x\|_\infty = \max(\|x_1\|, \|x_2\|)$, and for this norm $E_1 \times E_2$ is a normed vector space.
Assume x and y are non zero. Then one has: $\underbrace{\left\| B\left(\dfrac{\delta\,x}{\|x\|}, \dfrac{\delta\,y}{\|y\|}\right) \right\|}_{= \frac{\delta^2}{\|x\|\,\|y\|}\,\|B(x,y)\|} \leq 1$. So $\|B(x,y)\| \leq \dfrac{1}{\delta^2}\; \|x\|\; \|y\|$ and we take $M = \dfrac{1}{\delta^2}$. Then, one can write:

$$B(x+h, y+k) = B(x,y) + \underbrace{B(x,k) + B(h,y)}_{B'(x,y)(h,k)} + \underbrace{B(h,k)}_{o(\|(h,k)\|_\infty)}\,.$$

Indeed, $\|B(h,k)\| \leq M\; \|h\|\; \|k\| \leq M\; \|(h,k)\|_\infty^2$, so we have $B(h,k) = o(\|(h,k)\|_\infty)$.

Furthermore, $B'(x,y) : (h,k) \longmapsto B(x,k) + B(h,y)$ is clearly linear:

$$\begin{aligned} B'(x,y)\big(\underbrace{\lambda(h,k)+(u,v)}_{(\lambda h+u,\lambda k+v)}\big) &= B(x,\lambda k+v)+B(\lambda h+u,y) = \lambda B(x,k)+B(x,v)+\lambda B(h,y)+B(u,y) \\ &= \lambda B'(x,y)(h,k)+B'(x,y)(u,v) \end{aligned}$$

$B'(x,y)$ is continuous, indeed:

$$\big\|B'(x,y)(h,k)\big\| \leq M\ (\|x\|\ \|k\| + \|y\|\ \|h\|) \leq M\ (\|x\| + \|y\|)\ \|(h,k)\|_\infty$$

Application Let E, F, and G be three different normed vector spaces.

$$\begin{aligned} B : \mathcal{L}(E,F)\times\mathcal{L}(F,G) &\longrightarrow \mathcal{L}(E,G) \\ (u,v) &\longmapsto B(u,v) = v\circ u \end{aligned}$$

B is bilinear and continuous ($\|B(u,v)\| = \|v\circ u\| \leq \|v\|\,.\,\|u\|$). Thus B is differentiable and $B'(u,v)(h,k) = B(u,k)+B(h,v) = k\circ u + v\circ h$.

1.2 Differential of a composition of functions

Theorem 1 **(chain rule)**

Let E, F, and G three normed vector spaces. Let $U\subset E$ an open set, $x\in U$, $f : U \longrightarrow F$, $V\subset F$ an open set, $b = f(x)\in V$ and $g : V \longrightarrow G$. If f is differentiable at x and if g is differentiable at $b = f(x)$, then $g\circ f$ (defined on a neighborhood of x and continuous at x) is differentiable at x, and one has:

$$(g\circ f)'(x) = g'(f(x)).f'(x).$$

If f and g are C^p ($p\geq 1$) then $g\circ f$ also belongs to C^p.

proof Let us put $y = f(x)$ then

$$g(y+k) = g(y) + g'(y).k + o(\|k\|).$$

Let us define $o(\|k\|) = \|k\|r_1(k)$ with $\lim\limits_{\|k\|\to 0} r_1(k) = 0$, then set $k = f(x+h)-f(x) = f'(x)h + \|h\|r_2(h)$ with $\lim\limits_{\|h\|\to 0} r_2(h) = 0$. Then one can write:

$$\begin{aligned} g\circ f(x+h) &= g(f(x+h)) = g(y+k) = g(y)+g'(y).(f'(x)h+\|h\|r_2(h)) + o(\|h\|) \\ &= g\circ f(x) + g'(f(x)).f'(x)h + \|h\|g'(y).r_2(h) + o(\|h\|) \end{aligned}$$

But $\|g'(y).r_2(h)\| \leq \|g'(y)\|\|r_2(h)\|$. As $\|g'(y)\|$ is bounded and $\lim\limits_{\|h\|\to 0} r_2(h) = 0$, one obtains $\lim\limits_{\|h\|\to 0} g'(y)r_2(h) = 0$. Finally, as $g'(f(x)).f'(x)\in\mathcal{L}(E,F)$, we get the result.

Application : functions with values in a product of normed vector spaces

Proposition 1

Let $U \subset E$ an open set, and $f : U \longrightarrow F = F_1 \times \cdots \times F_k$ with F_i a normed vector space, $\|y\| = \max_{1 \le i \le k} \|y_i\|$ for $y = (y_1, \cdots, y_k) \in F$.

$$\begin{array}{rcl} u_i : F_i & \longrightarrow & F \\ y_i & \longmapsto & (0, \ldots, 0, y_i, 0, \ldots, 0) \end{array}$$

$$\begin{array}{rcl} p_i : F & \longrightarrow & F_i \\ y & \longmapsto & y_i \end{array}$$

f is differentiable at $x \in U$ if and only if for $i = 1, \ldots, k$, $f_i = p_i \circ f$ is differentiable at x. Then, we get:

$$f'(x) = \sum_{i=1}^{k} u_i \circ {f_i}'(x)$$

proof $(\Rightarrow)$ $p_i \in \mathcal{L}(F, F_i)$ (p_i is continuous since $\|p_i(x)\| = \|x_i\| \le \|x\|$), so p_i is differentiable. Therefore, $p_i \circ f = f_i$ is differentiable at x.

$(\Leftarrow)$ $\sum_{i=1}^{k} u_i \circ p_i = id_F \implies \sum_{i=1}^{k} u_i \circ \underbrace{p_i \circ f}_{f_i} = f$. Furthermore, as $u_i \in \mathcal{L}(F_i, F)$ f is differentiable as f_i is differentiable. Then, using the chain rule we obtain:

$$f'(x) = \sum_{j=1}^{k} (u_i \circ f_i)'(x) = \sum_{i=1}^{k} {u_i}'(f_i(x)) \circ {f_i}'(x) = \sum_{i=1}^{k} u_i \circ {f_i}'(x)$$

Illustration: Case where $E = \mathbb{R}^d$ et $F = \mathbb{R}^k$

From the previous section, one can write that

$$f'(x).h = \sum_{i=1}^{k} u_i \circ {f_i}'(x).h = (f_1'(x)h, \cdots, f_k'(x)h)$$

Let us then decompose h in the canonical base of $\mathbb{R}^d$, $h = \sum_{j=1}^{d} h_j e_j$. Then using the linearity of the differential, one may write:

$$f_i'(x)h = f_i'(x)(\sum_{j=1}^{d} h_j e_j) = \sum_{j=1}^{d} h_j f_i'(x)(e_j)$$

where $f_i'(x)(e_j) = \lim_{t \to 0} \frac{f_i(x+te_j) - f_i(x)}{t} = \frac{\partial f_i}{\partial x_j}(x)$. Thus one can write:

$$f'(x).h \quad = \quad (\sum_{j=1}^{d} \frac{\partial f_1}{\partial x_j}(x) h_j, \cdots, \sum_{j=1}^{d} \frac{\partial f_k}{\partial x_j}(x) h_j) = \mathrm{Jac}(f)(x)h$$

In other words, the matrix $\mathrm{J}ac(f)(x)$ is the matrix whose element indexed by (i,j), pour $i=1,...,k$ and $j=1,...,d$, is equal to $\frac{\partial f_i}{\partial x_j}(x)$. In the particular case $F=\mathbb{R}$, one may write

$$f'(x)h \quad = \quad f'(x)(\sum_{j=1}^{d} h_j e_j) = \sum_{j=1}^{d} h_j f'(x)(e_j) = \sum_{j=1}^{d} \frac{\partial f}{\partial x_j}(x) h_j = \nabla f(x)^t.h$$

One can finally rewrite the chain rule theorem in finite dimension under the following form:

Theorem 2

Let $d,n,p \geq 1$ three integers, and $U \subset \mathbb{R}^d$, $V \subset \mathbb{R}^n$ two open sets; Let $\varphi : V \to \mathbb{R}^p$ and $\psi : U \to \mathbb{R}^n$ two differentiable functions such that $\psi(U) \subset V$. Then $\varphi \circ \psi : U \to \mathbb{R}^p$ is defined and differentiable on U, and corresponds to its Jacobian matrix which reads:

$$\forall x \in U,\ \mathrm{Jac}(\varphi \circ \psi)(x) = \mathrm{Jac}(\varphi)(\psi(x))\mathrm{Jac}(\psi)(x). \tag{1.1}$$

which can be rewritten in the following way:

The Jacobian matrix of a composition of functions is the product of the Jacobian matrices.

Note that the matrix product in formula (1.1) makes sense , since the matrix $\mathrm{Jac}(\psi)(x)$ is with size $n \times m$ and $\mathrm{Jac}(\varphi)(\psi(x))$ is with size $p \times n$ (so that $\mathrm{Jac}(\varphi \circ \psi)(x)$ is with size $p \times m$). A particular case of this is the following ($p = d = 1$):

Corollary 1

Let $n \geq 1$ an integer and Let $\varphi : U \to \mathbb{R}$ a differentiable function on an open set U of $\mathbb{R}^n$. Let $\psi : I \to \mathbb{R}^n$ a differentiable function on an interval I of $\mathbb{R}$ (whose components are denoted by $\psi(t) = (\psi_1(t), ..., \psi_n(t)))$ such that $\psi(I) \subset U$. Then $f = \varphi \circ \psi : I \to \mathbb{R}$ is differentiable and

$$\forall t \in I,\ f'(t) = \frac{\partial \varphi}{\partial x_1}(\psi(t))\psi_1'(t) + ... + \frac{\partial \varphi}{\partial x_n}(\psi(t))\psi_n'(t).$$

1.3 Directional Derivative, first variation and local extrema

Definition 2 (Directional derivative) *Let E and F be two normed vector spaces, and U a non empty open set in E and $x \in U$. Let $f : U \longrightarrow F$, $v \in E$ and $t \in \mathbb{R}^*$. We say that f admits a directional derivative in the direction v at x if*

$$\lim_{t \to 0} \frac{f(x + t\,v) - f(x)}{t} \ \ \textit{exists}$$

Definition 3 (First variation) *Let E and F be two normed vector spaces. let U be a non empty open set of E and $a \in U$. Let $f : U \longrightarrow F$. Let $v \in E$. If f admits a directional derivative in each direction v at x, we call it first variation of f at x and denote it $\delta(x, v)$.*

Remark 2 *If f is differentiable at x and admits a first variation at x, but the converse is false.*

Remark 3 *If f is differentiable at x, $\delta(x, v) = f'(x)v$.*

Proposition 2 **(Necessary condition for the existence of an extremum)**

If f admits a first variation at x and an extremum at x, then $\delta(x,v)=0,\ \forall v\in E$. Thus, if f is differentiable at x and admits an extremum at x then $f'(x)=0$.

proof Consider that f admits a minimum at x then

$$\lim_{t\to 0^+}\frac{f(x+t\,v)-f(x)}{t}=\delta(x,v)\geq 0$$

but we also have:

$$\lim_{t\to 0^+}\frac{f(x-t\,v)-f(x)}{t}=-\delta(x,v)\geq 0$$

thus $\delta(x,v)=0$. The proof is the same for a maximum.

Definition 4 (Gâteaux-differentiability) *Let E and F be two normed vector spaces. Consider that f is defined in a neighborhood $V(x)$, for some x in E with values in F. f is said to be Gâteaux-differentiable (G-differentiable) at x if and only if:*

1. *f admits a first variation at x, $\delta f(x,h)$.*
2. *There exists $B\in\mathcal{L}(E,F)$ such that*

$$\delta f(x,h)=Bh$$

In this case, B will be called the G-differential of f at x. We will write $f'_G(x)=B$ this G-differential and will define the G-differential of f at x by

$$d_G f(x;h)=f'_G(x)h.$$

Proposition 3

If $f:V(x)\subset E\to F$ is differentiable at x, it is also G-differentiable at x and $f'(x)=f'_G(x)$.

proof Indeed, one can write:

$$f(x+h)=f(x)+f'(x)h+o(\|h\|)$$

Putting $h=tk$, $t\in\mathbb{R}$, $k\in E$, one obtains:

$$f(x+tk)=f(x)+tf'(x)k+o(|t|\|k\|)=f(x)+tf'(x)k+o(|t|).$$

So we deduce that $\delta f(x,h)=f'(x)h$, so f is G-differentiable and $f'_G(x)=f'(x)$.

The converse is of course wrong:
Let

$$\begin{array}{rcl} f:\mathbb{R}^2 & \to & \mathbb{R} \\ (x,y) & \to & \left\{\begin{array}{c} 1 \text{ si } y=x^2 \text{ et } x\neq 0 \\ 0 \text{ otherwise}\end{array}\right. \end{array}$$

This function is discontinuous at $(0,0)$ so it is not differentiable at this point, but:

$$\lim_{t \to 0} \frac{f(th_1, th_2) - f(0,0)}{t} = 0.$$

So the function is fonction is G-differentiable at 0.
exercise: Consider the function

$$f(x,y) \quad = \quad \left\{ \begin{array}{c} \frac{x^3 - 3xy^2}{x^2+y^2} \text{ si } x^2 + y^2 > 0 \\ 0 \text{ if } (x,y) = (0,0) \end{array} \right.$$

1. Compute $\delta f((0,0),(s,t))$

2. Is this function G-differentiable at $(0,0)$? What can you deduce from that?

However, we have the following proposition:

Proposition 4

If $f'_G(y)$ exists at every y in a neighborhood of x, and if $f'_G : X \to \mathcal{L}(E,F)$ is continuous at point x, then f is differentiable at x and $f'(x) = f'_G(x)$

proof Let us introduce the function $\varphi(t) = f(x + th)$, then $\varphi'(t) = f'_G(x + th)h$. Using the continuity of f'_G at x, we get:

$$\begin{aligned} \|\varphi(1) - \varphi(0) - \varphi'(0)\| &= \|\int_0^1 \varphi'(t) - \varphi'(0)\| \\ &\leq \sup_{t \in]0,1[} \|\varphi'(t) - \varphi'(0)\| \\ &= \sup_{t \in]0,1[} \|f'_G(x+th).h - f'_G(x).h\| \\ &\leq \sup_{t \in]0,1[} \|f'_G(x+th) - f'_G(x)\|\|h\| = o(\|h\|) \end{aligned}$$

Exercices

Ex.1
Study the differentiability of

$$f(x,y) = \left\{ \begin{array}{c} \frac{xy}{x^2+y^2} \text{ if } (x,y) \neq (0,0) \\ 0 \text{ if } (x,y) = (0,0) \end{array} \right.$$

Replace the numerator by x^2y, xy^2 et x^3y, and carry out the same study.
Ex. 2
Let $E = (C^1[0,1], \|u\| = \max(\|u\|_\infty, \|u'\|_\infty))$, $F = (C[0,1], \|u\|_\infty)$. Let $J : E \to F$ such that $J(u) = u'$. Show that $J \in C^\infty$, compute $J'(u)$.

Ex. 3
Let $E = (C[0,1], \|.\|_\infty)$, $J : E \to E$ and $J(u) = e^u$. Show that $J \in C^1$.
Ex. 4
Let $E = (C[0,1], \|.\|_\infty)$, $J : E \to \mathbb{R}$, $J(u) = \int_0^1 e^u$. Show that $J \in C^1$ using two different ways (without using the definition).
Ex. 5
We consider the following functional:

$$J(u) = e^{\int_0^1 u'(s)^2 ds}$$

defined on $C^1([0,1])$ and with values in $\mathbb{R}$. Compute the differential of J.
Ex. 6
Let E, F, G, H be normed vector spaces, $U \subset E$ open, $u : U \to F$, $v : U \to G$, $B : F \times G \to H$ bilinear and continuous. Let us put:

$$\omega(x) = B(u(x), v(x)) \quad x \in U$$

1. Show that if u and v are differentiable at $x \in U$, so is ω . Compute ω'.

2. Applications
   Consider $F = G = H = \mathbb{R}$, if $E = \mathbb{R}$, et $B(x, y) = xy$.
   If $F = \mathbb{R}$, $G = H$ et $B(x, y) = x.y$
   Finally, considering appropriate compositions, compute the differential of $z(x) = \frac{u(x)}{v(x)}$ with $G = \mathbb{R}$ and $v(x) \neq 0$ in U, F could be any Banach space, and $H = F$.

Ex. 7
Let:

$$\begin{gathered} f : \mathbb{R}^k \to \mathbb{R}^k \\ (x_1, \cdots, x_k) \to (x_1x_2, x_2x_3, \cdots, x_kx_1) \end{gathered}$$

and:

$$\begin{gathered} g : \mathbb{R}^k \to E = (C[0,1], \|.\|_\infty) \\ x \to (t \to \|x\|^2 \cos(t)) \end{gathered}$$

Show that $g \circ f$ is differentiable and compute its differential.
Ex. 8
Let E be a normed vector space and:

$$\begin{gathered} f : \mathbb{R}^k \times E \to E \\ ((x_1, \cdots, x_k), u) \to \|x\|^2 u \end{gathered}$$

Show that f is differentiable and compute this differential.

1.4 Local extrema on Convex Sets

Consider the normed vector space E, we have the following proposition:

Proposition 5

Let $X \subset E$ an open set, $C \subset X$ a convex set and $f : X \to \mathbb{R}$. If f admits a minimum at x in C and if f is differentiable at x then

$$f'(x)(y-x) \geq 0, \qquad \forall y \in C$$

proof One has for $\theta \in [0,1]$, $x + \theta(y-x) \in C$, and thus:

$$\frac{f(x+\theta(y-x)) - f(x)}{\theta} \geq 0$$

Making θ tend to zero we get the desired result.

1.5 Example of the computation of a differential in infinite dimension

Proposition 6

Recalling that $E = C^1([a,b],\mathbb{R})$ equipped with the norm $\|u\| = \max(\|u\|_\infty, \|u'\|_\infty)$. $(E, \|.\|)$ is a normed vector space. Considering $u \in E$ and $F \in C^1([a,b]\times\mathbb{R}\times\mathbb{R})$, we define for $u \in E$

$$J(u) = \int_a^b F\Big(x, u(x), u'(x)\Big)\, dx.$$

Then J belongs to C^1 and we have:

$$\forall\, h \in E,\; J'(u)(h) = \int_a^b \left(\frac{\partial F}{\partial y}\Big(x, u(x), u'(x)\Big)\, h(x) + \frac{\partial F}{\partial z}\Big(x, u(x), u'(x)\Big)\, h'(x) \right) dx$$

proof Let $y, z, s, t \in \mathbb{R}$, $x \in [a,b]$ and $r \in [0,1]$.

$$\begin{aligned} F(x, y+s, z+t) - F(x,y,z) &= \int_0^1 \frac{d}{dr}\Big(F(x, y+rs, z+rt)\Big) dr \\ &= \int_0^1 \left\{ \frac{\partial F}{\partial y}\Big(x, y+rs, z+rt\Big) s + \frac{\partial F}{\partial z}\Big(x, y+rs, z+rt\Big) t \right\} dr \end{aligned}$$

Let $\varepsilon > 0$ and $h \in E$. First, we can write

$$\begin{aligned}
J(u+h) - J(u) &= \int_a^b \Big\{F\Big(x, u(x)+h(x), u'(x)+h'(x)\Big) - F\Big(x, u(x), u'(x)\Big)\Big\}\,dx \\
&= \int_a^b \left(\int_0^1 \left\{\frac{\partial F}{\partial y}\Big(x, u(x)+rh(x), u'(x)+rh'(x)\Big)h(x)\right.\right. \\
&\quad \left.\left. + \frac{\partial F}{\partial z}\Big(x, u(x)+rh(x), u'(x)+rh'(x)\Big)h'(x)\right\}dr\right)dx \\
&= \underbrace{\int_a^b \left\{\frac{\partial F}{\partial y}\Big(x, u(x), u'(x)\Big)h(x) + \frac{\partial F}{\partial z}\Big(x, u(x), u'(x)\Big)h'(x)\right\}dx}_{L(h)} + \underbrace{\int_a^b A(x)\,dx}_{B}
\end{aligned}$$

We are going to show that $L \in \mathcal{L}(E, \mathbb{R}) = E'$ and that

$$\exists\, \eta > 0 \text{ tq } \|h\| < \eta \Longrightarrow |B| \le \varepsilon \|h\|.$$

L is clearly linear. Now let us show that it is continuous:

$$|L(h)| \le \sup_{x\in[a,b]} \left\{\left|\frac{\partial F}{\partial y}\Big(x, u(x), u'(x)\Big)\right| + \left|\frac{\partial F}{\partial z}\Big(x, u(x), u'(x)\Big)\right|\right\} \|h\|\,(b-a).$$

Furthermore,

$$\begin{aligned}
A(x) &= \int_0^1 \left\{\frac{\partial F}{\partial y}\Big(x, u(x)+rh(x), u'(x)+rh'(x)\Big) - \frac{\partial F}{\partial y}\Big(x, u(x), u'(x)\Big)\right\} h(x)\,dr \\
&\quad + \int_0^1 \left\{\frac{\partial F}{\partial z}\Big(x, u(x)+rh(x), u'(x)+rh'(x)\Big) - \frac{\partial F}{\partial z}\Big(x, u(x), u'(x)\Big)\right\} h'(x)\,dr
\end{aligned}$$

As F is C^1, there exists $\eta > 0$ such that $\|h\| \le \eta$:

$$\left|\frac{\partial F}{\partial y}\Big(x, u(x)+rh(x), u'(x)+rh'(x)\Big) - \frac{\partial F}{\partial y}\Big(x, u(x), u'(x)\Big)\right| \le \varepsilon$$
$$\left|\frac{\partial F}{\partial z}\Big(x, u(x)+rh(x), u'(x)+rh'(x)\Big) - \frac{\partial F}{\partial z}\Big(x, u(x), u'(x)\Big)\right| \le \varepsilon$$

for all $x \in [a,b]$ since u is fixed so $u([a,b])$ and $u'([a,b])$ are compact. Thus $|A(x)| \le \varepsilon\big(|h(x)| + |h'(x)|\big) \le 2\,\varepsilon\,\|h\|$ and $|B| = \left|\int_a^b A(x)\,dx\right| \le 2\,\varepsilon\,\|h\|\,(b-a)$

Application: Computation of a minimum

Now, consider the previous problem with $F \in C^2([a,b]\times\mathbb{R}^2, \mathbb{R})$, and assume we are looking for a minimum of J in

$$X = \big\{u \in C^2[a,b], u(a) = u_a,\ u(b) = u_b \text{ fixed}\big\}.$$

As X is a convex set , a minimum u_0 must satisfy:

$$J'(u_0).h = 0, \quad \forall h \in Y$$

with

$$Y = \left\{ h \in C^2[a,b], h(a) = 0,\ h(b) = 0 \right\}.$$

u_0 must satisfy $J'(u_0).h = 0$ for all h in Y. Taking into account the hypothesis made on F, one can write:

$$\forall\, h \in Y,\ J'(u_0).h = \int_a^b \left(\frac{\partial F}{\partial y}\Big(x, u_0(x), u_0'(x)\Big)\, h(x) + \frac{\partial F}{\partial z}\Big(x, u_0(x), u_0'(x)\Big)\, h'(x) \right) dx$$

By integrating by parts the second part of the integral, we get:

$$\forall\, h \in Y,\ J'(u_0).h = \int_a^b \left[\frac{\partial F}{\partial y}\Big(x, u_0(x), u_0'(x)\Big) - \frac{d}{dx}\left(\frac{\partial F}{\partial z}\Big(x, u_0(x), u_0'(x)\Big) \right) \right] h(x)\, dx \tag{1.2}$$

Furthermore, we have the following lemma

Lemma 1

let $f \in C^2([a,b])$ such that $\int_a^b f(t)h(t)dt = 0$, $\forall h \in C^2([a,b])$, satisfying $h(a) = h(b) = 0$, then $\forall t \in [a,b],\ f(t) = 0$.

proof Let us assume the contrary. If there exists a point $t_0 \in]a,b[$ such that $f(t_0) > 0$, then as f is continuous, there exists a small interval $[t_1, t_2] \subset]a,b[$ containing t_0 such that f is strictly positive on that interval. Let us then introduce the function:

$$h(t) = \begin{cases} (t-t_1)^3(t_2-t)^3 & t_1 \le t \le t_2 \\ 0 \text{ sinon.} \end{cases}$$

This function is in Y and one has:

$$\int_a^b f(t)h(t)dt = \int_{t_1}^{t_2} f(t)h(t)dt > 0$$

hence the contradiction.

One can then write that u satisfies the Euler-Lagrange (or Euler) equation associated with the problem, namely.:

$$\frac{\partial F}{\partial y}\Big(x, u_0(x), u_0'(x)\Big) - \frac{d}{dx}\left(\frac{\partial F}{\partial z}\Big(x, u_0(x), u_0'(x)\Big) \right) = 0, \quad \forall x \in [a,b].$$

This is a nonlinear second order differential equation, since one can rewrite the above equation under the following form:

$$\forall x \in [a,b]$$

$$\frac{\partial F}{\partial y}\Big(x, u_0(x), u_0'(x)\Big) - \frac{\partial^2 F}{\partial x \partial z}\Big(x, u_0(x), u_0'(x)\Big) - \frac{\partial^2 F}{\partial y \partial z}\Big(x, u_0(x), u_0'(x)\Big) u_0'(x) - \frac{\partial^2 F}{\partial^2 z}\Big(x, u_0(x), u_0'(x)\Big) u_0''(x) = 0.$$

Application: The length of a curve $y = y(t)$ (assumed to be $C^2([a,b])$) joining the points (a, y_a) and (b, y_b) is given by the integral:

$$J(y) = \int_a^b \sqrt{1 + y'(t)^2} dt$$

We try to find what the curves with minimal length in

$$X = \left\{y \in C^2([a,b]),\ y(a) = y_a,\ y(b) = y_b\right\}.$$

If y is a local minimum it must satisfy the Euler equation. Putting $L(t, y(t), y'(t)) = \sqrt{1 + y'(t)^2}$, which belongs to $C^2([a,] \times \mathbb{R}^2)$, such an equation reads in that case:

$$\frac{\partial L}{\partial y} L(t, y(t), y'(t)) - \frac{d}{dt}\frac{\partial L}{\partial y'} L(t, y(t), y'(t)) = 0 \Leftrightarrow \frac{d}{dt}\left(\frac{y'(t)}{\sqrt{1 + y'(t)^2}}\right) = 0$$

since $\frac{\partial L}{\partial y} L(t, y(t), y'(t)) = 0$. So, one has:

$$\frac{y'(t)}{\sqrt{1 + y'(t)^2}} = C \Leftrightarrow y'^2 = C^2(1 + y'^2) \Leftrightarrow y'^2 = \frac{C^2}{1 - C^2} \Leftrightarrow y' = D$$

Thus, $y = At + b$ (which are straight lines!!) and the unique potential candidate is

$$y = \frac{y_b - y_a}{b - a} t + \frac{b y_a - a y_b}{b - a}.$$

1.6 Mean value theorem

Theorem 3 (mean value theorem)

Let F be a normed vector space. Let $f : [a, b] \longrightarrow F$ and $g : [a, b] \longrightarrow \mathbb{R}$ two continuous applications on $[a, b]$ and differentiable on $]a, b[$. If

$$\forall\, t \in]a, b[,\ \left\|f'(t)\right\| \le g'(t)$$

then

$$\|f(b) - f(a)\| \le g(b) - g(a).$$

proof Let $\varepsilon > 0$. We are going to show that

$$\forall\, t \in [a, b],\ \|f(t) - f(a)\| \le g(t) - g(a) + \varepsilon(t - a) + \varepsilon \tag{1.3}$$

Using the continuity of f and g, (1.3) is true for $t \in [a, a + \eta]$ with $\eta > 0$ small.
$A = \{\eta \in]0, b - a]\ ;\ (1.3) \text{ is true for } t \in [a, a + \eta]\}$. A is non empty and bounded above, one can thus define $\widetilde{\theta} = \sup A$ and then set $\theta = \widetilde{\theta} + a$. Using a continuity argument, (1.3) is true for $t = \theta$.
Let us suppose that $\theta < b$, $\exists \delta > 0$ such that for all $t \in [\theta, \theta + \delta]$ and as f is differentiable in θ:

$$\left\|f(t) - f(\theta) - f'(\theta)(t - \theta)\right\| \le \frac{\varepsilon}{2}(t - \theta).$$

Similarly, as g is differentiable at θ,

$$\left|g(t) - g(\theta) - g'(\theta)(t - \theta)\right| \le \frac{\varepsilon}{2}(t - \theta)$$

Using triangular inequality, we get :

$$\|f(t)-f(\theta)\| \leq \|f'(\theta)\|(t-\theta)+\frac{\varepsilon}{2}(t-\theta) \leq g'(\theta)(t-\theta)+\frac{\varepsilon}{2}(t-\theta) \leq g(t)-g(\theta)+\varepsilon(t-\theta)$$

For $t \in [\theta, \theta+\delta]$:

$$\begin{aligned}\|f(t)-f(a)\| &\leq \|f(t)-f(\theta')\|+\|f(\theta')-f(a)\| \\ &\leq g(t)-g(\theta')+\varepsilon(t-\theta')+g(\theta')-g(a)+\varepsilon(\theta'-a)+\varepsilon \\ &= g(t)-g(a)+\varepsilon(t-a)+\varepsilon\end{aligned}$$

So (1.3) is true on $[\theta, \theta+\delta]$, which consists of a contradiction. So $\theta = b$ and for all $\varepsilon > 0$, one has:

$$\|f(b)-f(a)\| \leq g(b)-g(a)+\varepsilon(b-a)+\varepsilon$$

which entails $\|f(b)-f(a)\| \leq g(b)-g(a)$.

Proposition 7

Let E and F be two normed vector spaces. Let U be an open set in E, $f : U \longrightarrow F$ differentiable on U.
We assume that there exists a constant $k \geq 0$ such that $\forall\, x \in U$, $\|f'(x)\| \leq k$.
If $[x, y]$ is a segment included in U, one has :

$$\|f(y)-f(x)\| \leq k \;\|y-x\|$$

In particular, f is Lipschitz on a ball included in U and on a convex set included in U .

proof
Let us define $\tilde{f} : t \in [0,1] \longmapsto f(x+t(y-x))$ and then apply the mean value theorem to $\tilde{f}$:

$$\tilde{f}'(t) = f'(x+t(y-x))(y-x).$$

So,

$$\left\|\tilde{f}'(t)\right\| = \left\|f'(x+t(y-x))(y-x)\right\| \leq \left\|f'(x+t(y-x))\right\| \|y-x\| \leq k\,\|y-x\| = g'(t).$$

Taking $g(t) = k\,\|y-x\|\,t$, we get

$$\left\|\tilde{f}(1)-\tilde{f}(0)\right\| = \|f(y)-f(x)\| \leq g(1)-g(0) = k\,\|y-x\|$$

Corollary 2

Let $f : U \longrightarrow F$ a C^1 function. Then f is locally Lipschitz.

proof Let $x_0 \in U$, f' is continuous on a neighborhood V of x_0 and is bounded on V.

1.7 Higher order differentials

Let f be a function defined on an open set Ω of a normed vector space E taking its values in a normed vector space F. We assume that f is differentiable on Ω and thus

$$f' : \Omega \to \mathcal{L}(E, F).$$

If this function is also differentiable on Ω, f is said to be two times differentiable on Ω, and the second order differential is a function

$$f'' : \Omega \to \mathcal{L}(E, \mathcal{L}(E, F)).$$

The space $\mathcal{L}(E, \mathcal{L}(E, F))$ can be identified to $\mathcal{L}_2(E, F)$, the space of continuous bilinear applications from E onto F (in fact the spaces are isomorphic), so the second order differential in $a \in \Omega$, $f''(a)$ belongs to $\mathcal{L}_2(E, F)$, and one often denotes it by $D^2 f(a)$.
To understand well the notations, let us come back to the computation of the second order derivative. One can write

$$f'(a + h) = f'(a) + f''(a)h + o(\|h\|),$$

by definition. So $f''(a)h \in \mathcal{L}(E, F)$. This linear application evaluated at $k \in E$ should be written as $(f''(a)h)(k)$, which we rewrite as $f''(a)(h, k)$. If f is two time differentiable at a then $f''(a)$ is symmetric (Schwarz symmetry theorem). In practice, to compute $f''(a)(h, k)$ one differentiates the application $x \to f'(x)h$ for h in E, at $x = a$.
In the particular case where f is a function from $\mathbb{R}^n$ to $\mathbb{R}$, $f''(a)$ written in the canonical basis of $\mathbb{R}^n$ is a matrix called Hessian matrix of f at a and denoted by $Hf(a)$. The coefficients of the matrix $Hf(a)$ are the second order partial derivatives $\frac{\partial^2 f}{\partial x_i \partial x_j}(a)$, $1 \leq i, j \leq n$, and one has $\frac{\partial^2 f}{\partial x_i \partial x_j}(a) = \frac{\partial^2 f}{\partial x_j \partial x_i}(a)$.
Going further, if f is r times differentiable at a, the differential of order r at a is an r-linear application on E with values in F, which we denote by $D^r f(a)$, for all h in E we will denote $D^r f(a)(h, \cdots, h) = D^r f(a) h^r$.

1.7.1 Taylor-Young formula

One can generalize Taylor-Young formula for functions from E to F, assumed to be vector spaces. To start with we will need the following lemma:

Lemma 2

Let E and F be normed vector spaces, $\phi : E^k \to F$ continuous k-linear and symmetric and $\Phi : E \to F$ defined by $\Phi(x) = \phi(x^k)$. Then Φ is differentiable and

$$D\Phi(x)h = k\phi(x^{k-1}, h)$$

for x and h in E

proof We have

$$\begin{aligned} \Phi(x + h) &= \phi(x + h, \cdots, x + h) \\ &= \phi(x^k) + k\phi(x^{k-1}, h) + \text{ terms of the form } \phi(x^p, h^q) \end{aligned}$$

with $p+q=k$ and $q \geq 2$. The mapping $h \to k\phi(x^{k-1}, h)$ is linear and continuous. Also,

$$\|\phi(x^p, h^q)\| \leq \|\phi\|\|x\|^p\|h\|^q$$

and the result follows.

Theorem 4 (Taylor-Young)

Let $U \subset E$, an open set and f an application from U to F. If f is r times differentiable at $a \in U$, then it admits a Taylor-Young expansion of order k at point a, meaning there exists a function $\varepsilon : E \to F$, with $\lim_{\|h\| \to 0} \varepsilon(h) = 0$, such that:

$$f(a+h) = f(a) + \sum_{r=1}^{k} \frac{1}{r!} D^r f(a)(h^r) + o(\|h\|^k)$$

proof We will prove this result by induction on k. First, by the definition of the differential, it is true for $k=1$. We now suppose that it is true up to order $k-1$ and consider the case k. We set:

$$\phi(x) = f(a+x) - f(a) - Df(a)x - \frac{1}{2}D^2 f(a)(x^2) - \cdots - \frac{1}{k!}D^k f(a)(x^k).$$

Using Lemma 2, we obtain:

$$D\phi(x)h = Df(a+x)h - Df(a)h - D^2 f(a)(x,h) - \cdots - \frac{1}{(k-1)!}D^k f(a)(x^{k-1}, h).$$

By hypothesis, for the mapping $Df : U \to \mathcal{L}(E,F)$, we can write:

$$Df(a+x) = Df(a) + D(Df)(a)x + \frac{1}{2}D^2(Df)(a)(x^2) + \cdots + \frac{1}{(k-1)!}D^{k-1}(Df)(a)(x^{k-1}) + o(\|x\|^{k-1}),$$

therefore

$$Df(a+x)h = Df(a)h + D^2 f(a)(x,h) + \frac{1}{2}D^3 f(a)(x^2, h) + \cdots + \frac{1}{(k-1)!}D^k f(a)(x^{k-1}, h) + o(\|x\|^{k-1})h.$$

Hence $D\phi(x)h = o(\|x\|^{k-1})h$ and so $D\phi(x) = o(\|x\|^{k-1})$. Let us fix $\varepsilon > 0$. From what we have just seen, there exists $\delta > 0$ such that $\|D\phi(x)\|_{\mathcal{L}(E,F)} < \varepsilon\|x\|^{k-1}$ if $\|x\| < \delta$. From this we deduce that:

$$\|\phi(x)\| = \|\phi(x) - \phi(0)\| \leq \varepsilon\|x\|^k$$

or

$$\|f(a+h) - f(a) - \sum_{r=1}^{k} \frac{1}{r!} D^r f(a)(h^r)\| \leq \varepsilon\|h\|^k$$

It follows that

$$f(a+h) = f(a) + \sum_{r=1}^{k} \frac{1}{r!} D^r f(a)(h^r) + o(\|h\|^k)$$

Hence the result is true for k. This ends the proof.

1.7.2 Taylor-Lagrange formula

Before we introduce Taylor-Lagrange formula, we need the following lemma

Lemma 3

Let E and F be normed vector spaces, U an open subset of E and $f : U \to F$ a $(k+1)$-differentiable mapping. Suppose that $a \in U$ and $x \in E$ are such that the segment $[a, a+x] \in U$. Then the mapping

$$\begin{array}{ll} \phi : & [0,1] \to F, \\ & t \to f(a+tx) + \sum_{r=1}^{k} \frac{(1-t)^r}{r!} D^r f(a+tx)(x^r) \end{array}$$

is continuous on $[0,1]$ and differentiable on $]0,1[$ with

$$\phi'(t) = \frac{(1-t)^k}{k!} D^{k+1} f(a+tx)(x^{k+1})$$

proof There is no difficulty in seeing that ϕ is continuous on $[0,1]$. For $r = 1, \cdots, k$, we set

$$\phi_r(x) = D^r f(x)(x^r),$$

which is differentiable and

$$\frac{dD^r f(a+tx)(x^r)}{dt} = \frac{d}{dt}\phi_r(a+tx) = D\phi_r(a+tx)x = D^{r+1} f(a+tx)(x^{r+1}).$$

Finally since

$$\frac{d}{dt}\frac{(1-t)^r}{r!} D^r f(a+tx)(x^r) = \frac{(1-t)^r}{r!} D^{r+1} f(a+tx)(x^{r+1}) - \frac{(1-t)^{r-1}}{(r-1)!} D^r f(a+tx)(x^r).$$

Hence the result.

Theorem 5 (Taylor-Lagrange)

Let E and F be normed vector spaces. U an open subset of E and $f : U \to F$ a $(k+1)$-differentiable mapping. Suppose that $a \in U$ and that $x \in E$ is such that the segment $[a, a+x]$ is contained in U. Then

$$f(a+x) = f(a) + \sum_{r=1}^{k} \frac{1}{r!} D^r f(a)(x^r) + \mathcal{R}(a,x)$$

where

$$\|\mathcal{R}(a,x)\| \leq \frac{1}{(k+1)!} \sup_{0 \leq \lambda \leq 1} \|D^{k+1} f(a+\lambda x)\|_{\mathcal{L}_{k+1}(E,F)} \|x\|^{k+1}.$$

proof If $\sup_{0\leq\lambda\leq 1} \|D^{k+1}f(a+\lambda x)\|_{\mathcal{L}_{k+1}(E,F)} = \infty$, we have nothing to prove, so let us assume that this is not the case. Let ϕ be defined as in the preceding lemma. Then

$$\|\phi'(t)\| \leq \frac{(1-t)^k}{k!} \sup_{0\leq\lambda\leq 1} \|D^{k+1}f(a+\lambda x)\|_{\mathcal{L}_{k+1}(E,F)} \|x\|^{k+1} = \frac{(1-t)^k}{k!}C$$

If we set

$$\psi(t) = -\frac{(1-t)^{k+1}}{(k+1)!}C$$

then

$$\psi'(t) = \frac{(1-t)^k}{k!}C,$$

and thus using the mean value theorem we may write

$$\|\phi(1)-\phi(0)\| \leq \psi(1)-\psi(0) = \frac{C}{(k+1)!}$$

Observing that

$$\phi(1)-\phi(0) = f(a+x) - f(a) - \sum_{r=1}^{k} \frac{1}{r!} D^r f(a)(x^r),$$

we obtain the result.

When $F=\mathbb{R}$, we have a simpler form for the remainder.

Theorem 6 (Taylor-Lagrange 2)

Let E be a normed vector space, U an open subset of E and $f : U \to \mathbb{R}$ a $(k+1)$-differentiable function. Suppose that $a \in U$ and that $x \in E$ is such that the segment $[a, a+x]$ is contained in U. Then there is a real number $\theta \in]0,1[$ such that

$$f(a+x) = f(a) + \sum_{r=1}^{k} \frac{1}{r!} D^r f(a)(x^r) + \frac{1}{(k+1)!} D^{k+1}(a+\theta x)(x^{k+1})$$

proof If we set $g(t) = f(a+tx)$, then g has continuous derivatives up to order k on $[0,1]$ and a $(k+1)th$ derivative on $]0,1[$. It is easy to prove by induction that

$$g^{(r)}(t) = D^r f(a+tx)(x^r)$$

From Taylor's formula for a function defined on a compact interval of $\mathbb{R}$, we know that there is a real number $\theta \in]0,1[$ such that

$$g(1) = g(0) + \sum_{r=1}^{k} \frac{1}{r!} g^{(r)}(0) + \frac{1}{(k+1)!} g^{(k+1)}(\theta).$$

or

$$f(a+x) = f(a) + \sum_{r=1}^{k} \frac{1}{r!} D^r f(a)(x^r) + \frac{1}{(k+1)!} D^{k+1}(a+\theta x)(x^{k+1}).$$

This ends the proof.

1.7.3 Taylor formula with integral remainder

Taylor formula with integral remainder gives an explicit formula for the remainder.

Theorem 7

Let E and F be normed vector spaces with F complete. U an open subset of E and $f : U \to F$ of class C^{k+1}. If $a \in U$ and $x \in E$ is such that the segment $[a, a+x]$ is contained in U then

$$f(a+x) = f(a) + \sum_{r=1}^{k} \frac{1}{r!} D^r f(a)(x^r) + \int_0^1 \frac{(1-t)^k}{k!} D^{k+1} f(a+tx)(x^{k+1})dt$$

proof Let ϕ be the mapping defined as in Lemma 3. As f is of class C^{k+1}, ϕ is of class C^1 on an open interval containing $[0,1]$. Using the fundamental theorem of calculus, we have

$$\phi(1) = \phi(0) + \int_0^1 \phi'(t)dt$$

or

$$f(a+x) = f(a) + \sum_{r=1}^{k} \frac{1}{r!} D^r f(a)(x^r) + \int_0^1 \frac{(1-t)^k}{k!} D^{k+1} f(a+tx)(x^{k+1})dt$$

which is the result we are looking for.

Chapter 2

Image restoration

Image restoration is the process that corrects degraded images and reconstructs a good quality image from the latter. In this chapter, we are interested in the restoration of blurred images. Before we start with the description of the different techniques, we need to introduce some background on some quantities very often used in image processing.

2.1 Autocorrelation function and power spectral density

Definition 1 *One can associate to a stochastic process $f(t)$ its statistical autocorrelation defined by:*

$$R_f(t,\tau) = \mathbb{E}[f(t)f^*(t-\tau)].$$

For a deterministic signal, the autocorrelation is defined as:

$$R_f(\tau) = \int_{\mathbb{R}^2} f(t)f^*(t-\tau)dt.$$

A white noise is the simplest example of second order wide-sense stationary process, which corresponds to the following definition:

Definition 2 *A stochastic process is wide-sense stationary if it satisfies the following two properties:*

1. *Its expectation is independent of t.*
2. *The autocorrelation function depends only on τ but not on t.*

In the case of a wide-sense stationary process, one can thus write:

$$R_f(\tau) = \mathbb{E}[f(t)f^*(t-\tau)],$$

and for a white noise b, one has

$$R_b(\tau) = N_0^2\delta_0.$$

where N_0^2 is the power spectral density of the noise and δ_0 the Dirac distribution in 0.

Definition 3 *We have the following definitions:*

- *For a deterministic signal, one defines the energy spectral density* Γ_f *as the Fourier transform of the auto-correlation function, one can show that* $\Gamma_f(\xi) = |\hat{f}(\xi)|^2$.
- *For a stochastic signal, one defines the power spectral density* Γ_f *of the process* f *as the Fourier transform of the autocorrelation function.*

proof The above proposition is called the Wiener-Kintchine theorem. In the deterministic case, the proof is as follows:

$$\begin{aligned}
\Gamma_f(\xi) &= \int_{\mathbb{R}^2}\int_{\mathbb{R}^2} f(t)f^*(t-\tau)e^{-2i\pi\langle\xi,\tau\rangle}\,dt\,d\tau \\
&= \int_{\mathbb{R}^2}\left(\int_{\mathbb{R}^2} f(t+\tau)e^{-2i\pi\langle\xi,(t+\tau)\rangle}d\tau\right) f^*(t)e^{2i\pi\langle\xi,t\rangle}\,dt \\
&= \int_{\mathbb{R}^2}\left(\int_{\mathbb{R}^2} f(v)e^{-2i\pi\langle\xi,v\rangle}dv\right) f^*(t)e^{2i\pi\langle\xi,t\rangle}\,dt = \hat{f}(\xi)\int_{\mathbb{R}^2} f^*(t)e^{2i\pi\langle\xi,t\rangle}\,dt = |\hat{f}(\xi)|^2.
\end{aligned}$$

Remark 1 *In the case of a white noise, the autocorrelation function is a Dirac distribution so one has* $\Gamma_b = N_0^2$.

Similarly, we define $\Gamma_{f,g}$ the Fourier transform of the cross-correlation function defined in the deterministic case as:

$$C_{f,g} = \int_{\mathbb{R}^2} f(t)g^*(t-\tau)dt$$

and, in the stochastic case, as

$$C_{f,g}(t,\tau) = \mathbb{E}[f(t)g^*(t-\tau)].$$

Note that we say that f and g are jointly wide-sense stationary (WSS) processes when $C_{f,g}(t,\tau) = C_{f,g}(\tau)$.
Let us now consider that f is a bidimensional discrete stochastic process, then, as in the continuous case, we have the following definitions:

Definition 4 *The autocorrelation function of a discrete stochastic process* f_n *is defined as:*

$$R_f(n,k) = \mathbb{E}[f_k f_{k-n}^*].$$

For a deterministic signal, the autocorrelation corresponds to:

$$R_f(n) = \sum_{k\in\mathbb{Z}^2} f_k f_{k-n}^*$$

Definition 5 *Let* (f_n) *be a discrete signal in* $l_1(\mathbb{Z}^2)$, *one calls* discrete-space Fourier transform *(DSFT) the function:*

$$\hat{f}_d(\xi) = \sum_{n\in\mathbb{Z}^2} f_n e^{-2i\pi\langle n,\xi\rangle}, \tag{2.1}$$

Remark 2 *This corresponds to the Fourier transform in the sense of distributions of the bidimensional Dirac comb.*

Remark 3 *$\hat{f}_d$ is a continuous function and belongs to $L^2([0,1]^2)$ since from Fourier series theory one has:*

$$f_n = \int_{[0,1]^2} \hat{f}_d(\xi) e^{2i\pi\langle n,\xi\rangle} d\xi$$

As in the continuous space framework:

Definition 6 *one can then define:*

- *For a deterministic signal, the energy spectral density Γ_f is the DSFT of the autocorrelation function and one has: $\Gamma_f(\xi) = |\hat{f}_d(\xi)|^2$*
- *For a stochastic process, the power spectral density is the DSFT of the autocorrelation function.*

proof Assuming f in $l_1(\mathbb{Z}^2)$, we have

$$\begin{aligned} \Gamma_f(\xi) &= \sum_{n\in\mathbb{Z}^2}\sum_{p\in\mathbb{Z}^2} f_p f^*_{p-n} e^{-2i\pi\langle n,\xi\rangle} = \sum_{n\in\mathbb{Z}^2}\sum_{p\in\mathbb{Z}^2} f_{p+n} f^*_p e^{-2i\pi\langle n,\xi\rangle} \\ &= \sum_{p\in\mathbb{Z}^2} f^*_p (\sum_{n\in\mathbb{Z}^2} f_{p+n} e^{-2i\pi\langle p+n,\xi\rangle}) e^{2i\pi\langle p,\xi\rangle} = |\widehat{f}_d(\xi)|^2 \end{aligned}$$

Definition 7 *A discrete stochastic process f is said to be wide-sense stationary if it satisfies the following two properties:*

1. *$\mathbb{E}[f_n]$ is independent of n.*
2. *The autocorrelation function $R_f(n,k) = R_f(k)$.*

Remark 4 *In the case of a white noise, one has $R_b(k) = N_0^2\delta_{0,k}$ where $\delta_{i,j}$ is the kronecker symbol. In that case, we get $\Gamma_b(\xi) = N_0^2$.*

2.2 Image restoration using Wiener Filtering

One assumes an infinite image is damaged in the following way:

$$\tilde{u} = u * h + b,$$

where the image u is assumed to be a bidimensionnal wide-sense stationary process and h belongs to $L^1(\mathbb{R}^2) \bigcap L^2(\mathbb{R}^2)$, and b is a null average white noise, with *power spectral density* N_0^2. Note that in that framework $\tilde{u}$ is also WSS.

Coming back to the problem of recovering u from $\tilde{u}$, Wiener deconvolution consists in filtering $\tilde{u}$ to obtain $u_r = w * \tilde{u}$ (which is also WSS), where w is supposed to belong to $L^1(\mathbb{R}^2) \cap L^2(\mathbb{R}^2)$. The main idea is to find w that minimizes:

$$\begin{aligned} \mathbb{E}\left[(u(x) - u_r(x))^2\right] &= \mathbb{E}\left[e(x)^2\right] = \mathbb{E}\left[(u(x) - w * (u * h + b)(x))^2\right] \\ &= \mathbb{E}\left[(u - w * u * h)(x)^2 - 2(u(x) - w * u * h(x))(w * b)(x) + (w * b)(x)^2\right] \\ &= \mathbb{E}\left[(u - w * u * h)(x)^2\right] + \mathbb{E}\left[(w * b)(x)^2\right]. \end{aligned}$$

The last equality being obtained remarking b is with null average, and also because the noise is independent from the signal. Our goal is first to find a simpler expression for the above expectation, before finding the optimal w.
For that purpose let us first recall that for a real wide-sense stationary process f (assuming $\mathbb{E}\left[f(x)f(x-\tau)\right]$ has a sufficiently fast decay, typically is in $L^1(\mathbb{R}^2) \cap L^2(\mathbb{R}^2)$):

$$\Gamma_f(\xi) = \int_{\mathbb{R}^2} \mathbb{E}\left[f(x)f(x-\tau)\right] e^{-2i\pi\langle\tau,\xi\rangle} d\tau$$

Then, taking the inverse transform:

$$\mathbb{E}\left[f(x)f(x-\tau)\right] \quad = \quad \int_{\mathbb{R}^2} \Gamma_f(\xi) e^{2i\pi\langle\tau,\xi\rangle} d\xi$$

From this we may deduce that

$$\mathbb{E}\left[(u(x)-u_r(x))^2\right] \quad = \quad \int_{\mathbb{R}^2} \Gamma_{u-w*u*h}(\xi) + \Gamma_{w*b}(\xi) d\xi$$

Now we remark that

$$\mathbb{E}\left[(w*b)(x)(w*b)(x-\tau)\right] = \mathbb{E}\left[\int_{\mathbb{R}^2} w(q)b(x-q)dq \int_{\mathbb{R}^2} w(r)b(x-\tau-r)dr\right]$$
$$= \int_{\mathbb{R}^2}\int_{\mathbb{R}^2} w(q)w(r)\mathbb{E}\left[b(x-q)b(x-\tau-r)\right] drdq = \int_{\mathbb{R}^2}\int_{\mathbb{R}^2} w(q)w(r)R_b(x-q,\tau-(q-r))drdq$$
$$= \int_{\mathbb{R}^2}\int_{\mathbb{R}^2} w(q)w(r)R_b(\tau-(q-r))drdq = \int_{\mathbb{R}^2}\left(\int_{\mathbb{R}^2} w(q)w(q-p)dq\right) R_b(\tau-p)dp$$
$$= \int_{\mathbb{R}^2} R_w(p)R_b(\tau-p)dp = R_w * R_b(\tau).$$

From this we deduce that:

$$\Gamma_{w*b}(\xi) = \int_{\mathbb{R}^2} R_w * R_b(\tau) e^{-2i\pi\langle\tau,\xi\rangle} d\tau = \Gamma_w(\xi)\Gamma_b(\xi) = |\hat{w}(\xi)|^2 N_0^2.$$

Furthermore we may also write:

$$\Gamma_{u-w*u*h}(\xi) \quad = \quad \Gamma_u(\xi) - \Gamma_{u,w*u*h}(\xi) - \Gamma_{w*u*h,u}(\xi) + \Gamma_{w*u*h}(\xi).$$

Since one has:

$$\mathbb{E}\left[u(x)(w*u*h)(x-\tau)\right] = \int_{\mathbb{R}^2} w(q) \int_{\mathbb{R}^2} h(p)\mathbb{E}\left[u(x)u(x-\tau-q-p)\right] dpdq$$
$$= \int_{\mathbb{R}^2} w(q)\int_{\mathbb{R}^2} h(p)R_u(\tau+q+p)dpdq = \int_{\mathbb{R}^2} \bar{w}(q)\int_{\mathbb{R}^2} \bar{h}(p)R_u(\tau-q-p)dpdq$$
$$= \bar{w} * \bar{h} * R_u(\tau),$$

with $\bar{w}(x) = w(-x)$ (and similarly for h). One can then deduce that:

$$\Gamma_{u,w*u*h}(\xi) = \hat{w}(\xi)^*\hat{h}(\xi)^*\Gamma_u(\xi).$$

Similarly we may write:

$$\mathbb{E}\left[(w * u * h)(x)u(x-\tau)\right] = \bar{w} * \bar{h} * R_u(-\tau)$$

meaning that

$$\Gamma_{w*u*h,u}(\xi) = \hat{w}(\xi)\hat{h}(\xi)\Gamma_u(\xi).$$

This leads us to:

$$\begin{aligned}\Gamma_{u-w*u*h}(\xi) &= \Gamma_u(\xi) - 2\Re(\hat{w}(\xi)\hat{h}(\xi))\Gamma_u(\xi) + |\hat{w}(\xi)|^2|\hat{h}(\xi)|^2\Gamma_u(\xi)\\ &= |1-\hat{w}(\xi)\hat{h}(\xi)|^2\Gamma_u(\xi)\end{aligned}$$

and thus

$$\mathbb{E}\left[(u(x)-u_r(x))^2\right] = \int_{\mathbb{R}^2}\Gamma_{u-w*u*h}(\xi) + \Gamma_{w*b}(\xi)d\xi = \int_{\mathbb{R}^2}\Gamma_u(\xi)|1-\hat{w}(\xi)\hat{h}(\xi)|^2 + |\hat{w}(\xi)|^2N_0^2d\xi$$

Differentiating with respect to $\hat{w}$ we get:

$$\begin{aligned}D_{\hat{w}}\mathbb{E}\left[(u(x)-u_r(x))^2\right].\hat{v} &= -\int_{\mathbb{R}^2}\Gamma_u(\xi)2\Re\left((1-\hat{w}(\xi)\hat{h}(\xi))\hat{h}(\xi)^*\hat{v}(\xi)^*\right) + 2\Re(\hat{w}(\xi)\hat{v}(\xi)^*)N_0^2d\xi\\ &= 2\Re\left(\int_{\mathbb{R}^2}(-\Gamma_u(\xi)(1-\hat{w}(\xi)^*\hat{h}(\xi)^*)\hat{h}(\xi) + \hat{w}(\xi)^*N_0^2)\hat{v}(\xi)d\xi\right)\end{aligned}$$

which is null if, for all ξ:

$$-\Gamma_u(\xi)(\hat{h}(\xi) - \hat{w}(\xi)^*|\hat{h}(\xi)|^2) + \hat{w}(\xi)^*N_0^2 = 0$$

$$\Leftrightarrow \hat{w}(\xi)^* = \frac{\Gamma_u(\xi)\hat{h}(\xi)}{\Gamma_u(\xi)|\hat{h}(\xi)|^2+N_0^2} \Leftrightarrow \hat{w}(\xi) = \frac{\Gamma_u(\xi)\hat{h}(\xi)^*}{\Gamma_u(\xi)|\hat{h}(\xi)|^2+N_0^2}$$

Finally we remark that:

$$\Gamma_{\tilde{u}}(\xi) = \Gamma_u(\xi)|\hat{h}(\xi)|^2 + N_0^2$$

so that we finally get:

$$\hat{w}(\xi) = \frac{1}{\hat{h}(\xi)}\frac{\Gamma_{\tilde{u}}(\xi)-N_0^2}{\Gamma_{\tilde{u}}(\xi)}.$$

In the absence of noise, one obtains the so-called inverse filter. Finally, one must pay great attention to singular values of the inverse filter and proceed as previously to deal with that matter.

2.3 Wiener filtering: discrete space setting

We, this time, study the discrete formalism to Wiener filtering, that is we consider:

$$\tilde{u} = u * h + b$$

where $u * h_n = \sum_{k\in\mathbb{Z}^2} h_{n-k}u_k$, where h is supposed to belong to $l^1(\mathbb{Z}^2)$ (so also to $l^2(\mathbb{Z}^2)$), and b is a white noise, with spectral density N_0^2 and u is also supposed to be a wide sense stationary process (whose definition is just below).

Similarly to what was done in the continuous case, we seek a discrete filter (w_n) such that $u_r = w * \tilde{u}$ is the closest of u as possible in the following sense:

$$\varepsilon = \mathbb{E}\left[(u_n - (u_r)_n)^2\right].$$

Note that if $((R_f)_k)$ is in $l^1(\mathbb{Z}^2)$, $\Gamma_f(\xi)$ is continuous and using the theory of Fourier series we have that:

$$(R_f)_k = \int_{[0,1]^2} \Gamma_f(\xi) e^{2i\pi\langle k,\xi\rangle} d\xi$$

meaning that:

$$\begin{aligned}
\mathbb{E}\left[(u_n - (u_r)_n)^2\right] &= \int_{[0,1]^2} \Gamma_{u-u_r}(\xi) d\xi \\
&= \int_{[0,1]^2} \Gamma_{u-w*u*h}(\xi) + \Gamma_{w*b}(\xi) \\
&= \int_{[0,1]^2} |1 - \hat{w}_d(\xi)\hat{h}_d(\xi)|^2 \Gamma_u(\xi) + |\hat{w}_d(\xi)|^2 N_0^2 d\xi
\end{aligned}$$

Differentiating with respect to $\hat{w}_d$ (in the space $L^2([0,1]^2)$), we get as previously:

$$\hat{w}_d(\xi) = \frac{1}{\hat{h}_d(\xi)} \frac{\Gamma_{\tilde{u}}(\xi) - N_0^2}{\Gamma_{\tilde{u}}(\xi)}.$$

2.4 Wiener filter finite setting

The previous analysis assumes images are of infinite size which is definitely unrealistic. In image processing, one very often makes the assumption that the image is periodic in each of its directions. This enables us to replace the restoration problem on $\mathbb{Z}^2$ by the same type of analysis but on a finite domain. The key ingredient for this transition is the circular convolution.

Definition 8 *The circular convolution between sequences h et u both with size (N,N) is defined by:*

$$(h \circledast u)_n = \sum_{0 \le k_1, k_2 \le N-1} u_k h_{n \bmod (N,N) - k} = \sum_{0 \le k_1, k_2 \le N-1} u_{(n \bmod (N,N) - k)} h_k \tag{2.2}$$

Remark 5 *This sequence is itself periodic with period (N,N).*

Proposition 1

Let u a periodic signal with period (N,N), define $h_{N,n} = \sum_{k \in \mathbb{Z}^2} h_{n-kN}$, then one has the following property:

$$(h * u)_n = (h_N \circledast u)_n.$$

To summarize, one seeks to recover the image u, assumed to be periodic and wide sense stationary, from $\tilde{u}$ which has the following expression:

$$\tilde{u} = u \circledast h_N + b \tag{2.3}$$

and we look for a periodic filter w (period N in each direction) , and define $u_r = w \circledast \tilde{u}$ so as to minimize, the expectation:

$$\varepsilon_d = \mathbb{E}\left[(u_n - (u_r)_n)^2\right].$$

Image restoration in that context uses the discrete Fourier transform (DFT):

Definition 9 *The discrete Fourier transform of a bidimensional sequence f_n of size (N_1, N_2) is defined by:*

$$\hat{f}_k = \sum_{\substack{0 \leq n_1 \leq N_1 - 1 \\ 0 \leq n_2 \leq N_2 - 1}} f_{n_1,n_2} e^{-\frac{2i\pi n_1 k_1}{N_1}} e^{-\frac{2i\pi n_2 k_2}{N_2}}, \tag{2.4}$$

and its inverse discrete Fourier transform is equal to:

$$f_n = \frac{1}{N_1 N_2} \sum_{\substack{0 \leq k_1 \leq N_1 - 1 \\ 0 \leq k_2 \leq N_2 - 1}} \hat{f}_k e^{\frac{2i\pi n_1 k_1}{N_1}} e^{\frac{2i\pi n_2 k_2}{N_2}} \tag{2.5}$$

Proposition 2

The DFT satisfies the following property:

$$(f \circledast h)_n \overset{TFD}{\to} F_k H_k.$$

We have the following definitions:

Definition 10 *The autocorrelation function of a discrete stochastic periodic process f_n is defined as:*

$$R_f(n,k) = \mathbb{E}[f_k f^*_{k-n}].$$

and has the same period as f. For a deterministic signal with period (N,N), the autocorrelation corresponds to:

$$R_f(n) = \sum_{k \in S} f_k f^*_{k-n}$$

with $S = \{(k_1,k_2), 0 \leq k_1, k_2 \leq N-1\}$.

Definition 11 *In such a case, we have that:*

- *For a deterministic signal, the energy spectral density Γ_f is the DFT of the autocorrelation function and we have $(\Gamma_f)_k = |F_k|^2$.*

- *For a stochastic signal, the power spectral density Γ_f is the DFT of the autocorrelation function.*

So we have as in the previous two settings:

$$\begin{aligned}
\mathbb{E}\left[(u_n-(u_r)_n)^2\right] &= \frac{1}{N^2}\sum_{k\in S}(\Gamma_{u-u_r})_k \\
&= \frac{1}{N^2}\sum_{k\in S}(\Gamma_{u-w\circledast u\circledast h})_k+(\Gamma_{w\circledast b})_k \\
&= \frac{1}{N^2}\sum_{k\in S}(\Gamma_u)_k-2\Re(W_kH_k)(\Gamma_u)_k+|W_k|^2|H_k|^2(\Gamma_u)_k+|W_k|^2N_0^2
\end{aligned}$$

So if we compute the differential with respect to W_k we get:

$$W_k=\frac{1}{H_k}\frac{(\Gamma_{\tilde{u}})_k-N_0^2}{(\Gamma_{\tilde{u}})_k}.$$

In this context, one can write the Wiener filters using the FFT. Typically, the Matlab implementation is then as follows, assuming the support of h is $\{-m,\cdots,m\}$ and getting rid of singular values in the filtre H using a parameter n.
We give a simple illustration

```
load gatlin2;

% the image is loaded in the variable X
% other images clown; mandrill;
% visualization


imagesc(X);
colormap(gray);
[M,N] = size(X);

%white noise generation

sigma = 0.1;
J = sigma*randn(M,N);

% filter design

h = ones(5,5)/25;

%transformation to frequency domain

Freq_X = fft2(X);
Freq_h = fft2(h,M,N);
B = real(ifft2(Freq_X.*Freq_h)) + J;
figure
imagesc(B);
colormap(gray);

%Wiener filtering
```

```
Freq_B = fft2(B);
pow_B = abs(Freq_B).^2/(M*N);
```

%generation of the inverse filter

```
gamma = 50;
sFreq_h = Freq_h.* (abs(Freq_h) > 0)+ 1/gamma * (abs(Freq_h) == 0);
iFreq_h  = 1./sFreq_h;
iFreq_h  = iFreq_h.*(abs(sFreq_h)*gamma >1)...
                 +gamma*abs(sFreq_h).*iFreq_h.* (abs(sFreq_h)*gamma <= 1);
```

%construction of Wiener filter

```
W = iFreq_h.*(pow_B-sigma^2)./pow_B;
Xr = ifft2(Freq_B.*W);
figure
imagesc(Xr);
colormap(gray);
```

2.5 Relation between Wiener filter and orthogonality

Coming back to the initial formulation

$$\varepsilon_d = \mathbb{E}\left[(u_n - (u_r)_n)^2\right] = \mathbb{E}\left[(u_n - \sum_{k\in S} w_k \tilde{u}_{n-k})^2\right]$$

Finding the best filter corresponds to finding w_k such that:

$$\begin{aligned} \frac{\partial \varepsilon_d}{\partial w_k} &= 2\mathbb{E}\left[(u_n - \sum_{p\in S} w_p \tilde{u}_{n-p})\tilde{u}_{n-k}\right] \\ &= 2\mathbb{E}\left[e_n \tilde{u}_{n-k}\right] = 0 \end{aligned}$$

So Wiener filtering corresponds to the orthogonality of the error with the data.

2.6 Image Denoising with PDEs: isotropic operators

In the following we assume that $\tilde{u} = u + b$, where b is a white Gaussian noise.

2.6.1 On the relation between the heat equation and the Gaussian kernel

For any function, $u(t,.)$ in $L^1(\mathbb{R}^2)$, its Fourier transform is defined by:

$$F(t,\xi) = \int_{\mathbb{R}^2} u(t,x) e^{-2i\pi\langle \xi, x\rangle} dx. \tag{2.6}$$

Then, let us consider the heat equation:

$$\left\{ \begin{array}{rcl} \frac{\partial u(t,x)}{\partial t} &=& \Delta u(t,x) \text{ on } \mathbb{R}^2 \\ u(0,x) &=& \tilde{u}(x), \end{array} \right. \tag{2.7}$$

where $\tilde{u} \in L^1(\mathbb{R}^2)$. Now assume $\frac{\partial u(.,x)}{\partial t}$ is continuous for all t and almost every x, and that $|\frac{\partial u(.,x)}{\partial t}| \leq g(x) \in L^1(\mathbb{R}^2)$ for all t, then we can write that:

$$\frac{\partial F(t,\xi)}{\partial t} = \int_{\mathbb{R}^2} \frac{\partial u(t,x)}{\partial t} e^{-2i\pi\langle \xi,x\rangle} dx.$$

Then, if $u(t,.)$ belongs to $C^2(\mathbb{R}^2) \bigcap L^1(\mathbb{R}^2)$ and is such that $\frac{\partial u}{\partial x_1}$ and $\frac{\partial u}{\partial x_2}$ are in $L^1(\mathbb{R}^2)$ as well as $\frac{\partial^2 u}{\partial^2 x_1}$ and $\frac{\partial^2 u}{\partial^2 x_2}$, one can finally write that:

$$\int_{\mathbb{R}^2} \left(\frac{\partial^2 u}{\partial^2 x_1} + \frac{\partial^2 u}{\partial^2 x_2} \right) e^{-2i\pi\langle \xi,x\rangle} dx = -4\pi^2(\xi_1^2 + \xi_2^2) F(\xi,t).$$

Indeed, one has:

$$\int_{\mathbb{R}^2} \frac{\partial^2 u}{\partial^2 x_1} e^{-2i\pi\langle \xi,x\rangle} dx_1 dx_2 = \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \frac{\partial^2 u}{\partial^2 x_1} e^{-2i\pi\xi_1 x_1} dx_1 \right) e^{-2i\pi\xi_2 x_2} dx_2,$$

and then integrating by parts twice, one obtains the expected result.
Finally, the Fourier transform of the heat equation reads:

$$\frac{\partial F(t,\xi)}{\partial t} = -4\pi^2(\xi_1^2 + \xi_2^2) F(t,\xi), \tag{2.8}$$

so that we can finally write $F(t,\xi) = C(\xi)e^{-4\pi^2(\xi_1^2+\xi_2^2)t}$. Denoting $\widehat{\tilde{u}}$ the Fourier transform of $\tilde{u}$, we get that: $F(\xi,t) = \widehat{\tilde{u}}(\xi)e^{-4\pi^2(\xi_1^2+\xi_2^2)t}$. Using classical results on Fourier transforms of Gaussian functions, we deduce:

$$F(t,\xi) = \widehat{\tilde{u}}(\xi) \widehat{\frac{1}{4\pi t} e^{-\frac{x_1^2+x_2^2}{4t}}}(\xi).$$

Considering the inverse Fourier transform of the above expression, we end up writing:

$$u(t,x) = \tilde{u} * G_t(x)$$

where $G_t(x) = \frac{1}{4\pi t} e^{-\frac{x_1^2+x_2^2}{4t}}$.
In the following section, we are going to show that to smooth an image using a radial based kernel is asymptotically equivalent to using a Gaussian kernel. We first consider the averaging operator, and then will switch on to a more general case.
Matlab implementation of the heat equation:
We discretize the derivative in time as follows:

$$\frac{\partial u}{\partial t}(n\delta t, i, j) \approx \frac{u_{i,j}^{n+1} - u_{i,j}^n}{\delta t}$$

and the discretization of the Laplacian is given by:

$$\Delta u_{i,j}^n = u_{i+1,j}^n + u_{i,j+1}^n + u_{i-1,j}^n + u_{i,j-1}^n - 4u_{i+1,j}^n$$

Assuming Neumann conditions at the boundaries one can write the following Matlab program:

```
close all;
load gatlin2;
imagesc(X);
colormap(gray);
N = size(X);
Xextend = zeros(N(1)+2,N(2)+2);
```

% We extend the signal using miror extension

```
Xextend(2:N(1)+1,2:N(2)+1)  = X;
Xextend(2:N(1)+1,1) =  X(1:N,1);
Xextend(2:N(1)+1,N(2)+2) =  X(1:N(1),N(2));
Xextend(1,2:N(2)+1) =  X(1,1:N(2));
Xextend(N(1)+2,2:N(2)+1) =  X(N(1),1:N(2));
Xextend (1,1) = X(1,1);
Xextend (N(1)+2,1) = X(N(1),1);
Xextend (1,N(2)+2) = X(1,N(2));
Xextend (N(1)+2,N(2)+2) = X(1,1);
```

% we compute the iteration associated with the heat equation a certain number of times

```
delta_t = 0.1;
B = [0 delta_t 0; delta_t 1-4*delta_t delta_t; 0 delta_t 0];
n_iter = 5;
for k  = 1:n_iter
   C = conv2(Xextend,B,'same');
   X0 = C(2:N(1)+1,2:N(2)+1);
   Xextend(2:N(1)+1,2:N(2)+1) = X0;
   Xextend(2:N(1)+1,1)            = X0(1:N(1),1);
   Xextend(2:N(1)+1,N(2)+2)    = X0(1:N(1),N(2));
   Xextend(1,2:N(2)+1)            = X0(1,1:N(2));
   Xextend(N(1)+2,2:N(2)+1)   = X0(N(1),1:N(2));
   Xextend(1,1)                       = X0(1,1);
   Xextend(N(1)+2,1)               = X0(N(1),1);
   Xextend(1,N(2)+2)                = X0(1,N(2));
   Xextend(N(1)+2,N(2)+2)     = X0(1,1);
end
figure
imagesc(X0);
colormap(gray);
```

2.6.2 Property of the averaging operator

One defines the averaging operator on the disk centered at x and with radius h by:

$$m_h u_0(x) = \frac{1}{\pi h^2} \int_{D(x,h)} u_0(y) dy.$$

We are now going to show that the averaging operator satisfies the following property:

$$\frac{m_h u_0(x) - u_0(x)}{h^2} = \frac{1}{8} \Delta u_0(x) + \epsilon(h).$$

Looking at the above equation, we clearly see that it behaves like the heat equation. Indeed $\frac{\partial_t u(x,t)}{\partial t} \approx \frac{u(t+\Delta t)-u(t)}{\Delta t}$, which is similar to $\frac{m_h u_0(x)-u_0(x)}{h^2}$ considering the discretization parameter h^2.
PROOF: Without any loss of generality, let us consider that $x = 0$. A second order Taylor-Young expansion at 0 reads:

$$u_0(y) = u_0(0) + Du_0(0).y + \frac{1}{2}((u_0)_{xx}(0)y_1^2 + (u_0)_{yy}(0)y_2^2 + 2(u_0)_{xy}(0)y_1y_2) + o(\|y\|^2).$$

The average of u_0 computed on the disk reads:

$$(m_h u_0)(0) = u_0(0) + \frac{1}{2\pi h^2}((u_0)_{xx}(0)\int_{D(0,h)} y_1^2 dy_1 dy_2 + (u_0)_{yy}(0)\int_{D(0,h)} y_2^2 dy_1 dy_2) + o(h^2).$$

Then, we obtain the final result remarking that:

$$\frac{1}{2\pi h^2}\int_{D(0,h)} x_1^2 dx_1 dx_2 = \frac{1}{4\pi h^2}\int_{D(0,h)} (x_1^2 + x_2^2) dx_1 dx_2 = \frac{1}{4\pi h^2}\int_0^h 2\pi r^3 dr = \frac{h^2}{8}.\blacksquare$$

We are now going to generalize this property to more general kernels,

2.6.3 Convolution with radial based kernels

We define radial based kernels as $g(x) = g(\|x\|)$. In what follows, we assume that these kernels are normalized as:

$$\int_{\mathbb{R}^2} g(x)dx = 1 \qquad \int_{\mathbb{R}^2} x_1^2 g(x)dx = \int_{\mathbb{R}^2} x_2^2 g(x)dx = 2.$$

We would like to determine how these properties can still be satisfied when g undergoes a scale change, that is we consider $ag(\frac{x}{b})$ and seek the relations between a and b. For the first equality to be satisfied, one needs to have $ab^2 = 1$ that is to say $b = \frac{1}{\sqrt{a}}$. With such a b, the other equalities are satisfied (use a change of variables). Then, it is easy to remark that:

$$\int_{\mathbb{R}^2} x_1 g(x)dx = \int_{\mathbb{R}^2} x_2 g(x)dx = \int_{\mathbb{R}^2} x_1 x_2 g(x)dx = 0 \tag{2.9}$$

Indeed, since function g is radial based, one has:

$$\int (x_1 - x_2)^2 g(x)dx = \int (x_1 + x_2)^2 g(x)dx,$$

which enables us to deduce the last equality of (2.9). As far as the first two equalities are concerned, it suffices to remark that:

$$\begin{aligned} \int_{\mathbb{R}^2} x_1 g(x)dx &= \int_0^{+\infty} x_1 \int_{-\infty}^{\infty} g(x)dx_2 dx_1 + \int_{-\infty}^0 x_1 \int_{-\infty}^{\infty} g(x)dx_2 dx_1 \\ &= \int_0^{+\infty} x_1 \int_{-\infty}^{\infty} g(x)dx_2 dx_1 - \int_0^{+\infty} x_1 \int_{-\infty}^{\infty} g(x)dx_2 dx_1 = 0. \end{aligned}$$

We now consider a rescaled version of g as follows: $g_h(x) = \frac{1}{h}g(\frac{x}{h^{\frac{1}{2}}})$, and then denote by $g^{n*} = g * g * \cdots * g$ the function obtained by convolving g with itself n times. Similarly, one defines g_h^{n*}, of which we study the behaviour when $n \to +\infty$ and $h \to 0$. To do so, we first study the properties of the convolution by g_h, for which we have the following theorem:

Théorème 1 *Let $g(x) \in L^1(\mathbb{R}^2)$ a radial based function, normalized as explained above, if one further assumes that:*

$$\int_{\mathbb{R}^2} |g(z)||z|^3 dz = C < +\infty,$$

then for any function $u \in L^\infty(K) \bigcap C^3(K)$, K being a compact set of $\mathbb{R}^2$, one has:

$$(g_h * u)(x) - u(x) = h\Delta u(x) + O(h^{\frac{3}{2}})$$

PROOF: One can write using the fact that u is C^3:

$$\begin{aligned}(g_h * u)(x) - u(x) &= \int_{\mathbb{R}^2} h^{-1} g(\frac{y}{h^{\frac{1}{2}}})(u(x-y) - u(x))dy \\ &= \int_{\mathbb{R}^2} g(z)(u(x - h^{\frac{1}{2}} z) - u(x))dz\end{aligned}$$

$$= \int_{\mathbb{R}^2} g(z)(-h^{\frac{1}{2}} Du(x).z + \frac{h}{2} D^2u(x)(z,z))dz - \frac{1}{6} h^{\frac{3}{2}} \int_{\mathbb{R}^2} g(z) D^3 u(x - h^{\frac{1}{2}}\theta z)(z,z,z)dz,$$

where $\theta = \theta(x, z, h)$ belongs to $[0, 1]$. This last expression being obtained by using Taylor-Lagrange formula on the interval $[x - h^{\frac{1}{2}} z, x]$. Using the information on the moments of the function g given by equation (2.9) and its definition, we are able to write:

$$|(g_h * u)(x) - u(x) - h\Delta u(x)| \leq C h^{\frac{3}{2}} \max_{x \in K} \|D^3 u(x)\|.$$

2.7 Image restoration using PDEs: nonlinear diffusion equations

At each point (x, y) where $\nabla u(x, y) \neq 0$, we consider the local coordinates

$$\eta = \frac{\nabla u}{\|\nabla u\|} \text{ and } \xi = \frac{\nabla u^\perp}{\|\nabla u^\perp\|}.$$

Let us denote Φ the angle made by the gradient and the axis Ox. With this notation we have: $\eta = \begin{pmatrix} \cos(\Phi) \\ \sin(\Phi) \end{pmatrix}$. Then we write

$$\frac{\partial u}{\partial \eta} = \nabla u.\eta = \frac{\partial u}{\partial x}\cos(\Phi) + \frac{\partial u}{\partial y}\sin(\Phi).$$

Differentiating a second time we get:

$$u_{\eta\eta} = \frac{\partial^2 u}{\partial x^2}\cos^2(\Phi) + \frac{\partial^2 u}{\partial x \partial y}\cos(\Phi)\sin(\Phi) + \frac{\partial^2 u}{\partial^2 y}\sin^2(\Phi) = \eta^t D^2 u \eta = D^2 u(\frac{\nabla u}{\|\nabla u\|}, \frac{\nabla u}{\|\nabla u\|}).$$

Similarly, one has $D^2 u(\frac{\nabla u^\perp}{\|\nabla u\|}, \frac{\nabla u^\perp}{\|\nabla u\|}) = u_{\xi\xi}$.

2.7.1 Perona-Malik diffusion equation

One of the most annoying problem when using the heat equation for image restoration is that the smoothing process damages the edges. An approach proposed by Perona and Malik [10] aims at smoothing homogeneous regions while preserving the edges. It is essentially based on the use of the following equation:

$$\frac{\partial u}{\partial t} = div(g(\|\nabla u\|)\nabla u),$$

where $g(s) = \frac{1}{1+(\lambda s)^2}$. One easily shows that if $\|\nabla u\| \leq \frac{1}{\lambda}$, diffusion occurs while anti-diffusion in the opposite case. To prove this, we rewrite the Perona-Malik equation in terms of the local coordinates (η, ξ). We thus have:

$$
\begin{aligned}
div(g(\|\nabla u\|)\nabla u) &= \frac{\partial}{\partial x}(g(\|\nabla u\|)u_x) + \frac{\partial}{\partial y}(g(\|\nabla u\|)u_y) \\
&= Dg(|\nabla u|)\nabla u + g(\|\nabla u\|)\Delta u \\
&= g'(\|\nabla u\|)u_{\eta\eta}\|\nabla u\| + g(\|\nabla u\|)(u_{\eta\eta} + u_{\xi\xi}) \\
&= \frac{u_{\xi\xi}}{1+(\lambda\|\nabla u\|)^2} + \frac{(1-\lambda^2(\|\nabla u\|^2))u_{\eta\eta}}{(1+\lambda^2\|\nabla u\|^2)^2}.
\end{aligned}
$$

The first term appears to be a diffusion term in the direction orthogonal to the gradient while the second one is a smoothing term in the direction of the gradient. Applying this nonlinear diffusion equation, homogeneous regions should be smoothed while sharp edges should be preserved. However, since the diffusion operator depends on the gradient of u, the latter does not perform well in a noisy context. To improve the performance of the proposed operator in a noisy context, Catté, Lions and Morel have introduced the following model:

$$
\frac{\partial u}{\partial t} = div(g(\|\nabla(u * G_\sigma)\|)\nabla u), \tag{2.10}
$$

where G_σ is a Gaussian kernel. By making the diffusion depend on the gradient of the regularized image, the diffusion is still important on noisy homogeneous regions.
A example of Matlab implementation of Perona-Malick model is as follows
% Convert input image to double.

```
close all
load gatlin2;
```

% the image is loaded in the variable X % visualization %white noise generation

```
sigma = 1;
colormap(gray);
[M,N] = size(X);
J = sigma*randn(M,N);
im = X + J;
imagesc(im);
pause
im = double(im);
```

% PDE (partial differential equation) initial condition.

```
diff_im = im;
```

% Center pixel distances

```
dx = 1;
dy = 1;
```

% 2D convolution masks - finite differences.

```
hN = [0 1 0; 0 0 0; 0 -1 0]/(2*dx);
hE = [0 -1 0; 0 0 0; 0 1 0]/(2*dy);
```

```
num_iter = 10;
kappa = 2;
for t = 1:num_iter
        nablaN = conv2(diff_im,hN,'same');
        nablaE = conv2(diff_im,hE,'same');
        % Diffusion function.
        norm_nabla = sqrt(nablaN.^2+nablaE.^2);
        %mean(mean(norm_nabla))
        c = 1./(1 + (norm_nabla/kappa).^2);
        A =conv2(c.*nablaN,hN,'same')+conv2(c.*nablaE,hE,'same');
        % Discrete PDE solution.
        diff_im = diff_im + delta_t*A;
        % Iteration warning.
        fprintf('\rIteration %d\n',t);
end
figure
imagesc(diff_im);
colormap(gray)
```

2.7.2 Approaches using mean curvature motion

The main motivation to design filters based on the mean curvature motion (MCM) is that they correspond to a nonlinear diffusion operator that is able to smooth the image in the direction of an edge but not in the direction orthogonal to it. Let us first recall that the heat equation can be written in the local coordinates as:

$$\frac{\partial u}{\partial t} = u_{\xi\xi} + u_{\eta\eta}.$$

If we only keep the diffusion in the direction orthogonal to the gradient, we get:

$$\frac{\partial u}{\partial t} = u_{\xi\xi} = \Delta u - \frac{D^2u(\nabla u, \nabla u)}{\|\nabla u\|^2} = \|\nabla u\| div(\frac{\nabla u}{\|\nabla u\|}). \tag{2.11}$$

This last equality is obtained by means of the following computation:

$$\|\nabla u\| div(\frac{\nabla u}{\|\nabla u\|}) = \|\nabla u\| \left(\frac{\partial}{\partial x}(\frac{u_x}{\|\nabla u\|}) + \frac{\partial}{\partial y}(\frac{u_y}{\|\nabla u\|}) \right) = \Delta u -^t \nabla(\frac{1}{\|\nabla u\|})\nabla u = \Delta u - \frac{D^2u(\nabla u, \nabla u)}{\|\nabla u\|^2}.$$

We are now going to establish a relation between equation (2.11) and level set motion. Let us consider that $u(x, y, t)$ is the solution to (2.11) at time t, and then let us define the level set with level r at time t as:

$$F_{r,t} = \{(x, y),\ u(x, y, t) = r\}.$$

This curve can be parametrized by:

$$F_{r,t} = \{M(s, t),\ u(M(s, t), t) = r\}. \tag{2.12}$$

Differentiating (2.12) once with respect to s (we omit t for ease of notation), one gets:

$$\langle \frac{\partial M(s)}{\partial s}, \nabla u(M(s)) \rangle = 0. \tag{2.13}$$

Differentiating (2.12) twice with respect to s, one gets:

$$\langle \frac{\partial^2 M(s)}{\partial^2 s}, \nabla u(M(s))\rangle + \frac{\partial M(s)}{\partial s}^T D^2(u(M(s)))\frac{\partial M(s)}{\partial s} = 0. \tag{2.14}$$

We then remark that:

$$\frac{\partial^2 M}{\partial^2 s} = -K\eta, \tag{2.15}$$

where K is the curvature. We can thus write $K = -\frac{\langle \frac{\partial^2 M(s)}{\partial^2 s}, \nabla u(M(s))\rangle}{\|\nabla u\|}$, and finally since $\frac{\partial M(s)}{\partial s} = \frac{\nabla u^{\perp}}{\|\nabla u\|}$, we get from (2.14).

$$u_{\xi\xi} = D^2(u)(\frac{\nabla u^{\perp}}{\|\nabla u\|}, \frac{\nabla u^{\perp}}{\|\nabla u\|}) = D^2 u(M(s))(\frac{\partial M(s)}{\partial s}, \frac{\partial M(s)}{\partial s}) = \|\nabla u\| div(\frac{\nabla u}{\|\nabla u\|}) = K\|\nabla u\|,$$

hence the result. Based on this equation, a new model was proposed by Alvarez, Lions and Morel [1]:

$$\left\{ \begin{array}{rcl} \frac{\partial u}{\partial t} & = & g(\|\nabla(u * G_\sigma)\|)\|\nabla u\| div\left(\frac{\nabla u}{\|\nabla u\|}\right), \\ u(x,y,0) & = & u_0(x,y). \end{array} \right. \tag{2.16}$$

This model is based on a diffusion in the direction of the level sets except where the regularized gradient has a strong amplitude.

2.8 Image restoration as minimization process

2.8.1 Tychonov Minimization

The above mentioned nonlinear diffusion equations have uninteresting steady states, which means that if these equations are used for image restoration, they need to be stoped at a certain time, imposed by the user. Alternatively, image restoration can be viewed as an optimization problem. If u_0 is the original image, we would like the restored image $\hat{u}$ to be close to u_0 supposed to belong to $L^2(\Omega)$, and also such that the gradient of $\hat{u}$ is as small as possible on average. With this in mind, one can seek u can be viewed as the solution of the following optimization problem:

$$u = \underset{v \in H_0^1(\Omega)}{\text{argmin}} \left\{ \lambda \int_\Omega |v - u_0|^2 + \int_\Omega \|\nabla v\|_2^2 \right\}. \tag{2.17}$$

with

$$H_0^1(\Omega) = \left\{ u \in L^2(\Omega), \text{admitting first order weak derivatives} \in L^2(\Omega), \text{ and such that } u = 0 \text{ on } \partial\Omega \right\}.$$

The term "weak derivative" means u is not differentiable and that there exists unique functions $v_i \in L^2(\Omega)$, $i = 1, 2$ such that

$$\forall \varphi \in \mathcal{D}(\Omega), \int_\Omega u \frac{\partial \varphi}{\partial x_i} = -\int_\Omega v_i \varphi.$$

By an abuse of notation, we note $\nabla u = (v_1, v_2)^T$. The first question we need to answer is whether $\hat{u}$ exists and is unique. Let us put $J(u) = \lambda \int_\Omega |u - u_0|^2 + \int_\Omega \|\nabla u\|_2^2$. The minimum of this functional if it exists satisfies:

$$DJ(u).h = 0,\ \forall h \in H_0^1(\Omega)$$

Let us compute the differential of J, which has to be null at $\hat{u}$ whatever $h \in H_0^1(\Omega)$.

$$DJ(u).h = 0,\ \forall h \in H_0^1(\Omega) \Leftrightarrow \int_\Omega 2\lambda(u - u_0)h + 2\langle \nabla u, \nabla h\rangle = 0, \quad \forall\ h \in H_0^1(\Omega)$$

Note that $H_0^1(\Omega)$ is an equipped with the norm

$$\|u\|_{H_0^1(\Omega)} = \left(\|u\|_2^2 + \sum_{i=1}^{2} \|\frac{\partial u}{\partial x_i}\|_2^2\right)^{\frac{1}{2}},$$

which is associated with the inner product:

$$\langle u, v\rangle_{H_0^1(\Omega)} = \int_\Omega uv + \sum_{i=1}^{2} \int_\Omega \frac{\partial u}{\partial x_i}\frac{\partial v}{\partial x_i}.$$

Introducing:

$$H^1(\Omega) = \left\{u \in L^2(\Omega), \text{admitting first order weak derivatives} \in L^2(\Omega)\right\}.$$

is an Hilbert space (for the just mentioned norm) and $H_0^1(\Omega)$ can be be viewed as the closure of $C_0^1(\Omega)$ in this space, meaning $H_0^1(\Omega)$ is also an Hilbert space. We may then write:

$$\int_\Omega \langle \nabla u, \nabla h\rangle + \lambda uh = \int_\Omega \lambda u_0 h, \quad \forall\ h \in H_0^1(\Omega).$$

Considering the bilinear form $a(u, v) = \int_\Omega \langle \nabla u, \nabla h\rangle + \lambda uh$, which is continuous, i.e $|a(u, v)| \le C\|u\|\|v\|$, and coercive meaning $a(v, v) \ge \alpha\|v\|^2$, for some $\alpha > 0$. From classical functional analysis theory (Lax-Milgram theorem), it is known that there exists a unique $u \in H_0^1(\Omega)$, satisfying the above equation and that it corresponds to a minimum. Now if Ω is a C^2 open set, with $\partial\Omega$ bounded, one can show that the solution $u \in H^2$, and that one has:

Theorem 1

$$\boxed{\text{If } \Omega \text{ is } C^2 \text{ and } u_0 \in L^2(\Omega) \text{ then } u \in H^2(\Omega) \text{ and } \|u\|_{H^2(\Omega)} \le C\|u_0\|_{L^2(\Omega)}}$$

Note that if $u \in C^2(\Omega)$, using the Green-Ostrogradsky theorem we may write:

$$\int_\Omega \langle \nabla u, \nabla h\rangle = \int_{\partial\Omega} \langle \nabla u, n\rangle h - \int_\Omega \Delta uh = -\int_\Omega \Delta uh \quad \forall\ h \in C_0^1(\Omega),$$

where n is a vector normal to $\partial\Omega$ oriented outward. So u satisfies:

$$\int_\Omega (-\Delta u + \lambda(u - u_0))h = 0, \quad \forall\ h \in C_0^1(\Omega)$$

which means $-\Delta u+\lambda u=\lambda u_0$ almost everywhere and, as thus as u is C^2, this equation (called Euler equation) is satisfied everywhere. Therefore one can relate the Tychonov minimization problem to the following partial differential equation:

$$\begin{cases} \frac{\partial u}{\partial t} = -\lambda(u-u_0)+\Delta u \quad \text{in } \Omega \\ u=0 \text{ on } \partial\Omega \end{cases}$$

This equation can be viewed as a modified heat equation where the term $\lambda(u-u_0)$ prevents u to be too far from u_0. If one denotes u_∞ the solution u when t tends to $+\infty$, it has the following behaviour:
<u>If λ is large</u>, u_∞ is close to u_0.
<u>If λ is small</u>, the solution is close to that of the heat equation and the edges can be strongly damaged.

Ex: propose a Matlab code to solve this, investigate different values for λ.

The conditions that u be null at image boundaries are somewhat arbitrary, and we are going to show that Neumann conditions could be used instead. Let us reformulate the initial minimization problem as:

$$u \quad = \quad \underset{v\in H^1(\Omega)}{\operatorname{argmin}} \left\{ \lambda \int_\Omega |v-u_0|^2 + \int_\Omega \|\nabla v\|_2^2 \right\}. \tag{2.18}$$

The extrema of the function J should satisfy:

$$\begin{aligned} DJ(u).h=0,\ \forall h\in H^1(\Omega) &\Leftrightarrow \int_\Omega 2\lambda(u-u_0)h+2\langle\nabla u,\nabla h\rangle=0, \quad \forall\ h\in H^1(\Omega) \\ &\Leftrightarrow \int_\Omega \langle\nabla u,\nabla h\rangle+\lambda uh=\int_\Omega \lambda u_0 h, \quad \forall\ h\in H^1(\Omega), \end{aligned}$$

which admits a unique solution due to Lax-Milgram theorem. The theorem related to the regularity of u is similar to that associated with $u\in H_0^1(\Omega)$, that it is H^2 when Ω is C^2.

Assume that the solution is C^2, such that $\frac{\partial u}{\partial n}=0$ on $\partial\Omega$. From the Green-Ostrogradsky formula we may write:

$$\int_\Omega \langle\nabla u,\nabla h\rangle = \int_{\partial\Omega}\langle\nabla u,n\rangle h - \int_\Omega \Delta uh = -\int_\Omega \Delta uh \quad \forall\ h\in C^1(\Omega),\ \frac{\partial u}{\partial n}=0, \text{ on } \partial\Omega$$

Then, we may write

$$\begin{aligned} \int_\Omega \langle\nabla u,\nabla h\rangle+\lambda uh=\int_\Omega \lambda u_0 h, \quad \forall\ h\in C^1(\Omega),\ \frac{\partial u}{\partial n}=0, \text{ on } \partial\Omega \\ \int_\Omega (-\Delta u+\lambda(u-u_0))h=0, \quad \forall\ h\in C^1(\Omega),\ \frac{\partial u}{\partial n}=0, \text{ on } \partial\Omega \end{aligned}$$

From this we deduce that on Ω we have $-\Delta u+\lambda(u-u_0)=0$ (since the equality is also true for $C_0^1(\Omega)$). The associated evolution equation in that case is

$$\begin{cases} \frac{\partial u}{\partial t} = -\lambda(u-u_0)+\Delta u \quad \text{in } \Omega \\ \frac{\partial u}{\partial n}=0 \text{ on } \partial\Omega \end{cases}$$

The minimization problem can be rewritten, in a more general setting using more general Dirichlet conditions, as:

$$u = \underset{v\in K}{\operatorname{argmin}} \left\{ \lambda \int_\Omega |v-u_0|^2 + \int_\Omega \|\nabla v\|_2^2 \right\}. \tag{2.19}$$

where $K = \{u \in H^1(\Omega),\ u = g \text{ on } \partial\Omega\}$. Note that K is a convex set (it is even more than that, it is an affine space). In such a case, a critical point will satisfy

$$DJ(u)(h-u) \geq 0,\ \forall h \in K \Leftrightarrow DJ(u).h = 0\ \forall h \in H_0^1(\Omega)$$

So we end up with the following problem:

$$\int_\Omega \langle \nabla u, \nabla h\rangle + \lambda u h = \int_\Omega \lambda u_0 h, \quad \forall\ h \in H_0^1(\Omega)$$

which is known to admit a unique solution in K.
Instead of considering an evolution equation, one can seek to solve directly the Euler equation when $\Omega = [0,L]\times[0,1]$. So we seek u such that:

$$\begin{cases} -\frac{1}{\lambda}\Delta u + u = u_0 \text{ on } \Omega \\ u = 0 \text{ sur } \partial\Omega \end{cases}$$

We then recall that $(\frac{e^{i2\pi\frac{xn}{L}}}{\sqrt{L}} e^{i2\pi yk})_{n,k\in\mathbb{Z}}$ is an orthonormal basis of $L^2(\Omega)$, and proceed as follows:

- Extend u on $\tilde{\Omega} = [-L,L]\times[-1,1]$ into $\tilde{u}$, satisfying $\tilde{u}(-x,y) = -u(x,y)$, $\tilde{u}(x,-y) = -u(x,y)$, et $\tilde{u}(-x,-y) = u(x,y)$ for (x,y) in Ω. Remarking that $\tilde{u}$ is odd with respect to each of its variable, show that on $\tilde{\Omega}$, $\tilde{u}$ can be decomposed in the basis $(\sin(\pi\frac{xn}{L})\sin(\pi yk))_{n,k\in\mathbb{Z}^*}$.
- Do the same thing on u_0, to obtain $\tilde{u}_0$, which also decomposes the same way on the basis made of sine functions just introduced.
- Compute the Fourier coefficients of $\tilde{u}$ with respect to those of $\tilde{u}_0$, and then deduce u.

2.8.2 Mumford-Shah model and variants

An alternative approach to Tychonov minimization was proposed by Mumford and Shah in [8]. Their approach consists of minimizing the following functionnal:

$$E_{u_0}(u,K) = \beta \int_\Omega (u-u_0)^2 + \int_{\Omega\setminus K} \|\nabla u\|^2 + \alpha \operatorname{mes}(K).$$

This model admits many different variants. Among these, one could seek to write explicitly the edge function in the minimization, which leads to the so-called Nordström model [9]:

$$E_{u_0}(u,w) = \int_\Omega \beta(u-u_0)^2 + w\|\nabla u\|^2 + \lambda^2(w - \log(w))dx,$$

where $w : \Omega \to \mathbb{R}_*^+$ with $w \approx 1$ inside homogeneous regions and should be null at edge location, λ and β being two positive constants. Assuming that (u,v) lives in $H^1(\Omega)\times K$, with K the convex set of positive functions on Ω, the critical points will this time satisfy:

$$\begin{aligned} & DE_{u_0}(u,w).(v-u,k-w) \geq 0\ \forall (v,k) \in H^1(\Omega)\times K \\ \Leftrightarrow\ & \frac{\partial E_{u_0}}{\partial u}(u,w).(v-u) + \frac{\partial E_{u_0}}{\partial w}(u,w).(k-w) \geq 0\ \forall (v,k) \in H^1(\Omega)\times K. \end{aligned}$$

One of the local minimum satisfies:

$$\begin{aligned}\int_\Omega \beta(u-u_0)v - \langle w\nabla u, \nabla v\rangle &= 0,\ \forall v \in H^1(\Omega)\\ \int_\Omega (\lambda^2(1-\frac{1}{w}) + \|\nabla u\|^2)k &= 0,\ \forall k \in K\end{aligned}$$

Thus we get the Euler Lagrange equations:

$$\begin{cases}\beta(u-u_0) - div(w\nabla u) &= 0\\ \lambda^2(1-\frac{1}{w}) + \|\nabla u\|^2 &= 0.\end{cases}$$

The second equation enables us to write w under the same form as the function g using in the Perona-Malik equation:

$$w = \frac{1}{1+\frac{\|\nabla u\|^2}{\lambda^2}}.$$

Finally, we remark that the above minimization can be related to the following reaction-diffusion equation:

$$\begin{cases}\frac{\partial u}{\partial t} &= div(g(\|\nabla u\|)\nabla u) + \beta(u_0-u)\\ u(x,0) &= u_0(x).\end{cases}$$

This equation combines a diffusion term similar to that used in the Perona-Malik equation and another term that prevents the pre-processed image from being far from the original image.
Note that in that formalism the parameter β has to be tuned, and also that the minimization process can be viewed as minimizing

$$F_{u_0}(u) = \int_\Omega \beta(u-u_0)^2 + \lambda^2 \log(1+\frac{\|\nabla u\|}{\lambda^2})dx,$$

which is unfortunately non convex.

2.8.3 More general setting

We here study the restoration of the image u, defined on a domain Ω, and from u_0 by :

$$u_0 = k * u + n \tag{2.20}$$

where $*$ stands for the convolution product. For that purpose, one seeks u minimizing the following functional:

$$E(u) = \int_\Omega \Phi(\|\nabla u\|) + \lambda \int_\Omega (u_0 - k * u)^2 \tag{2.21}$$

1. Assuming Neumann on $\partial\Omega$, compute the differential of E.

2. Deduce the associated Euler equation and then that the minimisation can be solved by studying an certain evolution equation (we will denote $\bar{k}(x) = k(-x)$).

3. Write this equation in the local coordinates (η, ξ).

2.9 Formulation of the Rudin-Osher-Fatemi Model

If one assumes that an image u is a cartoon image, a natural approach is to consider that u is in BV, and $u_0 - u$, i.e. the noise is in L^2. BV functions correspond to:

$$BV = \{u \in L^1(\Omega), \int_\Omega \|\nabla u\| < +\infty\},$$

where the gradient has to be understood in the distributional sense. Then minimizing

$$u = \operatorname*{argmin}_{v \in BV} \left\{ \lambda \int_\Omega |v - u_0|^2 + \int_\Omega \|\nabla v\| \right\}$$

with the supplementary condition that $\int_\Omega u = \int_\Omega u_0$. The critical points are such that, the differential vanishes which admits the following differential, at point where $\|\nabla u\| \neq 0$:

$$DJ(u).h = \int_\Omega (\lambda(u - u_0) - div(\frac{\nabla u}{\|\nabla u\|}))h = 0, \ \forall h \in BV.$$

Using the Green-Ostrogradski theorem (assuming the gradient does not vanish) and imposing Neumann conditions on the boundaries (which amounts to drop the constraints on the constant mean), we obtain:

$$\begin{cases} \frac{\partial u}{\partial t} = div(\frac{\nabla u}{\|\nabla u\|}) - \lambda(u - u_0) \\ \frac{\partial u}{\partial n} = 0 \\ u(x,0) = u_0(x) \end{cases}$$

Then a solution is proposed by discretizing the above equation (an automatic determination of the parameter λ is proposed).
One of the limitation of the above approach is that the functional J is not differentiable everywhere, and an alternative technique was proposed by Chambolle in [2]. The minimization problem can be rewritten as:

$$\inf_v (\frac{1}{2\lambda} \|v - u_0\|_2^2 + F(v))$$

with $F(v) = \int_\Omega \|\nabla v\|$. Generally speaking, when v is not differentiable, $F(v)$ is defined as [2]:

$$F(v) = \sup\{ \int v(x) div(\xi(x)) dx, \xi \in C_c^1(\Omega, \mathbb{R}^2), \|\xi(x)\| \leq 1, \ \forall x \in \Omega\}$$

Note that if the gradient of v exists, then:

$$F(v) = \sup\{- \int_\Omega \langle \nabla v(x), \xi(x) \rangle dx, \xi \in C_c^1(\Omega, \mathbb{R}^2), \|\xi(x)\| \leq 1, \ \forall x \in \Omega\} = \int_\Omega \|\nabla v(x)\|.$$

The functional F is not differentiable (even though in the initial formulation by Osher, Fatemi, the formal derivation assumed it was the case), so to find the critical points, we use the concept of sub-differential

Definition 12 *(sub-gradient and sub-differential) Let $f : E \to \mathbb{R}$ a convex function. A vector $\eta \in E$ is called sub-gradient of f at x_0 if*

$$\forall x \in dom(f), f(x) \geq f(x_0) + \langle \eta, x - x_0 \rangle.$$

The set of the sub-gradients is called sub-differential of f at x_0 and is denoted by $\partial f(x_0)$.

Then, one needs to introduce the Legendre-Fenchel transform of F as

$$F^*(u) = \sup_{v \in L^2(\Omega)} \langle v, u \rangle - F(v).$$

One remarks that:

$$F^*(u) = \chi_K(u) = \begin{cases} 0 & \text{if } u \in K \\ +\infty & \text{otherwise,} \end{cases}$$

with $K = \{div(\xi) : \xi \in C_c^1(\Omega, \mathbb{R}^2), \|\xi(x)\| \leq 1, \ \forall x \in \Omega\}$. Indeed,

$$F^*(u) = \sup_{v \in L^2(\Omega)} \inf_{\xi \in C_c^1(\Omega,\mathbb{R}^2), \|\xi(x)\| \leq 1} \int_\Omega v(u - div(\xi)) = \inf_{\xi \in C_c^1(\Omega,\mathbb{R}^2), \|\xi(x)\| \leq 1} \sup_{v \in L^2(\Omega)} \int_\Omega v(u - div(\xi))$$

If $u - div(\xi) \neq 0$ for all ξ in the set, we take $v = \alpha(u - div(\xi))$ and notice that $F^*(u)$ is equal to $+\infty$ by making α tend to $+\infty$. On the contrary, if there exists ξ such that $u = div(\xi)$, then $F^*(u)$ is null.

Note that with that formalism, since Ω is finite, it can be proven that F is continuous and then $F^{**} = F$. With that formalism, the Euler equation associated with the minimization problem reads:

$$0 \in v - u_0 + \lambda \partial F(v) \Leftrightarrow \frac{u_0 - v}{\lambda} \in \partial F(v) \Leftrightarrow v \in \partial F^*(\frac{u_0 - v}{\lambda}) \tag{2.22}$$

To show that the Euler equation has the above form, let us denote $J(v) = \frac{1}{2\lambda}\|v - u_0\|_2^2 + F(v)$. Since J is convex we have :

$$u^* \in \partial J(u) \Leftrightarrow \forall v \in dom(J), \ J(v) \geq J(u) + \langle u^*, v - u \rangle$$

and thus

$$0 \in \partial J(u) \Leftrightarrow \forall v \in dom(J), \ J(v) \geq J(u)$$

Now $\partial J(u) = \frac{1}{\lambda}(u - u_0) + \partial F(u)$, since the two functions are convex and since the intersection of the domains of definition of $\frac{1}{2\lambda}\|v - u_0\|_2^2$ and of F are non empty.

Now to prove Eq. (2.22), we shall remark that

$$\begin{aligned} r \in \partial F(v) \quad & \Leftrightarrow \quad F(u) \geq F(v) + \langle r, u - v \rangle, \ \forall u \in dom(F) \\ & \Leftrightarrow \quad \langle r, v \rangle - F(v) \geq \langle r, u \rangle - F(u), \ \forall u \in dom(F) \\ & \Leftrightarrow \quad \langle r, v \rangle - F(v) = F^*(r) \\ & \Leftrightarrow \quad v \in \partial F^*(r) \end{aligned}$$

For the last equivalence, assume that $\langle r, v \rangle - F(v) = F^*(r)$, then we have:

$$\begin{aligned} v \in \partial F^*(r) \quad & \Leftrightarrow \quad F^*(u) \geq F^*(r) + \langle v, u - r \rangle \\ & \Leftrightarrow \quad F^*(u) \geq F^*(r) + \langle v, u \rangle - \langle v, r \rangle \\ & \Leftrightarrow \quad F^*(u) \geq \langle v, u \rangle - F(v) \end{aligned}$$

which is true. So we have proven that

$$r \in \partial F(v) \Rightarrow v \in \partial F^*(r)$$

Making the same kind of computation for F^* we get $v \in \partial F^*(r) \Rightarrow r \in \partial F^{**}(v) = \partial F(v)$. Hence the result. We can then rewrite the above equation as:

$$\frac{u_0}{\lambda} \in \frac{u_0 - v}{\lambda} + \frac{1}{\lambda}\partial F^*(\frac{u_0 - v}{\lambda}) \Leftrightarrow 0 \in \frac{u_0 - v}{\lambda} - \frac{u_0}{\lambda} + \frac{1}{\lambda}\partial F^*(\frac{u_0 - v}{\lambda})$$

we thus get that $w = \frac{u_0 - v}{\lambda}$ is the minimizer of

$$\frac{\|w - \frac{u_0}{\lambda}\|^2}{2} + \frac{1}{\lambda}F^*(w)$$

and due to the definition of F^* we deduce that $w = P_K(\frac{u_0}{\lambda})$ (K is a closed convex). Hence the solution of the minimization problem is given by:

$$v = u_0 - P_{\lambda K}(u_0).$$

We may write:

$$\begin{aligned}\frac{u_0 - v}{\lambda} = P_K(\frac{u_0}{\lambda}) \quad &\Leftrightarrow \quad v = u_0 - \lambda P_K(\frac{u_0}{\lambda}) \\ &\Leftrightarrow \quad v = u_0 - P_{\lambda K}(u_0).\end{aligned}$$

Indeed, from the characterization of the projection on a closed convex set, we may write:

$$\begin{gathered}\langle \frac{u_0}{\lambda} - P_K(\frac{u_0}{\lambda}), y - P_K(\frac{u_0}{\lambda})\rangle \leq 0,\ \forall y \in K \\ \langle u_0 - \lambda P_K(\frac{u_0}{\lambda}), \lambda y - \lambda P_K(\frac{u_0}{\lambda})\rangle \leq 0,\ \forall y \in K \\ \langle u_0 - \lambda P_K(\frac{u_0}{\lambda}), y - \lambda P_K(\frac{u_0}{\lambda})\rangle \leq 0,\ \forall y \in \lambda K\end{gathered}$$

From the unicity of the projection on a closed convex set we get: $\lambda P_K(\frac{u_0}{\lambda}) = P_{\lambda K}(u_0)$. In the discrete space setting, to compute the orthogonal projection amounts to determining the element $p \in \mathbb{R}^{N\times N} \times \mathbb{R}^{N\times N}$, minimizing

$$\min\left(\|\lambda div(p) - u_0\|_2^2,\ \text{s.t. } \|p_{i,j}\| \leq 1, \forall i,j = 1, \cdots, N\right)$$

which can be solved by a fixed point technique:

$$\begin{aligned} p^0 &= 0 \\ p^{n+1}_{i,j} &= \frac{p^n_{i,j} + \tau(\nabla(div(p^n) - u_0\lambda))_{i,j}}{1 + \tau|(\nabla(div(p^n)) - \frac{u_0}{\lambda})_{i,j}|}\end{aligned}$$

We then have the following theorem:

Theorem 2

$\boxed{\text{Assume that } \tau \leq 1/8, \text{ then } \lambda div(p^n) \text{ converges to } P_{\lambda K}(u_0) \text{ when } n \to +\infty}$

The solution to the minimization problem is thus given by:

$$u = u_0 - \lambda div(p^\infty).$$

where $p^\infty = \lim\limits_{n\to+\infty} p^n$.

2.10 Wavelet thresholding

An alternative technique to minimize the above mentioned problem is to consider the wavelet formalism. Let $u_0 \in L^2(\Omega)$ and $\{\Psi_{i,j}\}$ be an orthonormal wavelet basis of $L^2(\Omega)$, and write u_0 in that basis as follows:

$$u_0(x) = \sum_{j,k} c_{j,k} \Psi_{j,k}(x).$$

Since u also belongs to $L^2(\Omega)$, it admits a decomposition in the wavelet basis:

$$u(x) = \sum_{j,k} \tilde{c}_{j,k} \Psi_{j,k}(x).$$

Since the basis is orthonormal:

$$\|u - u_0\|^2 = \sum_{j,k} |c_{j,k} - \tilde{c}_{j,k}|^2.$$

Then, one would like to approximate $\int_\Omega \|\nabla u\|$ using wavelet coefficients. For that purpose, one needs to introduce an approximation of BV functions by means of Besov spaces.
Very commonly used spaces in wavelet analysis are Besov spaces $B^s_{p,q}$ which roughly speaking correspond to functions admitting s derivatives in $L^p(\Omega)$, the third parameter enabling to adjust with accuracy the regularity of the functions involved.
One can give an intrinsic definition of Besov spaces $B^s_{p,q}$ and of their norm $\|.\|_{B^s_{p,q}}$, assuming ψ has $s+1$ null moments and with regularity at least C^{s+1}, then if $f \in B^s_{p,q}$, $\|f\|_{B^s_{p,q}}$ is equivalent to:

$$\left(\sum_k \left(\sum_k 2^{skp} 2^{k(p-2)} |c_{j,k}|^p \right)^{\frac{p}{q}} \right)^{\frac{1}{q}} \tag{2.23}$$

The constant defining the equivalence depend on the chosen wavelet. The Besov spaces are defined up to a constant, therefore we introduce the homogeneous Besov space as:

$$\dot{B}^s_{p,q} = B^s_{p,q} \ / \ \{u \in B^s_{p,q}, \ \nabla u = 0\}$$

Definition 13 *$\dot{B}^1_{1,1}$ is the usual homogeneous Besov space. Let $\psi_{j,k}$ be an orthonormal basis of compactly supported regular wavelets. $\dot{B}^1_{1,1}$ is a subspace of functions of $L^2(\mathbb{R}^2)$, and a function f belongs to $\dot{B}^1_{1,1}$ if and only if:*

$$\sum_{j\in\mathbb{Z}} \sum_{k\in\mathbb{Z}^2} |c_{j,k}| < +\infty$$

BV functions are very close to $\dot{B}^1_{1,1}$ functions, so that we may write:

$$u \in BV \approx \sum_{j,k} |\tilde{c}_{j,k}| < \infty.$$

Written in terms of wavelet coefficients, the minimization problem then reads:

$$\operatorname*{argmin}_{\tilde{c}} \left\{ \lambda \sum_{j,k} |\tilde{c}_{j,k} - c_{j,k}|^2 + \sum_{j,k} |\tilde{c}_{j,k}| \right\}.$$

This problem can be solved directly without the use of partial differential equation (the functional is strictly convex and the coefficient of u can be found by simple differentiation). Annihilating the differential corresponds in this case to:

$$2\lambda(\tilde{c}_{j,k} - c_{j,k}) + sign(\tilde{c}_{j,k}) = 0.$$

so that $\tilde{c}_{j,k}$ has the same sign as $c_{j,k}$ if $|c_{j,k}| \geq \frac{1}{2\lambda}$, and then we get:

$$\begin{array}{rcl}
\tilde{c}_{j,k} & = & c_{j,k} - \frac{1}{2\lambda} \text{ if } c_{j,k} > \frac{1}{2\lambda} \\
\tilde{c}_{j,k} & = & c_{j,k} + \frac{1}{2\lambda} \text{ if } c_{j,k} < -\frac{1}{2\lambda} \\
\tilde{c}_{j,k} & = & 0 \text{ otherwise}
\end{array}$$

So , one translates by a factor of $-\frac{1}{2\lambda}$ the coefficients above $\frac{1}{2\lambda}$ and by a factor $\frac{1}{2\lambda}$ those under $-\frac{1}{2\lambda}$.
This restoration process is called soft-thresholding.
Ex: propose an algorithm to compute the image restored that way.

Chapter 3

Edge Detection

3.1 Edge detection: Marr-Hildreth and Canny approaches

A very intuitive definition of an edge would be to consider it is a closed curve on each side of which the grey level varies sharply, meaning the grey level gradient is strong at edges location. However, this definition of edges cannot be used as is since:

a) high grey level gradients do not necessarily correspond to closed curves.

b) high grey level gradients may also arise in noisy parts of an image and thus may not correspond to an edge.

To get rid of high grey level gradients associated with noise, a very classical technique is to filter the image. This step is carried out by convolution with Gaussian functions.
Then, two main definition are used to characterize an edge, removing the notion of closed curve. The first one called *Marr-Hildreth* approach consists in considering the locations where the Laplacian of the image u vanishes. Another approach, called Canny's approach, consists in seeking the location of the local maxima of the modulus of the gradient of u in the direction of the gradient of u. This amounts to considering a function g defined by:

$$g(t) = \|\nabla u(x + t\nabla u(x))\|,$$

and then computing points x such that $g'(0) = 0$. To compute this derivative, we remark that $g(t) = d(h(c(t)))$ with $c(t) = x + t\nabla u(x)$, $h(x) = \nabla u(x)$ and $d(x) = \|x\|$. Using the chain rule, one writes:

$$\begin{aligned} g'(t) &= d'(h(c(t))).h'(c(t)).c'(t) = \frac{\langle h(c(t)), h'(c(t)).c'(t)\rangle}{\|h(c(t))\|} \\ &= \frac{\langle \nabla u(x + t\nabla u(x)), D^2u(x + t\nabla u(x))\nabla u(x)\rangle}{\|\nabla u(x + t\nabla u(x))\|}. \end{aligned}$$

Finally, we write:

$$g'(0) = 0 \Leftrightarrow D^2u(\frac{\nabla u(x)}{\|\nabla u(x)\|}, \frac{\nabla u(x)}{\|\nabla u(x)\|}) = 0.$$

Now, if we are interested in the implementation of Canny edge detector, a potential implementation would be:
Algo 1 (Canny [3]):

- Convolve u_0 with a Gaussian kernel of increasing size to obtain u (associated with a scale s)
- Find points x such that $\|\nabla u(x)\| \neq 0$ and $D^2u(x)(\frac{\nabla u(x)}{\|\nabla u\|}, \frac{\nabla u(x)}{\|\nabla u(x)\|})$ passes through zero.
- At each scale s, only keep the selected points that corresponds to $\|\nabla u(x)\| > \theta(s)$, where $\theta(s)$ is some threshold.

Remark:

1. A simple implementation consists in using finite differences to compute the image gradients (but more complicated techniques can also be involved).
2. Due to the filtering step, the computation of the second order derivatives can always be carried out since u is C^∞ as soon as u_0 est bounded.

However, in practice Canny's edge detector is not implemented that way because to find the zeros of $D^2u(x)(\frac{\nabla u(x)}{\|\nabla u\|}, \frac{\nabla u(x)}{\|\nabla u(x)\|})$ is difficult. On the contrary, it is usually implemented by finding the location of the maxima of the norm of the grey level gradients in the direction of the grey level gradients. To do so, one proceeds as follows:
given a filter $\Lambda : \mathbb{R}^2 \to \mathbb{R}$, one defines the gradients of the smooth image by:

$$\nabla(u_0 * \Lambda) = \begin{pmatrix} u * \frac{\partial}{\partial x}\Lambda \\ u * \frac{\partial}{\partial y}\Lambda \end{pmatrix} = \begin{pmatrix} W_x u \\ W_y u \end{pmatrix}$$

Then one computes $Mu = \sqrt{|W_x u|^2 + |W_y u|^2}$ the norm of the gradient and then the orientation of the gradient is approximated by:

$$Au = \begin{cases} \alpha(u) \text{ if } & W_x u \geq 0 \\ \pi + \alpha(u) \text{ if } & W_x u < 0. \end{cases}$$

Then, to find the edge points we look for the maximum of the modulus of the image gradient in the direction of the gradient by considering the following approximation:

- If $Af \in [-\frac{\pi}{8}, \frac{\pi}{8}[$ (modulo π), we compare $Mu(x,y)$ to Mu evaluated at $(x+1,y)$ and $(x-1,y)$.
- If $Af \in [\frac{\pi}{8}, \frac{3\pi}{8}[$ (modulo π), we compare $Mu(x,y)$ to Mu evaluated at $(x+1,y+1)$ and $(x-1,y-1)$.
- If $Af \in [\frac{3\pi}{8}, \frac{5\pi}{8}[$ (modulo π), we compare $Mu(x,y)$ to Mu evaluated at $(x,y+1)$ and $(x,y-1)$.
- If $Af \in [\frac{5\pi}{8}, \frac{7\pi}{8}[$ (modulo π), we compare $Mu(x,y)$ to Mu evaluated at $(x-1,y+1)$ and $(x+1,y-1)$.

We then chain the maxima thus obtained by moving along the direction orthogonal to the local gradient and then remove the chains with small length.
In Matlab, a procedure is predefined to compute such type of edges:
% Convert input image to double.

```
close all
load gatlin2;
colormap(gray);
[M,N] = size(X);
edge(X,'Canny');
```

Marr-Hildreth edge detector

Alternatively, one can try and seek to compute the zeros of the Laplacian operator. This is called the Marr-Hildreth edge detector.
Algo 2 (Marr-Hildreth):

- Convolve u_0 (original image) with a Gaussian filter with increasing size.
- At each scale, spot the points x such that $\|\nabla u(x)\| \neq 0$ and $\Delta u(x)$ changes sign.

The Marr-Hildreth operator corresponds to the search for the zeros of the image Laplacian convolve with a Gaussian function. If one defines $b(x,y) = \Delta(u * g(x,y))$, with g a Gaussian kernel, then due to the property of Gaussian functions, $b(x,y)$ amounts to convolving directly the image with the Laplacian operator Δg (Laplacian of Gaussian).
A implementation is given in matlab through:
% Convert input image to double.

```
close all
load gatlin2;
colormap(gray);
[M,N] = size(X);
edge(X,'log');
```

This can be well approximated by a difference of two Gaussian functions (called *DoG*). Indeed, in the unidimensional case, the *DoG* operator reads:

$$DoG(x) = \frac{1}{\sigma}e^{-\frac{x^2}{2\sigma^2}} - \frac{1}{\sigma_i}e^{-\frac{x^2}{2\sigma_i^2}}.$$

Assuming $\sigma_i = \sigma + \delta\sigma$, one writes :

$$DoG(x) = \frac{1}{\sigma}e^{-\frac{x^2}{2\sigma^2}} - \frac{1}{\sigma+\delta\sigma}e^{-\frac{x^2}{2(\sigma+\delta\sigma)^2}}.$$

Writing a first order Taylor expansion with respect to σ, we get that:

$$\begin{aligned} DoG(x) &\approx \delta\sigma\frac{\partial}{\partial\sigma}\left(\frac{1}{\sigma}e^{-\frac{x^2}{2\sigma^2}}\right) \\ &= -\left(\frac{1}{\sigma^2} - \frac{x^2}{\sigma^4}\right)e^{-\frac{x^2}{2\sigma^2}} \\ &= \sigma\frac{\partial^2 g(x)}{\partial x^2}. \end{aligned}$$

This operator naturally extends to the bidimensional case through:

$$DoG(x,y) = \frac{1}{(\sigma+\delta\sigma)^2}e^{-\frac{x^2+y^2}{2(\sigma+\delta\sigma)^2}} - \frac{1}{\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}}.$$

By means of a first order Taylor expansion with respect to σ, we obtain:

$$\begin{aligned} DoG(x,y) &= \delta\sigma\frac{\partial}{\partial\sigma}\left(\frac{1}{\sigma^2}e^{-\frac{x^2+y^2}{2\sigma^2}}\right) \\ &= \delta\sigma\left(-\frac{1}{\sigma^3} + \frac{1}{\sigma^2}\left(\frac{x^2+y^2}{\sigma^3}\right)\right)e^{-\frac{x^2+y^2}{2\sigma^2}} \end{aligned}$$

Differentiating twice the Gaussian function with respect to x, one obtains:

$$\frac{\partial^2}{\partial x^2} g(x,y) = \frac{1}{\sigma^2}\left(-\frac{1}{\sigma^2}+\frac{x^2}{\sigma^4}\right) e^{-\frac{x^2+y^2}{2\sigma^2}}.$$

One would obtain the same type of expression by differentiating twice with respect to y so that we finally get:

$$\frac{1}{\sigma\delta\sigma} DoG(x,y) \approx \Delta g(x,y).$$

This formulation is very much used in computer vision because it allows for a very fast computation of the Laplacian of the image. As we will see in the next chapter, the DoG algorithm is very much used to detect structure like blobs.

3.2 Edge detection with active contour method: energy based approach

A very different approach to the previous ones for edge detection is known as *active contour edge detection* and was initially proposed by Kass [6]. In the literature, this technique is sometimes refered to as edge detection based on *snakes*. It is a semi-interactive mehod in which the operator places in the image, and in the vicinity of the object to be detected, an initial contour line. This contour line is going to undergo different transformations under the action of several forces:

- An internal energy corresponding to tension and torsion forces.
- A potential energy aiming at sticking the contour onto the object of interest.

3.2.1 Direct Resolution

We first try to find whether it is possible to solve the problem directly, that is by finding directly the contour associated with the minimal energy. To do so, one first defines a contour using a parametric formulation involving the curvilinear abscissa s and time t:

$$v(s) = [x(s), y(s)]^t \quad s \in [0,1].$$

Then, welook for v that minimize an energy containing 2 different terms, as mentionned before which are as follows:

$$E_{global} = E_{internal}(v) + E_{image}(v).$$

To obtain a C^2 contour one considers [6]:

$$E_{internal} = \int_0^1 \alpha \|\frac{dv}{ds}\|^2 + \beta\|\frac{d^2v}{ds^2}\|^2 ds.$$

The first order derivative takes into account the length variations of the contour (it is a tension term which controls the elasticity of the studied contour), while the second order derivative monitors the curvature variations (this is a flexion term controlling the stiffness of the contour). The contour has to be both smooth and stiff and these two terms assume a C^2 regularity for the curve.
The second term E_{image} characterizes the lines in the image one wants to follow. In the case of edge detection, the lines correspond to high image gradient, and the energy is taken equal to:

$$E_{image} = -\int_0^1 \|\nabla u(v(s))\|^2 ds.$$

The global energy can be written as:

$$E_{global} = \int_0^1 \left[-\|\nabla u(v(s))\|^2 + \alpha \|\frac{\partial v}{\partial s}\|^2 + \beta \|\frac{\partial^2 v}{\partial^2 s}\|^2 \right] ds.$$

To compute the minimum of the energy, we make use of the important result of differential calculus that follows:

Theorem 1 *Let $s \in [0,1]$ and v a function belonging to $C_p^n(0,1)$, p being for periodic (with period 1). Let $F(s,v,v',v'',\cdots,v^{(n)})$ a C^1 function with respect to each of its variables, periodic with period 1, and such that F_{v^m} is C^m with respect to s. With these hypotheses, $J(v)$ the functional defined by:*

$$J(v) = \int_0^1 F(s,v,v',v'',\cdots,v^{(n)})ds,$$

is such that its extrema must satisfy the following equation:

$$\sum_{m=0}^{n} (-1)^m \frac{\partial^m F_{v^{(m)}}}{\partial s^m}(s,v,v',v'',\cdots,v^{(n)}) = 0,$$

where $F_{v^{(m)}}$ stands for the partial derivative of F with respect to $v^{(m)}$.

Proof: $C_p^n(0,1)$ equipped with $\|u\| = \max_i(\|u^{(i)}\|_{\infty,[0,1]})$ is a Banach space, so the extrema satisfy $J'(v) = 0$. Let us compute, the differential of J:

$$\begin{array}{rcl} J(v+k) & = & \int_0^1 F(s, v+k, v'+k', v''+k'', \cdots, v^{(n)}+k^{(n)})ds \\ & = & \int_0^1 F(s,v,v',v'',\cdots,v^{(n)})ds + \int_0^1 F_v k + F_{v'}k' + \cdots + F_{v^{(n)}}k^{(n)}ds + \|k\|\epsilon(k) \\ & = & \int_0^1 F(s,v,v',v'',\cdots,v^{(n)})ds + \sum_{m=0}^{n} \int_0^1 (-1)^m \frac{\partial^m F_{v^{(m)}}}{\partial s^m} k + \|k\|\epsilon(k) \\ & = & \int_0^1 F(s,v,v',v'',\cdots,v^{(n)})ds + \int_0^1 \sum_{m=0}^{n} (-1)^m \frac{\partial^m F_{v^{(m)}}}{\partial s^m} k + \|k\|\epsilon(k). \end{array}$$

The last equality is obtained by integration by parts using the fact that the expression under consideration are equal in 0 and 1. The final result is then obtained by remarking that $C_p^n(0,1)$ is dense in $L^2(0,1)$ (One can also notice that the property is true for $k \in \mathcal{D}(0,1)$) $\blacksquare$.
In our case, s is the curvilinear abscissa and, to simplify, we will note:

$$\begin{array}{rclrcl} x' & = & \frac{dx(s)}{ds} & y' & = & \frac{dy(s)}{ds} \\ x'' & = & \frac{d^2x(s)}{d^2s} & y'' & = & \frac{d^2y(s)}{d^2s} \\ v' & = & \frac{\partial v(s)}{\partial s} & v'' & = & \frac{\partial^2 v(s)}{\partial s^2}. \end{array}$$

The energy to be minimized has thus the following form:

$$E_{global} = \int_0^1 \alpha(x'^2 + y'^2) + \beta(x''^2 + y''^2) - \|\nabla u(v(s))\|^2 ds.$$

Differentiating E_{global}, on $C_p^2(0,1)$ assuming F satisfies the hypotheses of the theorem, we get:

$$\begin{array}{rcl} E_{global}(v+h) & = & \int_0^1 \alpha\|v'+h'\|^2 + \beta\|v''+h''\|^2 - \|\nabla u(v(s)+h(s))\|^2 ds \\ & = & E_{global}(v) + \int_0^1 2\alpha\langle v', h'\rangle + 2\beta\langle v'', h''\rangle - 2\langle D^2u(v)\nabla u(v), h\rangle ds + \|h\|\epsilon(h) \\ & = & E_{global}(v) + 2\int_0^1 \langle -\alpha v'' + \beta v^{(4)} - 2D^2u\nabla u, h\rangle ds + \|h\|\epsilon(h), \end{array}$$

where $\|h\| = \max(\|h\|_{2,[0,1]}, \|h'\|_{2,[0,1]}, \|h''\|_{2,[0,1]})$. In such a case, v satisfies the following equation:

$$-\alpha v''(s) + \beta v^{(4)}(s) = D^2u(v(s))\nabla u(v(s)).$$

Since the term on the right hand side also depends on $v(s)$ a direct resolution is not feasible, and the sought minimum is going to be defined as the steady sate of a certain evolution equation as explained hereafter. That is why we need the time parameter.

3.2.2 Active contour: formulation using finite differences

To find out the curves associated with the minimal energy requires to solve a partial differential equation which involves a discretization step by using finite differences.
So, we first consider a discretized version $\{v_i = (x_i, y_i), i = 0, \cdots, N-1\}$ of v. At location v_i, the first and second order derivatives of the first component x of vector v are respectively approximated by $\frac{x_i - x_{i-1}}{h}$ and $\frac{x_{i+1} - 2x_i + x_{i-1}}{h^2}$ (and similarly for y), with $h = \frac{1}{N}$. The equations for x and y are independent. In that discrete framework, the equation $-\alpha x'' + \beta x^{(4)} = f$ becomes:

$$\begin{gathered}-\alpha \tfrac{1}{h^2}\left(x_{i+1} - 2x_i - x_{i-1}\right) + \beta \tfrac{1}{h^4}\left((x_{i+2} - 2x_{i+1} + x_i) - 2(x_{i+1} - 2x_i + x_{i-1}) + (x_i - 2x_{i-1} + x_{i-2})\right) \\ = f_{i+1} \\ \Leftrightarrow \beta x_{i-2} + (-4\beta - h^2\alpha)x_{i-1} + (6\beta + 2h^2\alpha)x_i + (-4\beta - h^2\alpha)x_{i+1} + \beta x_{i+2} = h^4 f_{i+1}.\end{gathered}$$

We then obtain a linear sytem of the form $BX = F$, where:

$$\begin{gathered} X = (x_i)_{i=0,\cdots,N-2} \\ F = (h^4 f_{i+1})_{i=0,\cdots,N-2}, \end{gathered}$$

where B is the matrix defined just below. In the case of a closed curve, and taking into account the fact that the later is periodic, the matrix B is circulant and can be written as (we take h equal 1 for the sake of simplicity):

$$B = \begin{pmatrix} 2\alpha+6\beta & -\alpha-4\beta & \beta & 0 & \cdots & \beta & -\alpha-4\beta \\ -\alpha-4\beta & 2\alpha+6\beta & -\alpha-4\beta & \beta & \cdots & 0 & \beta \\ \beta & -\alpha-4\beta & 2\alpha+6\beta & -\alpha-4\beta & \beta & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \\ 0 & \cdots & \beta & -\alpha-4\beta & 2\alpha+6\beta & -\alpha-4\beta & \beta \\ \beta & 0 & \cdots & \beta & -\alpha-4\beta & 2\alpha+6\beta & -\alpha-4\beta \\ -\alpha-4\beta & \beta & 0 & \cdots & \cdots & -\alpha-4\beta & 2\alpha+6\beta \end{pmatrix}.$$

As already mentioned above, since the right hand term depends on v, a direct resolution os not possible. The problem is thus solved iteratively by studying the following partial differential equation:

$$\frac{\partial v}{\partial t} - \alpha v'' + \beta v^{(4)} = D^2u(v)\nabla u(v) = F(v), \tag{3.1}$$

to which one can associate the following so-called explicit scheme:

$$h^4 \frac{x_i^{t+\Delta t} - x_i^t}{\Delta t} + \beta x_{i-2}^t - (4\beta + h^2\alpha)x_{i-1}^t + (6\beta + 2h^2\alpha)x_i^t - (4\beta + h^2\alpha)x_{i+1}^t + \beta x_{i+2}^t = h^4 F(x_i^t),$$

which can be written in the following matrix form:

$$X(t + \Delta t) = (I + \Delta t B)X(t) + \Delta t F(X(t)),$$

which may be unstable depending on the choice for Δt with respect to h. If one considers the resolution of the problem under its implicit form, one finally gets:

$$(I + \Delta t B)X(t + \delta t) = (X(t) + \Delta t F(X(t))).$$

3.2.3 Active contour: variational approach

In this formulation, we consider the original expression of the differential, i.e. without integrating by parts and simplifying using periodic conditions. The problem can be written as follows:

$$a(v,w) = \int_0^1 \alpha\langle v', w'\rangle + \beta\langle v'', w''\rangle ds = \int_0^1 \langle F(v), w\rangle = L_v(w).$$

in which $\langle .,. \rangle$ denotes the inner product on $\mathbb{R}^2$. a is a bilinear form on the Sobolev space $H_p^2(0,1)$ (p being for periodic). Then, we write the variational formulation corresponding to our problem. We denote $\langle .,. \rangle_2$ the inner product on $L^2(0,1)$, and then the variational problem reads (v is now a function belonging to $L^2(0,T,H_p^2(0,1))$:

$$\frac{\partial}{\partial t}\langle v, w\rangle_2 + a(v,w) = L_v(w), \forall w \in H_p^2(0,1)$$

We now need to find the appropriate subspace V_h of $H_p^2(a,b)$ satisfying relevant limit conditions and in which we can solve the following variational problem.
That is we determine $v_h \in V_h$ such that :

$$\forall w \in V_h, \frac{\partial}{\partial t}\langle v_h(t), w\rangle + a(v_h(t), w) = L_{v_h}(w). \tag{3.2}$$

This formulation enables the resolution of the problem using the finite elements method. In practice, the problem is solved that way: one first remark that $a(v,w) = L_v(w)$ admits a unique solution in $H_p^2(]0,1[)$, provided the bilinear form a is coercive ($a(x,x) \geq \|x\|^2 \ \forall\ x \in V$), which is the case as soon as $\alpha(s)$ and $\beta(s)$ are positive (to prove it, one computes for each coordinate its mean over $]0,1[$ and then subtract it to the considered coordinate. Then one applies the Poincaré-Wirtinger inequality).
Having shown the existence of a solution to the stationary problem, one approaches this solution in a finite dimension sub-space V_h of V (Galerkin method):

$$a(v_h, w_h) = L_{v_h}(w_h)\ \forall w_h \in V_h,$$

where (w_h) is a basis for V_h. Decomposing v_h onto the basis (w_h), one can write the previous equality as a linear system: $A\mathcal{V} = L$, where $\mathcal{V}$ corresponds to the coordinates of v_h in the basis of finite elements (w_h). Nevertheless, since the right hand side term also depends on v, one cannot ensure that v_h converges to v. One therefore finally discretizes the evolution equation (3.2) using finite differences to obtain:

$$\frac{\mathcal{V}^t - \mathcal{V}^{t-1}}{\Delta t} + A\mathcal{V}^{t-1} = L_{\mathcal{V}^{t-1}}$$

3.3 Edge detection with active contour method: geodesic based approach

In many different papers it was remarked that the term involving the second order derivative of the curve was somehow useless, and a variant was proposed through the minimization of [4]:

$$E(v) = \alpha\int_0^1 \|v'(s)\|^2 ds - \lambda\int_0^1 \|\nabla u(v(s))\| ds. \tag{3.3}$$

Note that another difference with the previous minimization is that the modulus of the norm is considered instead of the squared modulus. The above minimization can be generalized to:

$$E(v) = \alpha \int_0^1 \|v'(s)\|^2 ds - \lambda \int_0^1 g(\|\nabla u(v(s))\|)^2 ds.$$

with $g : [0, +\infty[\to \mathbb{R}^+$ a strictly decreasing function such that $g(r) \to 0$ when $r \to \infty$. One remaining issue is that the energy to minimize depend on the parametrization of the curve. To find an intrinsic description of the curve the concept of geodesic is used. Indeed, it is possible to show that the solution to (3.3) is given by the geodesic curve in a Riemannian space induced from the image u (A geodesic is a (local) minimal path between given points).
We can rewrite:

$$E(v) = \int_0^1 \mathcal{L}(v(s), v'(s)) ds,$$

where $\mathcal{L}$ is the Lagrangian. We recall that for such an energy the Euler-Lagrange equation:

$$\frac{\partial \mathcal{L}}{\partial v} - \frac{d}{ds} \frac{\partial \mathcal{L}}{\partial v'} = 0, \ \forall s$$

Now has the Lagrangian does not depend explicitly on s, we have the following proposition:

Theorem 1 (Beltrami)

If the Lagrangian does not depend on s then the Euler-Lagrange equation is equivalent to the following Beltrami equation:

$$\mathcal{L}(v(s), v'(s)) - \frac{\partial \mathcal{L}}{\partial v'}(v(s), v'(s)) v'(s) = C$$

proof Remarking that we have

$$\frac{d}{ds} \mathcal{L}(v(s), v'(s)) = \frac{\partial \mathcal{L}}{\partial v}(v(s), v'(s)) v'(s) + \frac{\partial \mathcal{L}}{\partial v'}(v(s), v'(s)) v''(s)$$

we deduce from the Euler Lagrange equation that

$$\frac{\partial \mathcal{L}}{\partial v}(v(s), v'(s)) v'(s) = \frac{d}{ds} \frac{\partial \mathcal{L}}{\partial v'} v'(s) = \frac{d}{ds} L(v(s), v'(s))) - \frac{\partial \mathcal{L}}{\partial v'}(v(s), v'(s)) v''(s)$$

Thus

$$\frac{d}{ds} \mathcal{L}(v(s), v'(s))) = \frac{d}{ds} \frac{\partial \mathcal{L}}{\partial v'}(v(s), v'(s)) v'(s) + \frac{\partial \mathcal{L}}{\partial v'}(v(s), v'(s)) v''(s) = \frac{d}{ds} \left(\frac{\partial \mathcal{L}}{\partial v'}(v(s), v'(s)) v'(s) \right).$$

Hence the result.

The Hamiltonian is then given by $H = \frac{\partial \mathcal{L}}{\partial v'}(v(s), v'(s)) v'(s) - \mathcal{L}(v(s), v'(s))$is a constant independent from s in that case. Thus the Hamiltonian can be rewritten as:

$$H = \alpha \|v'\|^2 - \lambda g(\|\nabla u(v(s))\|)^2$$

So we are looking for the minimizing the energy above under the constraint that $H = E_0$. Choosing $E_0 = 0$, one then easily see that the energy minimization amounts to minimizing the following problem:

$$\int_0^1 g(\|\nabla u(v(s))\|)\|v'(s)\|ds.$$

This approach to active contour is implemented in Matlab. Here is an illustration of how the procedure works:

```
I = imread('coins.png');
imshow(I)
title('Original Image')
```

% Computation of the mask

```
mask = zeros(size(I));
mask(25:end-25,25:end-25) = 1;
imshow(mask)
title('Initial Contour Location')
pause

bw = activecontour(I,mask);
imshow(bw)
title('Segmented Image, 100 Iterations')
pause

bw = activecontour(I,mask,300);
imshow(bw)
title('Segmented Image, 300 Iterations')
pause
```

Bibliography

[1] L. Alvarez, P.L. Lions, and J.M. Morel, *Image Selective Smoothing and Edge Detection by Nonlinear Diffusion. II*,SIAM Journal Numer. Anal., vol 29, no. 3, pp. 845-866, 1992.

[2] A. Chambolle, *An Algorithm for Total Variations Minimization and Applications*, Journal of Mathematical imaging and Vision, pp. 89-97, 2004.

[3] J. Canny, *A Computational Approach to Edge Detection*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 8, 679-698, 1986.

[4] V. Caselles, R. Kimmel, and G. Sapiro, *Geodesic active contours*, International journal of computer Vision, vol. 22, pp. 61-79, 1997

[5] Deriche R., *Using Canny's criteria to derive recursively implemented optimal edge detector*, International Journal of Computer Vision, vol. 1, pp. 167-187, 1987.

[6] Kass M., Witkin A., Terzopoulos D., *Active Contour Models*, International Journal of computer Vision,vol. 1, pp. 321-331, 1988.

[7] Mackworth A, Mokhtarian F., *A theory of Multiscale, curvature-based shape representation for planar curves*, IEEE Transactions on Pattern Analysis and Machine Intelligence,vol. 14, pp. 789-805, 1992.

[8] D. Mumford and J. Shah, *Boundary Detection by Minimizing Functionals*, IEEE Conf. Computer Vision and Pattern Recognition (CVPR '85), san Francisco, pp. 22-26, 1985.

[9] N. Nordström, *Biased Anisotropic Diffusion- a unified regularization and diffusion approach to edge detection*, Image Vision Comput., vol. 8, pp. 318-327, 1990.

[10] Perona. P and malik. J, *Scale Space and Edge Detection using Anisotropic Diffusion*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 12, pp. 629-639, 1990.

[11] A. Witkin, *Scale-space Filtering*, International Joint Conference on Artificial Intellignece, pp. 1019-1023, 1983.

[12] S. Benzoni-Gavage, *Calcul différentiel et équations différentielles : Cours et exercices corrigés*, Dunod, (2010).

[13] J.-M. Bony, *Cours d'analyse, théorie des distributions et analyse de Fourier*, Éditions de l'École Polytechnique, (2001).