Calculus of variations in image processing

Jean-François AUJOL

CMLA, ENS Cachan, CNRS, UniverSud,
61 Avenue du Président Wilson, F-94230 Cachan, FRANCE

Email : Jean-Francois.Aujol@cmla.ens-cachan.fr
http://www.cmla.ens-cachan.fr/membres/aujol.html

18 september 2008

Note: This document is a working and uncomplete version, subject to errors and changes. Readers are invited to point out mistakes by email to the author.

This document is intended as course notes for second year master students. The author encourage the reader to check the references given in this manuscript. In particular, it is recommended to look at [9] which is closely related to the subject of this course.

Contents

1 Inverse problems in image processing **4**
1.1 Introduction . . . 4
1.2 Examples . . . 4
1.3 Ill-posed problems . . . 6
1.4 An illustrative example . . . 8
1.5 Modelization and estimator . . . 10
1.5.1 Maximum Likelihood estimator . . . 10
1.5.2 MAP estimator . . . 11
1.6 Energy method and regularization . . . 11

2 Mathematical tools and modelization **14**
2.1 Minimizing in a Banach space . . . 14
2.2 Banach spaces . . . 14
2.2.1 Preliminaries . . . 14
2.2.2 Topologies in Banach spaces . . . 17
2.2.3 Convexity and lower semicontinuity . . . 19
2.2.4 Convex analysis . . . 23
2.3 Subgradient of a convex function . . . 23
2.4 Legendre-Fenchel transform: . . . 24
2.5 The space of funtions with bounded variation . . . 26
2.5.1 Introduction . . . 26
2.5.2 Definition . . . 27
2.5.3 Properties . . . 28
2.5.4 Decomposability of $BV(\Omega)$: . . . 28
2.5.5 SBV . . . 30
2.5.6 Sets of finite perimeter . . . 32
2.5.7 Coarea formula and applications . . . 33

3 Energy methods **35**
3.1 Introduction . . . 35
3.2 Tychonov regularization . . . 35
3.2.1 Introduction . . . 35
3.2.2 Sketch of the proof (to fix some ideas) . . . 36
3.2.3 General case . . . 36
3.2.4 Minimization algorithms . . . 39
3.3 Rudin-Osher-Fatemi model . . . 40
3.3.1 Introduction . . . 40
3.3.2 Interpretation as a projection . . . 45
3.3.3 Euler-Lagrange equation for (3.67): . . . 47
3.3.4 Other regularization choices . . . 48
3.3.5 Half quadratic minimmization . . . 50
3.4 Wavelets . . . 53
3.4.1 Besov spaces . . . 53

3.4.2 Wavelet shrinkage . . . 54
3.4.3 Variational interpretation . . . 55

4 Advanced topics: Image decomposition **56**
4.1 Introduction . . . 56
4.2 A space for modeling oscillating patterns in bounded domains . . . 57
4.2.1 Definition and properties . . . 57
4.2.2 Study of Meyer problem . . . 60
4.3 Decomposition models and algorithms . . . 61
4.3.1 Which space to use? . . . 61
4.3.2 Parameter tuning . . . 62
4.3.3 $TV - G$ algorithms . . . 64
4.3.4 $TV - L^1$ . . . 65
4.3.5 $TV - H^{-1}$ . . . 66
4.3.6 TV-Hilbert . . . 67
4.3.7 Using negative Besov space . . . 68
4.3.8 Using Meyer's E space . . . 69
4.3.9 Applications of image decomposition . . . 71

A Discretization **71**

1. Inverse problems in image processing

For this section, we refer the interested reader to [71]. We encourage the reader not familiar with matrix to look at [29].

1.1 Introduction

In many problems in image processing, the goal is to recover an ideal image u from an observation f.

u is a perfect original image describing a real scene.

f is an observed image, which is a degraded version of u.

The degradation can be due to:

- Signal transmission: there can be some noise (*random* perturbation).
- Defects of the imaging system: there can be some blur (*deterministic* perturbation).

The simplest modelization is the following:

$$f = Au + v \tag{1.1}$$

where v is the noise, and A is the blur, a linear operator (for example a convolution).

The following assumptions are classical:

- A is known (but often not invertible)
- Only some statistics (mean, variance, ...) are know of n.

1.2 Examples

Image restoration (Figure 1)

$$f = u + v \tag{1.2}$$

with n a white gaussian noise with standard deviation σ.

Radar image restoration (Figure 2)

$$f = uv \tag{1.3}$$

with v a gamma noise with mean one.

Poisson distribution for tomography.

Image decomposition (Figure 3)

$$f = u + v \tag{1.4}$$

u is the *geometrical* component of the original image f (u can be seen a s a sketch of f),

and v is the *textured* component of the original image f (v contains all the details of f).

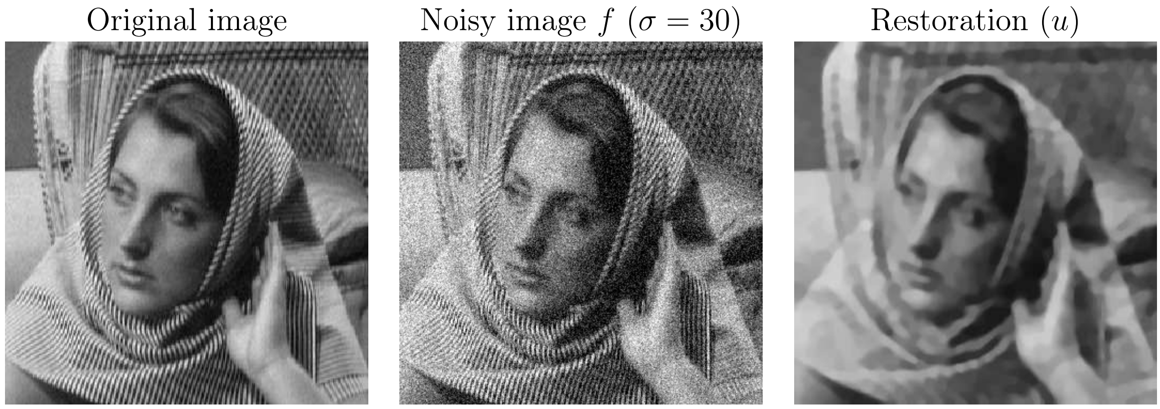

Figure 1: Denoising

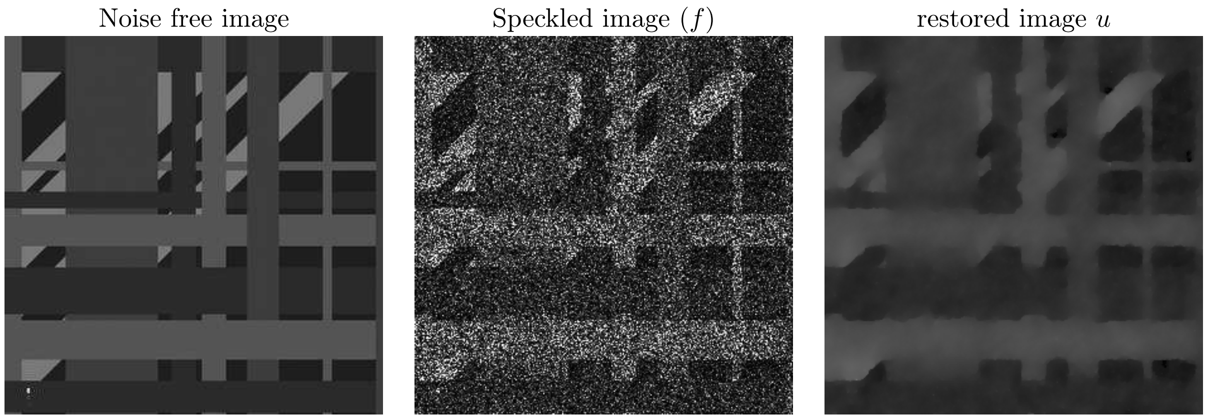

Figure 2: Denoising of a synthetic image with gamma noise. f has been corrupted by some multiplicative noise with gamma law of mean one.

Figure 3: Image decomposition

Figure 4: Example of TV deconvolution

Figure 5:

Image deconvolution (Figure 4)

$$f = Au + v \tag{1.5}$$

Image inpainting [56] (Figure 5)

Zoom [54] (Figure 6)

1.3 Ill-posed problems

Let X and Y be two Hilbert spaces. Let $A : X \to Y$ a continous linear application (in short, an operator).

Consider the following problem:

$$\text{Given } f \in Y, \text{find } u \in X \text{ such that } f = Au \tag{1.6}$$

The problem is said to be *well-posed* if

(i) $\forall f \in Y$ there exists a unique $u \in X$ such that $f = Au$.

(ii) The solution u depends continuously on f.

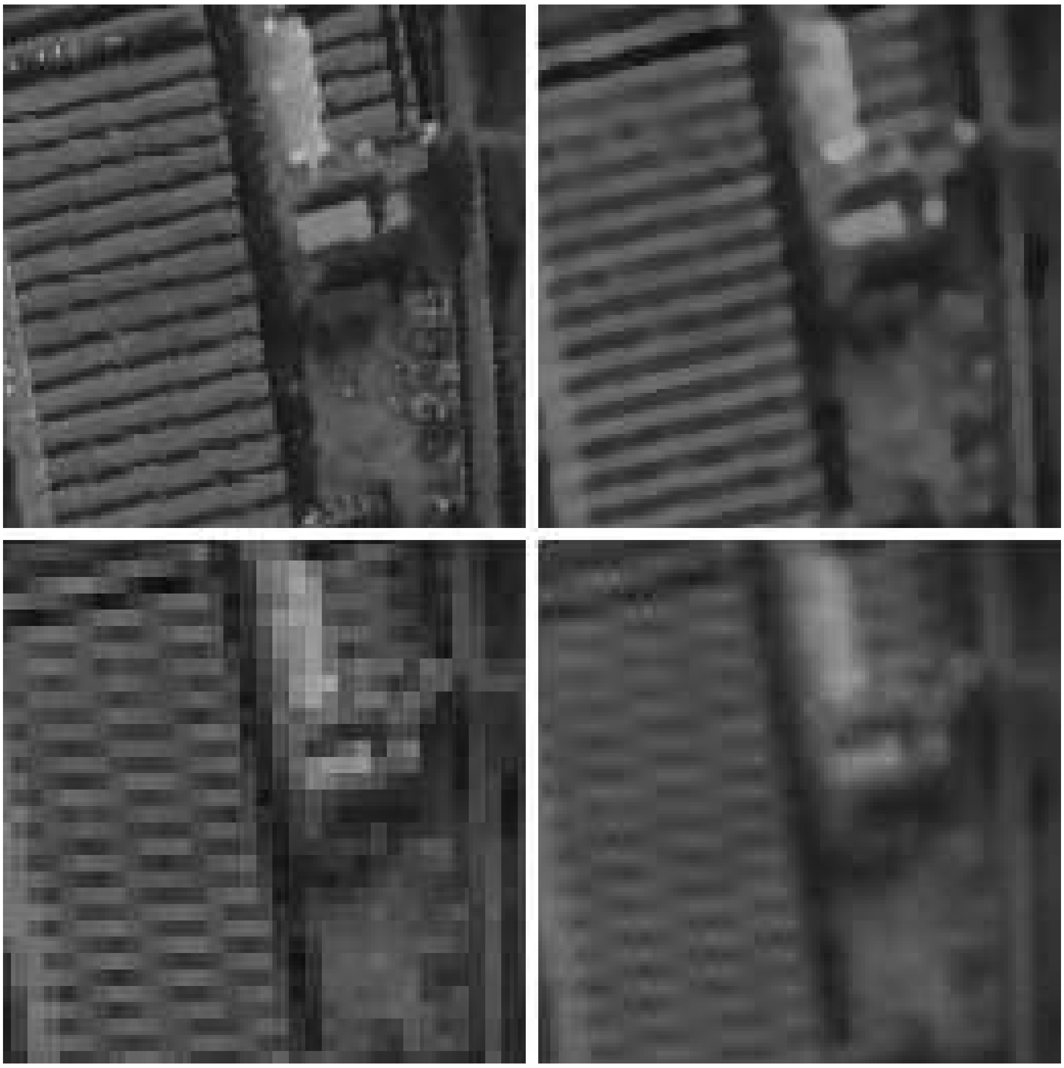

Figure 6: Top left: ideal image; top right: zoom with total variation minimization; bottom left: zoom by pixel duplication; bottom right: zoom with cubic splines

In other words, the problem is well-posed if A is invertible and its inverse $A^{-1} : Y \to X$ is continuous.

Conditions (i) and (ii) are referred to as the *Hadamard* conditions.

A problem that is not well-posed is said to be *ill-posed*.

Notice that a mathematically well-posed problem may be ill-posed in practice: the solution may exist, be unique, and depend continuously on the data, but still be very sensitive to small perturbations of it. An eror δf produces the error $\delta u = A^{-1}\delta f$, which may have dramatic consequences on the interpretation of the solution. In particular, if $\|A^{-1}\|$ is very large, errors may be strongly amplified by the action of A^{-1}. There can also be some computational time issues.

1.4 An illustrative example

See [29] for the defintion of $\|.\|_2$ of a vector, a matrix, a positive symmetric matrix, an orthogonal matrix, ...

We consider the following problem:

$$f = Au + v \tag{1.7}$$

$\|v\|$ is the the amount of noise.

We assume that A is a *real symetric positive matrix*, and has some small eigenvalues. $\|A^{-1}\|$ is thus very large. We want to compute a solution by filtering.

Since A is symetric, there exists an orthogonal matrix P (i.e. $P^{-1} = P^T$) such that:

$$A = PDP^T \tag{1.8}$$

with $D = \text{diag}(\lambda_i)$ a diagonal matrix, and $\lambda_i > 0$ for all i.

We have:

$$A^{-1}f = u + A^{-1}v = u + PD^{-1}P^T v \tag{1.9}$$

with $D^{-1} = \text{diag}(\lambda_i^{-1})$. It is easy to see that instabilities arise from small eigenvalues λ_i.

Regularization by filtering: One way to overcome this problem consists in modifying the λ_i^{-1}: we multiply them by $w_\alpha(\lambda_i^2)$. w_α is chosen such that:

$$w_\alpha(\lambda^2)\lambda^{-1} \to 0 \text{ when } \lambda \to 0. \tag{1.10}$$

This filters out singular components from $A^{-1}f$ and leads to an approximation to u by u_α defined by:

$$u_\alpha = PD_\alpha^{-1}P^T f \tag{1.11}$$

where $D_\alpha^{-1} = \text{diag}(w_\alpha(\lambda_i^2)\lambda_i^{-1})$.

To obtain some accuracy, one must retain singular components corresponding to large singular values. This is done by choosing $w_\alpha(\lambda^2) \approx 1$ for large values of λ.

For w_α, we may chose (truncated SVD):

$$w_\alpha(\lambda^2) = \begin{cases} 1 \text{ if } \lambda^2 > \alpha. \\ 0 \text{ if } \lambda^2 \leq \alpha. \end{cases} \tag{1.12}$$

We may also choose a smoother function (Tychonov filter function):

$$w_\alpha(\lambda^2) = \frac{\lambda^2}{\lambda^2 + \alpha} \tag{1.13}$$

An obvious question arises: *can the regularization parameter α be selected to guarantee convergence as the error level $\|v\|$ goes to zero?*

Error analysis: We denote by R_α the regularization operator:

$$R_\alpha = PD_\alpha^{-1}P^T \tag{1.14}$$

We have $u_\alpha = R_\alpha f$. The regularization error is given by:

$$e_\alpha = u_\alpha - u = R_\alpha Au - u + R_\alpha v = e_\alpha^{trunc} + e_\alpha^{noise} \tag{1.15}$$

where:

$$e_\alpha^{trunc} = R_\alpha Au - u = P(D_\alpha^{-1}D - Id)P^T u \tag{1.16}$$

and:

$$e_\alpha^{noise} = R_\alpha v = PD_\alpha^{-1}P^T v \tag{1.17}$$

e_α^{trunc} *is the error due to the regularization (it quantifies the loss of information due to the regularizing filter).* e_α^{noise} *is called the noise amplification error.*

We first deal with e_α^{trunc}. Since $w_\alpha(\lambda^2) \to 1$ as $\alpha \to 0$, we have $D_\alpha^{-1} \to D^{-1}$ as $\alpha \to 0$ and thus:

$$e_\alpha^{trunc} \to 0 \text{ as } \alpha \to 0. \tag{1.18}$$

To deal with the noise amplification error, we use the following inequality for $\lambda > 0$:

$$w_\alpha(\lambda^2)\lambda^{-1} \leq \frac{1}{\sqrt{\alpha}} \tag{1.19}$$

Remind that $\|P\| = 1$ since P orthogonal. We thus deduce that:

$$e_\alpha^{noise} \leq \frac{1}{\sqrt{\alpha}}\|v\| \tag{1.20}$$

where we recall that $\|v\| = \|v\|$ is the amount of noise. To conclude, it suffice to choose $\alpha = \|v\|^p$ with $p < 2$, and let $\|v\| \to 0$: we get $e_\alpha^{noise} \to 0$.

Now, if we choose $\alpha = \|v\|^p$ with $0 < p < 2$, we have:

$$e_\alpha \to 0 \text{ as } \|v\| \to 0. \tag{1.21}$$

For such a regularization parameter choice, we say that the method is *convergent*.

Rate of convergence: We assume a range condition:

$$u = A^{-1}z \tag{1.22}$$

Since $A = PDP^T$, we have:

$$e_\alpha^{trunc} = P(D_\alpha^{-1}D - Id)P^T u = P(D_\alpha^{-1} - D^{-1})P^T z \tag{1.23}$$

Hence:

$$\|e_\alpha^{trunc}\|_2^2 \leq \|D_\alpha^{-1} - D^{-1}\|^2\|z\|^2 \tag{1.24}$$

Since $D_\alpha^{-1} - D^{-1} = \text{diag}((w_\alpha(\lambda_i^2) - 1)\lambda_i^{-1})$, we deduce that:

$$\|e_\alpha^{trunc}\|_2^2 \leq \alpha\|z\|^2 \tag{1.25}$$

We thus get:

$$\|e_\alpha\| \leq \sqrt{\alpha}\|z\| + \frac{1}{\sqrt{\alpha}}\|v\| \tag{1.26}$$

The right-hand side is minimized by taking $\alpha = \|v\|/\|z\|$. This yields:

$$\|e_\alpha\| \leq 2\sqrt{\|z\|\|v\|} \tag{1.27}$$

Hence the convergence order of the method is $0(\sqrt{\|v\|})$.

1.5 Modelization and estimator

We consider the following additive model:

$$f = Au + v \tag{1.28}$$

Idea: We want to comput the most likely u with respect to the observation f. Natural probabilistic quantities: $P(F|U)$, and $P(u|F)$.

1.5.1 Maximum Likelihood estimator

Let us assume that the noise v follows a Gaussian law with zero mean:

$$P(V = v) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(v)^2}{2\sigma^2}\right) \tag{1.29}$$

We have $P(F = f|U = u) = P(F = Au + v|U = u) = P(V = f - Au)$.

We thus get:

$$P(F = f|U = u)(f|u) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(f - Au)^2}{2\sigma^2}\right) \tag{1.30}$$

We want to maximize $P(F|U)$. Let us remind the reader that the image is discretized, and that we denote by $\mathcal{S}$ the set of the pixels of the image. We also assume that the samples of the noise on each pixel $s \in S$ are mutually independent and identically distributed (i.i.d). We therefore have:

$$P(F|U) = \prod_{s \in \mathcal{S}} P(F(s)|U(s)) \tag{1.31}$$

where $F(s)$ (resp. $U(s)$) is the instance of the variable F (resp. U) at pixel s.

Maximizing $P(F|U)$ amounts to minimizing the log-likelihood $-\log(P(F|U)$, which can be written:

$$-\log\left(P(F|U)\right) = -\sum_{s \in S} \log\left(P(F(s)|U(s))\right) \tag{1.32}$$

We eventually get:

$$-\log\left(P(F|U)\right) = \sum_{s \in \mathcal{S}} \left(-\log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) + \frac{(F(s) - AU(s))^2}{2\sigma^2}\right) \tag{1.33}$$

We thus see that minimizing $-\log\left(P(F|U)\right)$ amounts to minimizing:

$$\sum_{s \in \mathcal{S}} \left(F(s) - AU(s)\right)^2 \tag{1.34}$$

Getting back to continuous notations, the data term we consider is therefore:

$$\int (f - Au)^2 \tag{1.35}$$

1.5.2 MAP estimator

As before, we have:

$$P(F = f|U = u)(f|u) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(f - Au)^2}{2\sigma^2}\right) \tag{1.36}$$

We also assume that u follows a Gibbs prior:

$$P(U = u) = \frac{1}{Z} \exp\left(-\gamma\phi(u)\right) \tag{1.37}$$

where Z is a normalizing constant.

We aim at maximizing $P(U|F)$. This will lead us to the classical Maximum a Posteriori estimator. From Bayes rule, we have:

$$P(U|F) = \frac{P(F|U)\, P(U)}{P(F)} \tag{1.38}$$

Maximizing $P(U|F)$ amounts to minimizing the log-lihkelihood $-\log(P(U|F)$:

$$-\log(P(U|F) = -\log(P(F|U)) - \log(P(U)) + \log(P(F)) \tag{1.39}$$

As in the previous section, the image is discretized. We denote by $\mathcal{S}$ the set of the pixel of the image. Moreover, we assume that the sample of the noise on each pixel $s \in \mathcal{S}$ are mutually independent and identically distributed (i.i.d) with density g_V. Since $\log(P(F))$ is a constant, we just need to minimize:

$$-\log\left(P(F|U)\right) - \log(u) = -\sum_{s\in\mathcal{S}} \left(\log\left(P(F(s)|U(s))\right) - \log(P(U(s)))\right) \tag{1.40}$$

Since Z is a constant, we eventually see that minimizing $-\log\left(P(F|U)\right)$ amounts to minimizing:

$$\sum_{s\in\mathcal{S}} \left(-\log\left(\frac{1}{\sqrt{2\pi\sigma^2}}\right) + \frac{(F(s) - AU(s))^2}{2\sigma^2} + \gamma\phi(U(s))\right) \tag{1.41}$$

Getting back to continuous notations, this lead to the following functional:

$$\int \left(\frac{(f - Au)^2}{2\sigma^2}\right) dx + \gamma \int \phi(u)\, dx \tag{1.42}$$

1.6 Energy method and regularization

From the ML method, one sees that many image processing problems boil down to the following minimization problem:

$$\inf_u \int_\Omega |f - Au|^2\, dx \tag{1.43}$$

If a minimizer u exists, then it satisfies the following equation:

$$A^* f - A^* A u = 0 \tag{1.44}$$

where A^* is the adjoint operator to A.

This is in general an *ill-posed* problem, since A^*A is not always one-to-one, and even in the case when it is one-to-one its eigenvalues may be small, causing numerical instability.

A classical approach in inverse problems consists in introducing a regularization term, that is to consider a related problem which admits a unique solution:

$$\inf_u \int_\Omega |f - Au|^2 + \lambda L(u) \tag{1.45}$$

where L is a non-negative function.

The choice of the regularization is influenced by the following points:

- *Well-posedness* of the solution u_λ.
- *Convergence:* when $\lambda \to 0$, one wants $u_\lambda \to u$.
- *Convergence rate.*
- *Qualitative stability estimate.*
- *Numerical algorithm.*
- *Modelization:* the choice of L must be in accordance with the expected properties of u $\Longrightarrow$ link with MAP approach.

Relationship between Tychonov regularization and Tychonov filtering: Let us consider the following minimization problem:

$$\inf_u \|f - Au\|_2^2 + \alpha \|u\|_2^2 \tag{1.46}$$

We denote by u_α its solution. We want to show that u_α is the same solution as the one we got with the Tychonov filter in subsection 1.4. We propose two different methods.

1. Let us set:
$$F(u) = \|f - Au\|_2^2 + \alpha \|u\|_2^2 \tag{1.47}$$
We first compute ∇F. We make use of: $F(u+h) - F(u) = \langle h, \nabla F \rangle + o(\|h\|)$. We have indeed:
$$F(u+h) - F(u) = 2\langle h, \alpha u + A^T A u - A^T f \rangle + o(\|h\|) \tag{1.48}$$
We also have $A^T = A$ since A symmetric. Hence:
$$\nabla F(u) = (\alpha Id + A^2)u - Af = (P^T(\alpha Id + D^2)Pu - Af \tag{1.49}$$
where we remind that $A = P^T D P$, and $D = \mathrm{diag}(\lambda_i)$ (consider also the case $\alpha = 0$, what happens?). We have: $\alpha Id + D^2 = \mathrm{diag}(\alpha + \lambda_i^2)$.
But it is easy to see that u_α is a solution of $\nabla F(u) = 0$, and then to conclude.

2. We have:

$$\|f - Au\|^2 + \alpha\|u\|^2 = \|f\|^2 + \|Au\|^2 + \alpha\|u\|^2 - 2\langle f, Au\rangle \tag{1.50}$$

But $Au = PDP^T u = PDw$ with

$$w = P^T u \tag{1.51}$$

And thus $\|Au\| = \|Dw\|$ since P orthogonal. Moreover, we have $u = PP^T u = Pw$ and therefore $\|u\| = \|w\|$. We also have:

$$\langle f, Au\rangle = \langle f, PDP^T u\rangle = \langle P^T f, DP^T u\rangle = \langle g, Dw\rangle \tag{1.52}$$

with

$$g = P^T f \tag{1.53}$$

Hence we see that minimizing (1.50) with respect to u amounts to minimizing (with respect to w):

$$\|Dv\|^2 + \alpha\|w\|^2 - 2\langle g, Dw\rangle = \sum_i F_i(w_i) \tag{1.54}$$

where:

$$F_i(w_i) = (\lambda_i^2 + \alpha)w_i^2 - 2\lambda_i g_i w_i \tag{1.55}$$

We have $F'(v_i) = 0$ when $(\lambda_i^2 + \alpha)v_i - \lambda_i g_i = 0$, i.e. $v_i = \lambda_i g_i/(\lambda_i^2 + \alpha)$. Hence (1.54) is minimized by

$$w_\alpha = D_\alpha^{-1} g \tag{1.56}$$

We eventually get that:

$$u_\alpha = Pw_\alpha = PD_\alpha^{-1} P^T f \tag{1.57}$$

which is the solution we had computed with the Tychonov filter in subsection 1.4.

2. Mathematical tools and modelization

Throughout our study, we will use the following classical distributional spaces. $\Omega \subset R^N$ will denote an open bounded set of R^N with regular boundary.

For this section, we refer the reader to [23], and also to [2, 37, 40, 52, 46, 68] for functional anlaysis, to [64, 48, 36] for convex analysis, and to [5, 38, 45] for an introduction to BV functions.

2.1 Minimizing in a Banach space

2.2 Banach spaces

A Banach space is a normed space in which Cauchy sequences have a limit.

Let $(E, |.|)$ be a real Banach space. We denote by $E^{'}$ the topological dual space of E (i.e. the space of linear form continuous on E):

$$E^{'} = \left\{ l : E \to \mathbb{R} \text{ linear such that } |l|_{E'} = \sup_{x>0} \frac{|l(x)|}{|x|} < +\infty \right\} \tag{2.1}$$

If $f \in E'$ and $x \in E$, we note $\langle f, x \rangle_{E',E} = f(x)$.

If $x \in E$, then $J_x : f \mapsto \langle f, x \rangle_{E',E}$ is a continuous linear form on E', i.e. an elemenet of $E^{''}$. Hence $\langle J_x, f \rangle_{E'',E'} = \langle f, x \rangle_{E',E}$ for all $f \in E'$ and $x \in E$. $J : E \to E^{''}$ is a linear isometrie.

We say that E is reflexive if $J(E) = E^{''}$ (in general, J may be non surjective).

2.2.1 Preliminaries

We will use the following classical spaces.

Test functions: $\mathcal{D}(\Omega) = C_c^\infty(\Omega)$ is the set of functions in $C^\infty(\Omega)$ with compact support in Ω. We denote by $\mathcal{D}'(\Omega)$ the dual space of $\mathcal{D}(\Omega)$, i.e. the space of distributions on Ω.

$\mathcal{D}(\bar{\Omega})$ is the set of restriction to Ω of functions in $\mathcal{D}(\mathbb{R}^N) = C_c^\infty(\mathbb{R}^N)$.

Notice that a sequence (v_n) in $\mathcal{D}(\Omega)$ converges to v in $\mathcal{D}(\Omega)$ if the two following conditions are satisfied:

1. There exists a compact subset K in Ω such that support of v_n is embeded in K for all n and support of v is embeded in K.
2. For all multi-index $p \in \mathbb{N}^N$, $D^p v_n \to D^p v$ uniformly on K.

Notice that a sequence (v_n) in $\mathcal{D}'(\Omega)$ converges to v in $\mathcal{D}'(\Omega)$ if as $n \to +\infty$, we have for all $\phi \in \mathcal{D}(\Omega)$:

$$\int_\Omega v_n \phi \to \int_\Omega v \phi \tag{2.2}$$

Radon measure A Radon measure μ is a linear form on $C_c(\Omega)$ such that for each compact $K \subset \Omega$, the restriction of μ to $C_K(\Omega)$ is continuous; that is, for each compact $K \subset \Omega$, there exists $C(K) \geq 0$ such that :

$$\forall v \in C_c(\Omega) \ , \quad \text{with support of } v \text{ embeded in } K \ \ , \ |\mu(v)| \leq C(K) \|v\|_\infty$$

L^p **spaces:** Let $1 \leq p < +\infty$.

$$L^p(\Omega) = \left\{ f : \Omega \to \mathbb{R} \text{ such that } \left(\int_\Omega |f|^p \, dx \right)^{1/p} < +\infty \right\} \tag{2.3}$$

$$L^\infty(\Omega) = \{ f : \Omega \to \mathbb{R} \text{ , f measurable, such that there exists a constant } C \text{ and } |f(x)| \leq C \text{ p.p. on } \Omega \} \tag{2.4}$$

Properties:

1. If $1 \leq p \leq +\infty$, then $L^p(\Omega)$ is a Banach space.
2. If $1 \leq p < +\infty$, then $L^p(\Omega)$ is a separable space (i.e. it has a countable dense subset). But $L^\infty(\Omega)$ is not separable.
3. If $1 \leq p < +\infty$, then the dual space of $L^p(\Omega)$ is $L^q(\Omega)$ with $\frac{1}{p} + \frac{1}{q} = 1$. But $L^1(\Omega)$ is strictly included in the dual space of $L^\infty(\Omega)$.

   Remark: Let $E = C_c(\Omega)$ embeded with the norm $\|u\| = \sup_{x \in \Omega} |u(x)|$. Let us denote $E' = M(\Omega)$ the space of Radon measure on Ω. Then $L^1(\Omega)$ can be identified with a subspace of $M(\Omega)$. Indeed, consider the application $T : L^1(\Omega) \to M(\Omega)$. If $f \in L^1(\Omega)$, then if $u \in C_c(\Omega)$, $u \mapsto \int fu$ is a linear continuous form on $C_c(\Omega)$, so that: $\langle Tf, u \rangle_{E',E} = \int fu$. It is easy to see that T is a linear application from $L^1(\Omega)$ onto $M(\Omega)$, and:

   $$\|Tf\|_{M(\Omega)} = \sup_{u \in C_c(\Omega), \|u\| \leq 1} \int fu = \|f\|_{L^1(\Omega)}$$

4. If $1 < p < +\infty$, then $L^p(\Omega)$ is reflexive.

We have the following density result:

Proposition 2.1. *Ω being an open subset of $\mathbb{R}^N$, then $C_c^\infty(\Omega)$ is dense in $L^p(\Omega)$ for $1 \leq p < \infty$.*

The proof relies on the use of mollifiers.

Theorem 2.1. Lebesgue's theorem
Let (f_n) a sequence in $L^1(\Omega)$ such that:

(i) $f_n(x) \to f(x)$ p.p. on Ω.

(ii) There exists a function g in $L^1(\Omega)$ such that for all n, $|f_n(x)| \leq g(x)$ p.p; on Ω.

Then $f \in L^1(\Omega)$ and $\|f_n - f\|_{L^1} \to 0$.

Theorem 2.2. Fatou's lemma
Let f_n a sequence in $L^1(\Omega)$ such that:

(i) For all n, $f_n(x) \geq 0$ p.p. on Ω.

(ii) $\sup \int_\Omega f_n < +\infty$.

For all x in Ω, we set $f(x) = \lim_{n\to+\infty} \inf f_n(x)$. Then $f \in L^1(\Omega)$, and:

$$\int f \leq \lim_{n\to+\infty} \inf \int f_n \tag{2.5}$$

($\liminf u_n$ *is the smallest cluster point of* u_n*).*

Theorem 2.3. Gauss-Green formula

$$\int_\Omega (\Delta u) v = \int_\Gamma \frac{\partial u}{\partial N} v \, d\sigma - \int_\Omega \nabla u \nabla v \tag{2.6}$$

for all $u \in C^2(\overline{\Omega})$ and for all $v \in C^1(\overline{\Omega})$.

This can be seen as a generalization of the integration by part.

In image processing, we often deal with Neumann boundary conditions, that is $\frac{\partial u}{\partial N} = 0$ on Γ.

Another formulation is the following:

$$\int_\Omega v \operatorname{div} u = \int_\Gamma u.N \, v - \int_\Omega u.\nabla v \tag{2.7}$$

for all $u \in C^1(\overline{\Omega}, \mathbb{R}^N)$ and for all $v \in C^1(\overline{\Omega}, \mathbb{R})$, with N unitary normal outward vector of Γ.

We recall that $\operatorname{div} u = \sum_{i=1}^N \frac{\partial u_i}{\partial x_i}$, and $\Delta u = \operatorname{div} \nabla u = \sum_{i=1}^N \frac{\partial^2 u_i}{\partial x_i^2}$.

In the case of Neumann or Dirichlet boundary conditions, (2.7) reduces to:

$$\int_\Omega u \nabla v = - \int_\Omega v \operatorname{div} u \tag{2.8}$$

In this case, we can define $\operatorname{div} = -\nabla^*$. Indeed, we have:

$$\int_\Omega u \nabla v = \langle u, \nabla v \rangle = \langle \nabla^* u, v \rangle = - \int_\Omega v \operatorname{div} u \tag{2.9}$$

Sobolev spaces: Let $p \in [1, +\infty)$.

$$\begin{aligned} W^{1,p}(\Omega) &= \{ u \in L^p(\Omega) \ / \ \text{there exist } g_1, \ldots, g_N \text{ in } L^p(\Omega) \text{ such that} \\ & \left. \int_\Omega u \frac{\partial \phi}{\partial x_i} = - \int_\Omega g_i \phi \ \forall \phi \in C_c^\infty(\Omega) \ , \ \forall i = 1, \ldots, N \right\} \end{aligned}$$

We can denote by $\frac{\partial u}{\partial x_i} = g_i$ and $\nabla u = \left(\frac{\partial u}{\partial x_1}, \ldots, \frac{\partial u}{\partial x_N} \right)$.

Equivalently, we say that u belongs to $W^{1,p}(\Omega)$ if u is in $L^p(\Omega)$ and if u has a derivative in the distibutional sens also in $L^p(\Omega)$.

This is a Banach space endowed with the norm:

$$\|u\|_{W^{1,p}(\Omega)} = \left(\|u\|_{L^p(\Omega)} + \sum_{i=1}^N \left\| \frac{\partial u}{\partial x_i} \right\|_{L^p(\Omega)} \right)^{\frac{1}{p}} \tag{2.10}$$

We denote by $H^1(\Omega) = W^{1,2}(\Omega)$. This is a Hilbert space embed with the inner product:

$$\langle u, v \rangle_{H^1} = \langle u, v \rangle_{L^2} + \langle \nabla u, \nabla v \rangle_{L^2 \times L^2}$$

and the associated norm is $\|u\|_{H^1}^2 = \|u\|_{L^2}^2 + \|\nabla u\|_{L^2 \times L^2}^2$.

$W_0^{1,p}(\Omega)$ denotes the space of functions in $W^{1,p}(\Omega)$ with compact support in Ω (it is the closure of $C_c^1(\Omega)$ in $W^{1,p}(\Omega)$).

Let $q = \frac{p}{p-1}$ (so that $\frac{1}{p} + \frac{1}{q} = 1$). We denote by $W^{-1,q}(\Omega)$ the dual space of $W_0^{1,p}(\Omega)$.

Properties: If $1 < p < +\infty$, then $W^{1,p}(\Omega)$ is reflexive.
If $1 \leq p < +\infty$, then $W^{1,p}(\Omega)$ is separable.

Proposition 2.2. *Ω being an open subset of $\mathbb{R}^N$, then $C_c^\infty(\Omega)$ is dense in $W^{1,p}(\Omega)$ for $1 \leq p < \infty$.*

Theorem 2.4. Poincaré inequality
Let Ω a bounded open set. Let $1 \leq p < \infty$. Then there exists $C > 0$ (depending on Ω and p) such that, for all u in $W_0^{1,p}(\Omega)$:

$$\|u\|_{L^p} \leq C\|\nabla u\|_{L^p} \tag{2.11}$$

Theorem 2.5. Poincaré-Wirtinger inequality *Let Ω be open, bounded, connected, with a C^1 boundary. Then for all u in $W^{1,p}(\Omega)$, we have:*

$$\left\| u - \frac{1}{\Omega}\int_\Omega u\,dx \right\|_{L^p(\Omega)} \leq C\|\nabla u\|_{L^p} \tag{2.12}$$

We have the following Sobolev injections:

Theorem 2.6. *Ω bounded open set with C^1 boundary. We have:*

- *If $p < N$, then $W^{1,p}(\Omega) \subset L^q(\Omega)$ for all $q \in [1, p^*)$ where $\frac{1}{p^*} = \frac{1}{p} - \frac{1}{N}$.*
- *If $p = N$, then $W^{1,p}(\Omega) \subset L^q(\Omega)$ for all $q \in [1, +\infty)$.*
- *If $p > N$, then $W^{1,p}(\Omega) \subset C(\overline{\Omega})$.*

with compact injections (in particular, a compact injection from X to Y turns a bounded sequence in X into a compact sequence in Y).
In particular, one always have: $W^{1,p}(\Omega) \subset L^p(\Omega)$ with compact injection for all p.

We recall that a linear operator $L : E \to F$ is said to be compact if $L(B_E)$ is relatively compact in F (i.e. its closure is compact), B_E being the unitary ball in E.

Particular case of dimension 1: Here we consider the case when $\Omega = I = (a, b)$, a and b finite or not. We have the following result (roughly speaking, functions in $W^{1,p}(I)$ are primitives of functions in $L^p(I)$.

Proposition 2.3. *Let $u \in W^{1,p}(I)$. Then there exists $\tilde{u} \in C(\bar{I})$ such that: $u = \tilde{u}$ a.e. in I, and:*

$$\tilde{u}(x) - \tilde{u}(y) = \int_y^x u'(t)\,dt \tag{2.13}$$

for all x and y in $\bar{I}$.

2.2.2 Topologies in Banach spaces

Let $(E, |.|)$ be a real Banach space. We denote by E' the topological dual space of E (i.e. the space of linear form continuous on E):

$$E' = \left\{ l : E \to \mathbb{R} \text{ linear such that } |l|_{E'} = \sup_{x>0} \frac{|l(x)|}{|x|} < +\infty \right\} \tag{2.14}$$

E can be endowed with two topologies:

(i) The strong topology:

$$x_n \to x \text{ if } |x_n - x|_E \to 0 \text{ (as } n \to +\infty) \tag{2.15}$$

(ii) The weak topology:

$$x_n \rightharpoonup x \text{ if } l(x_n) \to l(x) \text{ (as } n \to +\infty) \ \forall l \in E' \tag{2.16}$$

Remark: Weak convergence does not imply strong convergence.

Consider for instance: $\Omega = (0,1)$, $f_n(x) = \sin(2\pi n x)$, and $L^2(\Omega)$. We have $f_n \rightharpoonup 0$ in $L^2(0,1)$ (integration by part with $\phi \in C^1(0,1)$, but $\|f_n\|^2_{L^2(0,1)} = \frac{1}{2}$ (by using $\sin^2 x = \frac{1-\cos 2x}{2}$).

More precisely, to show that $f_n \rightharpoonup 0$ in $L^2(0,1)$, we first take $\phi \in C^1(0,1)$. We have

$$\int_0^1 f_n(x)\phi(x)\,dx = \left[\frac{\cos(2\pi n x)}{2\pi n}\right]_0^1 + \frac{1}{2\pi}\int_0^1 \cos(2\pi n x)\phi'(x)\,dx \tag{2.17}$$

Hence $\langle f_n, \phi\rangle \to 0$ as $n \to +\infty$. By density of $C^1(0,1)$ in $L^2(0,1)$, we get that $\langle f_n, \phi\rangle \to 0$ for all $\phi \in L^2(\Omega)$. We thus deduce that $f_n \rightharpoonup 0$ in $L^2(0,1)$ (since $L^{2'} = L^2$ thanks to Riesz theorem).

Now we observe that

$$\|f_n\|^2_{L^2(0,1)} = \int_0^1 \sin(2\pi n x)\,dx = \int_0^1 \frac{1-\cos(4\pi n x)}{2}\,dx = \frac{1}{2} \tag{2.18}$$

and thus f_n cannot go to 0 in $L^2(0,1)$ strong.

The dual E' can be endowed with three topologies:

(i) The strong topology:

$$l_n \to l \text{ if } |l_n - l|_{E'} \to 0 \text{ (as } n \to +\infty) \tag{2.19}$$

(ii) The weak topology:

$$l_n \rightharpoonup l \text{ if } z(l_n) \to z(l) \text{ (as } n \to +\infty) \ \forall z \in \left(E'\right)', \text{ the bidual of } E. \tag{2.20}$$

(iii) The weak-* topology:

$$l_n \rightharpoonup_* l \text{ if } l_n(x) \to l(x) \text{ (as } n \to +\infty) \ \forall x \in E \tag{2.21}$$

Examples: If $E = L^p(\Omega)$, if $1 < p < +\infty$, E is *reflexive*, i.e. $\left(E'\right)' = E$ and *separable*. The dual of E is $L^{p'}(\Omega)$ with $\frac{1}{p} + \frac{1}{p'} = 1$.

If $E = L^1(\Omega)$, E is nonreflexive and $E' = L^\infty(\Omega)$. The bidual $\left(E'\right)'$ is a very complicated space.

Main property *(weak sequential compactness):*

Proposition 2.4.

- *Let E be a* reflexive *Banach space, $K > 0$, and $x_n \in E$ a sequence such that $|x_n|_E \le K$. Then there exists $x \in E$ and a subsequence x_{n_j} of x_n such that $x_{n_j} \rightharpoonup x$ as $n \to +\infty$.*
- *Let E be a* separable *Banach space, $K > 0$, and $l_n \in E'$ a sequence such that $|l_n|_{E'} \le K$. Then there exists $l \in E'$ and a subsequence l_{n_j} of l_n such that $l_{n_j} \rightharpoonup_* l$ as $n \to +\infty$.*

The first point can be used for instance with $E = L^p(\Omega)$, with $1 < p < +\infty$. The second point can be used for instance with $E' = L^\infty(\Omega)$ (and thus $E = L^1(\Omega)$)

2.2.3 Convexity and lower semicontinuity

Let E be a banach space, and $F : E \to \mathbb{R}$. Let $(E, |.|)$ a real Banach space, and $F : E \to \mathbb{R}$.

Definition 2.1.

(i) F is convex if

$$F(\lambda x + (1-\lambda)y) \le \lambda F(x) + (1-\lambda)F(y) \tag{2.22}$$

for all x, y in E and $\lambda \in [0,1]$.

(ii) F is lower semi-continuous (l.s.c.) if

$$\liminf_{x_n \to x} F(x_n) \ge F(x) \tag{2.23}$$

Equivalently, F is l.s.c if for all λ in $\mathbb{R}$, the set $\{x \in E; F(x) \le \lambda\}$ is closed.

Proposition 2.5.

1. *If F_1 and F_2 are lsc, then $F_1 + F_2$ is also lsc.*
2. *If F_i are lsc, then $\sup_i F_i$ is also lsc.*
3. *If F_1 and F_2 are convex, then $F_1 + F_2$ is also convex.*
4. *If F_i are convex, then $\sup_i F_i$ is also convex.*

Proposition 2.6. *F C^1 is convex on E iff*

$$F(x+y) \ge F(x) + \langle \nabla F(x), y \rangle\rangle \tag{2.24}$$

In particular, if F is convex and $\nabla F(x) = 0$, then x is a minimizer of F. Notice also that the above result remains true when assuming that F is Gateau differentiable.

Proposition 2.7. *F C^2 is convex on E iff $\nabla^2 F$ is non negative on E.*

Proposition 2.8. *Let $F : E \to \mathbb{R}$ be convex. Then F is weakly l.s.c. if and only if F is strongly l.s.c.*

In particular, if $F : E \to \mathbb{R}$ convex strongly l.s.c., if $x_n \rightharpoonup x$, then

$$F(x) \leq \liminf F(x_n) \tag{2.25}$$

Notice also that if $x_n \rightharpoonup x$, then

$$|x|_E \leq \liminf |x_n|_E \tag{2.26}$$

Proposition 2.9. *Let E and F be two Banach spaces. If L is a continuous linear operator from E to F, then L is strongly continuous if and only if L is weakly continuous.*

Minimization: ***the Direct method of calculus of variations***

Consider the following minimization problem

$$\inf_{x \in E} F(x) \tag{2.27}$$

(a) One constructs a minimizing sequence $x_n \in E$, i.e. a sequence satisfying

$$\lim_{n \to +\infty} F(x_n) = \inf_{x \in E} F(x) \tag{2.28}$$

(b) If F is *coercive* (i.e. $\lim_{|x| \to +\infty} F(x) = +\infty$), one can obtain a uniform bound: $|x_n| \leq K$.

(c) If E is *reflexive* (i.e. $E'' = E$), then we deduce the existence of a subsequence x_{n_j} and of $x_0 \in E$ such that $x_{n_j} \rightharpoonup x_0$.

(d) If F is lower semi-continuous, we deduce that:

$$\inf_{x \in E} F(x) = \liminf F(x_n) \geq F(x_0) \tag{2.29}$$

which obviously implies that:

$$F(x_0) = \min_{x \in E} F(x) \tag{2.30}$$

Remark that convexity is used to obtain l.s.c; while coercivity is related to compactness.

Remark: The above method can be extended at once to the case:

$$\inf_{x \in C} F(x) \tag{2.31}$$

where C is a nonempty closed convex set of E (we remind the reader that a convex set is weakly closed iff it is strongly closed).

Remark: case when F is an integral functional Let $f : \Omega \times \mathbb{R} \times \mathbb{R}^2 \to \mathbb{R}$ (with $\Omega \subset \mathbb{R}^2$) For $u \in W^{1,p}(\Omega)$, we consider the functional:

$$F(u) = \int_\Omega f(x, u(x), Du(x))\, dx \tag{2.32}$$

If f is l.s.c., convex (with respect to u and ξ), and coercive, then so is F. Moreover, if f satisfies a growth condition $0 \leq f(x, u, \xi) \leq a(x, |u|, |\xi|)$ with a increasing with respect to $|u|$ and $|\xi|$., then we have: *F is weakly l.s.c. on $W^{1,}(\Omega)$ iff f is convex in ξ.*

Examples Let $\Omega = (0,1) = I$.

We remind the reader that we have the following result:

Let $u \in W^{1,p}(I)$. Then there exists $\tilde{u} \in C(\bar{I})$ such that: $u = \tilde{u}$ a.e. in I, and:

$$\tilde{u}(x) - \tilde{u}(y) = \int_y^x u'(t)\,dt \tag{2.33}$$

for all x and y in $\bar{I}$.

(a) *Weiertrass.* Let us consider the problem when $f(x,u,\xi) = x\xi^2$:

$$m = \inf\left\{\int_0^1 x(u'(x))^2\,dx \text{ with } u(0) = 1 \text{ and } u(1) = 0\right\} \tag{2.34}$$

It is possible to show that $m = 0$ but that this problem does not have any solution. The function f is convex, but the $W^{1,2}$ *coercivity* with respect to u is not satisfied because the integrand $f(x,\xi) = x\xi^2$ vanishes at $x = 0$.

To show that $m = 0$, one can consider the minimizing sequence:

$$u_n(x) = \begin{cases} 1 & \text{if } x \in \left(0, \frac{1}{n}\right) \\ -\frac{\log x}{\log n} & \text{if } x \in \left(\frac{1}{n}, 1\right) \end{cases} \tag{2.35}$$

We have $u_n \in W^{1,\infty}(0,1) \subset W^{1,2}(\Omega)$, and

$$F(u_n) = \int_0^1 x(u_n'(x))^2\,dx = \frac{1}{\log n} \to 0 \tag{2.36}$$

So $m = 0$. Now, if a minimizer $\hat{u}$ exists, then $\hat{u}' = 0$ a.e. in $(0,1)$, which is clearly not compatible with the boundary conditions.

(b) *Minimal surface.* Let $f(x,u,\xi) = \sqrt{x^2 + \xi^2}$. We thus have: $F(u) \geq \frac{1}{2}\|u\|_{W^{1,1}}$ (straightforward consequence of the fact that $\sqrt{a^2+b^2} \geq \frac{1}{2}(||a + |b|)$ since $a^2 + b^2 - (a+b)^2/4 = (a^2 + b^2 + 2(a-b)^2)/4 \geq 0$. The associated functional F is convex and coercive on the *non reflexive* Banach space $W^{1,1}$. Let us set:

$$m = \inf\left\{\int_0^1 \sqrt{u^2 + (u')^2}\,dx \text{ with } u(0) = 0 \text{ and } u(1) = 1\right\} \tag{2.37}$$

It is possible to show that $m = 1$ but that there is no solution.

Let us prove $m = 1$. First, we observe that:

$$F(u) = \int_0^1 \sqrt{u^2 + (u')^2}\,dx \geq \int_0^1 |u'|\,dx \geq \int_0^1 u'\,dx = 1 \tag{2.38}$$

So we see that $m \geq 1$. Now, let us consider the sequence:

$$u_n(x) = \begin{cases} 0 & \text{if } x \in \left(0, 1 - \frac{1}{n}\right) \\ 1 + n(x-1) & \text{if } x \in \left(1 - \frac{1}{n}, 1\right) \end{cases} \tag{2.39}$$

It is easy to check that $F(u_n) \to 1$ as $n \to +\infty$. This implies $m = 1$.

Now, if a minimizer $\hat{u}$ exists, then we should have:

$$1 = F(\hat{u}) = \int_0^1 \sqrt{\hat{u}^2 + (\hat{u}')^2}\, dx \geq \int_0^1 |\hat{u}'|\, dx \geq \int_0^1 \hat{u}'\, dx = 1 \tag{2.40}$$

which implies $\hat{u} = 0$, which does not satisfy the boundary conditions.

(c) *Bolza* Let $f(x, u, \xi) = u^2 + (\xi^2 - 1)^2$. The Bolza problem is:

$$m = \inf \left\{ \int_0^1 \left(u^2 + (1 - (u')^2)^2\right) dx \text{ with } u(0) = u(1) = 0 \right\} \tag{2.41}$$

and u in $W^{1,4}(\Omega)$. The functional is clearly *nonconvex*, and it is possible to show that $m = 0$ and that there is no solution.

Indeed, we have $\inf_u F(u) \geq 0$, where $F(u) = \int_0^1 f(x, u(x), u'(x))\, dx$. Now, if $n \geq 1$, if $0 \leq k \leq n_1$, we can choose:

$$u_n(x) = \left\{ \begin{array}{ll} x - \frac{k}{n} & \text{if } x \in \left(\frac{2k}{2n}, \frac{2k+1}{2n}\right) \\ -x + \frac{k+1}{n} & \text{if } x \in \left(\frac{2k+1}{2n}, \frac{2k+2}{2n}\right) \end{array} \right. \tag{2.42}$$

We have u_n in $W^{1,\infty}(0,1) \subset W^{1,4}(0,1)$, $0 \leq u_n(x) \leq \frac{1}{2n}$ for $x \in (0,1)$, $|u'{}_n(x)| = 1$ a.e. in $(0,1)$, $u_n(0) = u_n(1) = 0$.

Therefore $F(u_n) \leq \frac{1}{4n^2}$, and we thus deduce that $m = 0$.

However, there exists no $\hat{u}$ in $W^{1,4}(0,1)$ such that $F(\hat{u}) = 0$ (and thus such that $|\hat{u}'| = 1$ a.e. and $u = 0$ a.e. and $u(0) = u(1) = 0$).

Characterization of a minimizer: *(Euler-Lagrange equation)*

Definition 2.2. *Gâteaux derivative*

$$F^{'}(u; \nu) = \lim_{\lambda \to 0^+} \frac{F(u + \lambda\nu) - F(u)}{\lambda} \tag{2.43}$$

is called the directional derivative of F at u in the direction ν if this limit exists. Moreover, if there exists $\bar{u} \in E^{'}$ such that $F^{'}(u; \nu) = \bar{u}(\nu) = \langle \bar{u}, \nu \rangle$ for all $\nu \in E$, we say that F is Gâteaux differentiable at u and we write $F^{'}(u) = \bar{u}$.

Notice that F is said Frechet differentiable on a Banach space if there exists some linear continuous operator A_x such that

$$\lim_{h \to 0} \frac{\|F(x + h) - F(x) - Ax(h)\|}{\|h\|} = 0 \tag{2.44}$$

On an open subset, if F is Gateau differentiable, then F is Frechet differentiable if the derivative is linear and continuous, and the Gateau derivative is a continuous map.

Application: If F is Gâteaux differentiable and if problem $\inf_{x \in E} F(x)$ has a solution u, then necessarily we have the optimiality condition:

$$F^{'}(u) = 0 \tag{2.45}$$

(the controverse is true if F is convex). This last equation is called *Euler-Lagrange* equation.

Indeed, if $F(u) = \min_x F(x)$, then $F(u + \lambda v) - F(u) \geq 0$, i.e. $\langle v, F'(u) \rangle \geq 0$. But if we consider $-v$, we also get: $F(u + \lambda(-v)) - F(u) \geq 0$, and thus $\langle -v, F'(u) \rangle \geq 0$. Hence $\langle v, F'(u) \rangle = 0$ for all v. We thus deduce that $F'(u) = 0$.

2.2.4 Convex analysis

2.3 Subgradient of a convex function

Definition :

Let $F : E \to \mathbb{R}$ a convex proper function. The subgradient of F at position x is defined as:

$$\partial F(u) = \left\{ v \in E^{'} \text{ such that } \forall w \in E \text{ we have } F(w) \geq F(u) + \langle v, w - u \rangle \right\} \tag{2.46}$$

Equivalently, F is said to be subdifferentiable in u if F has a continuous affine minorante, exact in u. The slope of such a minorante is called a subgradient of F in u., and the set of all subgradients in u is called the subdifferential of F in u.

It can be seen as a generalization of the concept of derivative for convex function.

Proposition 2.10. *x is a solution of the problem*

$$\inf_E F \tag{2.47}$$

if and only if $0 \in \partial F(x)$.

This is another version of the Euler-Lagrange equation.

Roughly speaking, F convex is Gateau differentiable in u (plus some technical assumptions) iff $\partial F(u) = \{F'(u)\}$.

Monotone operator :

Proposition 2.11. *F a convex proper function on E. Then ∂F is a monotone operator, i.e.*

$$\langle \partial F(u_1) - \partial F(u_2), u_1 - u_2 \rangle \geq 0 \tag{2.48}$$

Proof :

Let v_i in $\partial F(u_i)$. We have:

$$F(u_2) \geq F(u_1) + \langle v_1, u_2 - u_1 \rangle$$

and:

$$F(u_1) \geq F(u_2) + \langle v_2, u_1 - u_2 \rangle$$

hence:

$$0 \geq \langle v_2 - v_1, u_1 - u_2 \rangle$$

■

Proposition 2.12. *F Gateaux differentiable on E. Then F is convex iff F' is monotone.*

Subdifferential calculus :

Proposition 2.13.

- *If* $\lambda > 0$ *then:* $\partial(\lambda F)(u) = \lambda \partial F(u)$.
- F_1 *and* F_2 *two convex proper functions. Then:*

$$\partial F_1(u) + \partial F_2(u) \subset \partial(F_1 + F_2)(u) \tag{2.49}$$

The reverse inclusion does not always hold. A sufficient condition is the following:

Proposition 2.14. *Let* F_1 *and* F_2 *two convex proper functions. If there exists* $\bar{u}$ *in Dom* F_1 $\bigcap$ *Dom* F_2 *where* F_1 *is continuous, then:*

$$\partial F_1(u) + \partial F_2(u) = \partial(F_1 + F_2)(u) \tag{2.50}$$

In particular, if F_1 is differentiable, then:

$$\partial(F_1 + F_2)(u) = \nabla F_1(u) + \partial F_2(u) \tag{2.51}$$

2.4 Legendre-Fenchel transform:

Definition :
Let $F : E \to \mathbb{R}$. We define $F^* : E' \to \mathbb{R}$ by:

$$F^*(v) = \sup_{u \in E} \left(\langle v, u \rangle - F(u) \right) \tag{2.52}$$

It is easy to see that F is a convex lsc function (sup of convex lsc functions).

Properties

- $F^*(0) = -\inf_u F(u)$
- If $F \leq G$, then $F^* \geq G^*$.
- $\left(\inf_{i \in I} F_i\right)^* = \sup_{i \in I} F_i^*$
- $\left(\sup_{i \in I} F_i\right)^* \leq \inf_{i \in I} F_i^*$

Proposition 2.15. $v \in \partial F(u)$ *iff:*

$$F(u) + F^*(v) = \langle u, v \rangle \tag{2.53}$$

Proof: $v \in \partial F(u)$ means that for all w:

$$F(w) \geq F(u) + \langle v, w - u \rangle \tag{2.54}$$

i.e. for all w:

$$\langle v, u \rangle - F(u) \geq \langle v, w \rangle - F(w) \tag{2.55}$$

and thus: $\langle v,u\rangle - F(u) \geq F^*(v)$. But by definition, $F^*(v) \geq \langle v,u\rangle - F(u)$ Hence:

$$F^*(v) = \langle v,u\rangle - F(u) \tag{2.56}$$

Theorem 2.7. *If F is convex l.s.c., and $F \neq +\infty$, then $F^{**} = F$.*

In particular, one always has $F^{***} = F^*$. Remark that in general one always has $F^{**} \leq F$.

Theorem 2.8. *If F is convex l.s.c., and $F \neq +\infty$, then*

$$v \in \partial F(u) \quad \textit{iff} \quad u \in \partial F^*(v) \tag{2.57}$$

Indeed, if $v \in \partial F(u)$, then:

$$F(u) + F^*(v) = \langle u,v\rangle \tag{2.58}$$

And since $F^{**} = F$, we have:

$$F^{**}(u) + F^*(v) = \langle u,v\rangle \tag{2.59}$$

which means that $u \in \partial F^*(v)$.

Theorem 2.9. *(Fenchel-Rockafellar)*
Let F and G two convex functions. Assume that $\exists x_0 \in E$ such that $F(x_0) < +\infty$, $G(x_0) < +\infty$, and F continuous in x_0. Then:

$$\inf_{x\in E}\{F(x)+G(x)\} = \sup_{f\in E'}\{-F^*(-f) - G^*(f)\} = \max_{f\in E'}\{-F^*(-f) - G^*(f)\} \tag{2.60}$$

Proposition 2.16. *Let $K \subset E$ a closed and non empty convex set. We call* indicator function *of K:*

$$\chi_K(u) = \begin{cases} 0 & \textit{if } u \in K \\ +\infty & \textit{otherwise} \end{cases} \tag{2.61}$$

χ_K is convex, l.s.c., and $\chi_K \neq +\infty$.
The conjugate function χ_K^* is called support function *of K.*

$$\chi_K^*(v) = \sup_{u\in K}\langle v,u\rangle \tag{2.62}$$

Remark that then the conjugate function of a support function is an indicator function.

Proposition 2.17. *Assume $E = L^2$. Let $F(x) = \frac{1}{2}\|x\|_2^2$. Then $F^* = F$.*

More generally, if $E = L^p$ with $1 < p < +\infty$, if

$$F(u) = \frac{1}{p}\|u\|_p^p \tag{2.63}$$

then we have:

$$F^*(v) = \frac{1}{q}\|u\|_q^q \tag{2.64}$$

with $\frac{1}{p} + \frac{1}{q} = 1$.

Important example :

Let $E = L^2(\Omega)$, with $\Omega \subset \mathbb{R}^2$. We consider the non empty closed convex set:

$$K = \left\{ u \in L^2(\Omega) \,/\, u = \operatorname{div} \xi \,,\, \xi \in C_c^\infty(\Omega, \mathbb{R}^2) \,,\, \|\xi\|_\infty \leq 1 \right\} \tag{2.65}$$

After what we said above, the *indicator function* of K, χ_K, is convex, l.s.c., and proper. Now, consider its Legendre-Fenchel transform:

$$\chi_K^*(v) = \sup_{u \in K} \langle v, u \rangle = \int_\Omega |Dv| = J(v) \tag{2.66}$$

We recognize the definition of the total variation. Moreover, since $\chi_K^{**} = \chi_K$, we get that $J^*(v) = \chi_K(v)$.

2.5 The space of funtions with bounded variation

For a full introduction to $BV(\Omega)$, we refer the reader to [5].

2.5.1 Introduction

We first give a few definitions.

Definition 2.3. Let X be a non empty set, and let $\mathcal{I}$ be a collection of subsets of X.

- $\mathcal{I}$ is an algebra if $\emptyset \in \mathcal{I}$, and $E_1 \bigcup E_2 \in \mathcal{I}$, $X \backslash E_1 \in \mathcal{I}$, whenever E_1, $E_2 \in \mathcal{I}$.
- An algebra $\mathcal{I}$ is a σ-algebra if for any sequences $(E_h) \subset \mathcal{I}$, their union $\bigcup_h E_h$ belongs to $\mathcal{I}$. σ-algebra are closed under countable intersections.
- If (X, τ) is a topological space, we note $B(X)$ the σ-algebra generated by the open subsets of X.

Definition 2.4.

- Let $\mu : \mathcal{I} \rightarrow [0, +\infty]$ with $\mathcal{I}$ σ-algebra. μ is said to be a positive measure if $\mu(\emptyset) = 0$ and μ σ additive on $\mathcal{I}$, i.e. for any sequences (E_h) of pairwise disjoint elements of $\mathcal{I}$:

$$\mu\left(\bigcup_{h=0}^{+\infty} E_h\right) = \sum_{h=0}^{+\infty} \mu(E_h) \tag{2.67}$$

- μ is said bounded if $\mu(X) < +\infty$.
- μ is said to be a signed or real measure if $\mu : \mathcal{I} \rightarrow \mathbb{R}$.
- μ is said to be a vector-valued measure if $\mu : \mathcal{I} \rightarrow \mathbb{R}^m$.

Definition 2.5. If $X = \mathbb{R}^N$, μ is called Radon measure if $\mu(K) < +\infty$ for all compact K of X.

Definition 2.6. If μ is a measure, we define its total variation $|\mu|$ for every $E \in \mathcal{I}$ as follows:

$$|\mu|(E) = \sup \left\{ \sum_{h=0}^{+\infty} |\mu(E_h)| \,;\, E_h \in \mathcal{I} \text{ pairwise disjoint, } E = \bigcup_{h=0}^{+\infty} E_h \right\} \tag{2.68}$$

The $|\mu|$ is a bounded measure.

Definition 2.7. Let μ be a positive measure. $A \subset X$ is said μ negligible if there exists $E \in \mathcal{I}$ such that $A \subset E$ and $\mu(E) = 0$.

Definition 2.8. Let μ be a positive measure, and let ν be a measure. ν is said absolutely continuous with respect to μ and we write $\nu << \mu$ if $\mu(E) = 0 \Longrightarrow \nu(E) = 0$.

μ and ν are said mutually singular and we write $\mu \perp \nu$ if there exists a set E such that $\mu(\mathbb{R}^N \backslash E) = \nu(E) = 0$.

Theorem 2.10. *Lebesgue theorem:*

Let μ be a positive bounded measure on $(\mathbb{R}^N, B(\mathbb{R}^N))$ (typically the Lebesgue measure)*, and ν a measure on $(\mathbb{R}^N, B(\mathbb{R}^N))$. Then there exists a unique pair of measures ν_{ac} and ν_s such that:*

$$\nu = \nu_{ac} + \nu_s \ , \ \ \nu_{ac} << \mu \ , \ \ \nu_s \perp \mu \tag{2.69}$$

2.5.2 Definition

Definition 2.9. $BV(\Omega)$ is the subspace of functions $u \in L^1(\Omega)$ such that the following quantity is finite:

$$\int_\Omega |Du| = J(u) = \sup \left\{ \int_\Omega u(x) \mathrm{div}\,(\phi(x)) dx / \phi \in C_c^\infty(\Omega, \mathbb{R}^N), \|\phi\|_{L^\infty(\Omega,\mathbb{R}^N)} \leq 1 \right\} \tag{2.70}$$

$BV(\Omega)$ endowed with the norm

$$\|u\|_{BV} = \|u\|_{L^1} + J(u) \tag{2.71}$$

is a Banach space.

If $u \in BV(\Omega)$, the distributional derivative Du is a bounded Radon measure (consequence of the Riesz representation theorem) and (2.70) corresponds to the total variation $|Du|(\Omega)$, i.e. $J(u) = \int_\Omega |Du|$.

Examples:

- If $u \in C^1(\Omega)$, then $\int_\Omega |Du| = \int_\Omega |\nabla u|$. It is a straightforward consequence of the Gauss-Green formula: $\int_\Omega u \mathrm{div}\,(\phi) = \int_\Omega \nabla u.\phi$.
- Let u be defined in $(-1, +1)$ by $u(x) = -1$ if $-1 \leq x < 0$ and $u(x) = +1$ if $0 < x \leq 1$. We have $\int_\Omega u \mathrm{div}\, \phi = \int_{-1}^1 u\phi' = \int_{-1}^0 \phi' + \int_0^1 \phi' = 2\phi(0)$. Then $Du = 2\delta_0$ and $\int_\Omega |Du| = 2$. In fact, $Du = 0\, dx + = 2\delta_0$. Notice that u dos not belong to $W^{1,1}$ since the Dirac mass δ_0 is not in L^1.
- If $A \subset \Omega$, if $u = \mathbf{1}_A$ the characteristic function of the set A, then $\int_\Omega |Du| = \mathrm{Per}_\Omega(A)$ which coincides with the clasical perimeter of A if the boundary of A is smooth (i.e. the lenght if $N = 2$ or the surface if $N = 3$).

  Notice that $\int_\Omega \mathbf{1}_A \mathrm{div}\, \phi = \int_{\partial A} \phi.N$ with N outer unit normal along ∂A.

  See [45] page 4 for more details.

A function belonging to BV may have jumps along curves (in dimension 2; more generally, along surfaces of codimension $N - 1$).

2.5.3 Properties

• *Lower semi-continuity:* Let u_j in $BV(\Omega)$ and $u_j \to_{L^1(\Omega)} u$. Then $\int_\Omega |Du| \leq \lim_{j\to+\infty} \inf \int_\Omega |Du_j|$.
• The strong topology of $BV(\Omega)$ does not have good compactness properties. Classically, in $BV(\Omega)$, one works with the *weak -* topology on* $BV(\Omega)$, defined as:

$$u_j \rightharpoonup_{BV-w*} u \Leftrightarrow u_j \to_{L^1(\Omega)} u \text{ and } Du_j \rightharpoonup_M Du \tag{2.72}$$

where $Du_j \rightharpoonup_M Du$ is a convergence in the sens of measure, i.e. $\langle Du_j, \phi\rangle \to \langle Du, \phi\rangle$ for all ϕ in $(C_c^\infty(\Omega))^2$.

Equipped with this topology, $BV(\Omega)$ has some interesting compactness properties.
• *Compactness:*

If (u_n) is a bounded sequence in $BV(\Omega)$, then up to a subsequence, there exists $u \in BV(\Omega)$ such that: $u_n \to u$ in $L^1(\Omega)$ strong, and $Du_n \rightharpoonup_M Du$.

Let us set $N^* = \frac{N}{N-1}$ ($N^* = +\infty$ if $N = 1$). For $\Omega \subset R^N$, if $1 \leq p \leq N^*$, we have:

$$BV(\Omega) \subset L^p(\Omega) \tag{2.73}$$

Moreover, for $1 \leq p < N^*$, this embedding is compact.

Notice that $N^* = 2$ in the case when $N = 2$.
• If $N = 2$, since $BV(\Omega) \subset L^2(\Omega)$, we can extend the functional J (which we still denote by J) over $L^2(\Omega)$:

$$J(u) = \begin{cases} \int_\Omega |Du| \text{ if } u \in BV(\Omega) \\ +\infty \text{ if } u \in L^2(\Omega) \backslash BV(\Omega) \end{cases} \tag{2.74}$$

We can then define the subdifferential ∂J of J [64]: $v \in \partial J(u)$ iff for all $w \in L^2(\Omega)$, we have $J(u + w) \geq J(u) + \langle v, w\rangle_{L^2(\Omega)}$ where $\langle ., .\rangle_{L^2(\Omega)}$ denotes the usual inner product in $L^2(\Omega)$.
• *Approximation by smooth functions*: If u belongs to $BV(\Omega)$, then there exits a sequence u_n in $C^\infty(\Omega) \bigcap BV(\Omega)$ such that $u_n \to u$ in $L^1(\Omega)$ and $J(u_n) \to J(u)$ as $n \to +\infty$.

The notion of *strict convergence* is useful to prove several identities in BV by smoothing arguments.

Definition 2.10. [Strict convergence] Let u in BV(Ω), and a sequence (u_n) in BV(Ω). Then we say that (u_n) strictly converges in BV(Ω) to u if (u_n) converges to u in $L^1(\Omega)$ and the variations $|Du_n|(\Omega)$ converge to $|Du|(\Omega)$ as $n \to +\infty$.

Notice that strict convergence implies weak-$*$ convergence but the converse is false in general.
• *Poincaré-Wirtinger inequality*

Proposition 2.18. *Let Ω be open, bounded, conneceted, with a C^1 boundary. Then for all u in $BV(\Omega)$, we have:*

$$\left\| u - \frac{1}{|\Omega|} \int_\Omega u\, dx \right\|_{L^p(\Omega)} \leq C \int_\Omega |Du| \tag{2.75}$$

for $1 \leq p \leq N/(N-1)$ (i.e. $1 \leq p \leq 2$ when $N = 2$).

2.5.4 Decomposability of $BV(\Omega)$:

Hausdorff measure :

Definition 2.11. Let $k \in [0, +\infty]$, and $A \subset \mathbb{R}^N$. The k dimensional Hausdorff measure of A is given by

$$\mathcal{H}^k(A) = \lim_{\delta \to 0} \mathcal{H}^k_\delta(A) \tag{2.76}$$

where for $0 < \delta \leq \infty$:

$$\mathcal{H}^k_\delta(A) = \frac{w_k}{2^k} \inf \left\{ \sum_{i \in I} |diam(A_i)|^k, diam(A_i) \leq \delta, A \subset \bigcup_{i \in I} A_i \right\} \tag{2.77}$$

for finite or countable covers $(A_i)_{i \in I}$, diam(A_i) stands for the diameter of A_i, w_k is a normalization factor equal to $\pi^{k/2}\Gamma(1 + k/2)$, where $\Gamma(t) = \int_0^{+\infty} s^{t-1} e^{-s} \, ds$ is the gamma function (w_k coincides with the Lebesgue measure of the unit ball of $\mathbb{R}^k$ if $k \geq 1$ is an integer).

We define the Hausdorff measure of A by:

$$\inf \left\{ k \geq 0 \, ; \, \mathcal{H}^k(A) = 0 \right\} \tag{2.78}$$

$\mathcal{H}^k$ is a measure on $\mathbb{R}^N$.

$\mathcal{H}^N$ coincides with the Lebesgue measure dx, and for $1 \leq k \leq N$, k integer, $\mathcal{H}^k$ is the classical k dimensional area of A if Â is a C^1 k dimensional manifold embedded in $\mathbb{R}^N$. Moreover, if $k > k' \geq 0$, then $\mathcal{H}^k(A) > 0 \Longrightarrow \mathcal{H}^{k'}(A) = +\infty$.

Consequence of Lebesgue theorem: If u in $BV(\Omega)$, then (Radon-Nikodim theorem):

$$Du = \nabla u \, dx + D_s u \tag{2.79}$$

where $\nabla u \in L^1(\Omega)$ and $D_s u \perp dx$. ∇u is called the regular part of Du.

In fact, it is possible to make this decomposition more precise. Let $u \in BV(\Omega)$, we define the approximate upper limit u^+ and approximate lower limit u^-:

$$u^+(x) = \inf \left\{ t \in [-\infty, +\infty]; \lim_{r \to 0} \frac{dx\left(\{u > t\} \bigcap B(x,r)\right)}{r^N} = 0 \right\} \tag{2.80}$$

$$u^-(x) = \sup \left\{ t \in [-\infty, +\infty]; \lim_{r \to 0} \frac{dx\left(\{u < t\} \bigcap B(x,r)\right)}{r^N} = 0 \right\} \tag{2.81}$$

If $u \in L^1(\Omega)$, then:

$$\lim_{r \to 0} \frac{1}{|B(x,r)|} \int_{B(x,r)} |u(x) - u(y)| \, dy = 0 \text{ a.e. x} \tag{2.82}$$

A point x satisfying (2.82) is called a Lebesgue point of u, for such a point we have $u(x) = u^+(x) = u^-(x)$ and:

$$u(x) = \lim_{r \to 0} \frac{1}{|B(x,r)|} \int_{B(x,r)} u(y) \, dy \tag{2.83}$$

We denote by S_u the *jump set* of u, that is, the complement, up to a set of $\mathcal{H}^{N-1}$ measure zero, of the set of Lebesgue points:

$$S_u = \left\{ x \in \Omega; u^-(x) < u^+(x) \right\} \tag{2.84}$$

Then S_u is countably rectifiable, and for $\mathcal{H}^{N-1}$-a.e. $x \in \Omega$, we can define a normal $n_u(x)$ as $\lim_{r \downarrow 0} \frac{Du(B_r(x))}{|Du|(B_r(x))}$

The complete decomposition of Du ($u \in BV(\Omega)$) is thus:

$$Du = \nabla u \, dx + (u^+ - u^-) n_u \mathcal{H}^{N-1}_{|S_u} + C_u \tag{2.85}$$

Here, $J_u = (u^+ - u^-) n_u \mathcal{H}^{N-1}_{|S_u}$ is the jump part, and C_u the Cantor part. We have $C_u \perp dx$, and C_u is diffuse, i.e. $C_u(\{x\}) = 0$. More generally, $C_u(B) = 0$ for all B such that $\mathcal{H}^{N-1}(B) < +\infty$, i.e. the Hausdorff dimension of the support of C_u is strictly greater than $N-1$.

We finally have:

$$\int_\Omega |Du| = \int_\Omega |\nabla u| \, dx + \int_{S_u} |u^+ - u^-| \, d\mathcal{H}^{N-1} + \int_{\Omega \backslash S_u} |C_u| \tag{2.86}$$

Notice that the subset of $BV(\Omega)$ function for which the Cantor part is zero is called $SBV(\Omega)$ and has also some interesting compactness properties.

Example (Devil's staircase) :

For an example of BV function with Du reduced to its Cantor part, see fig 10.3 p 408 in [7] (with the Cantor-Vitali function). $\Omega = (0,1)$, $C = \bigcap_{n \in N} C_n$, where C_n is the union of 2^n intervals of size 3^{-n}. We define:

$$f_n(x) = (2/3)^{-n} \mathbf{1}_{C_n} \ , \ u_n(x) = \int_0^x f_n(t) \, dt$$

For all n, u_n is in $C([0,1])$. Moreover, with the Cauchy criterion, one can show that u_n uniformly converges to some u (which is thus continuous).

Thanks to the lsc of the total variation, we have:

$$\int_{(0,1)} |Du| \leq \lim \inf_{n \to +\infty} \int_{(0,1)} |Du_n| \, dx = 1$$

Thus u is in $BV(0,1)$, and since u continuous, J_u is empty. Moreover, since u is locally constant on $(0,1) \backslash C$, and $\mathcal{L}^1(C) = 0$, one has $\nabla u = 0$ and $Du = C_u$. Finally, the support of C_u is the Cantor set C whose Hausdorff dimension is $\log(2)/\log(3)$.

Remark: $\sin(1/x)$ on $\Omega = (0,1)$ is a continuous function, but does not bleong to $BV(\Omega)$.

Chain rule: if u in $BV(\Omega)$, $g : \mathbb{R} \mapsto \mathbb{R}$ Lipshitz, then $g \circ u$ belongs to $BV(\Omega)$, and the regular part of Dv is given by $\nabla v = g^{'}(u) \nabla u$.

2.5.5 *SBV*

Definition 2.12. [SBV] A function $u \in \mathrm{BV}(\Omega)$ is a special function of bounded variation if its distributional derivative can be decomposed as

$$Du = \nabla u \mathcal{L}^N + (u^+ - u^-) \nu_u \mathcal{H}^{N-1} \llcorner S_u$$

where S_u denotes the approximate discontinuity set, $u^{\pm}$ the approximate upper and lower limits of u on S_u, ν_u the generalized normal to S_u defined as $\lim_{r \downarrow 0} \frac{Du(B_r(x))}{|Du|(B_r(x))}$, ∇u the approximate gradient of u and $\mathcal{H}^{N-1}$ the $N-1$-dimensional Hausdorff measure.

The space of special functions of bounded variation in Ω is denoted as $\mathrm{SBV}(\Omega)$.

Again, this definition can be extended to vector-valued functions and we say that $u \in \mathrm{SBV}(\Omega, \mathbb{R}^m)$ if $u \in \mathrm{BV}(\Omega, \mathbb{R}^m)$ and

$$Du = \nabla u \mathcal{L}^N + (u^+ - u^-) \otimes \nu_u \mathcal{H}^{N-1} \llcorner S_u.$$

A very useful compactness theorem due to L. Ambrosio holds in SBV:

Theorem 2.11. *[Compactness in* SBV*] Let $(u_n)_{n\in\mathbb{N}}$ be a sequence of functions in* $\mathrm{SBV}(\Omega)$ *such that*

$$\sup_{n\in N} \left[\|u_n\|_\infty + \int_\Omega \varphi(|\nabla u_n|)dx + \mathcal{H}^{N-1}(S_{u_n}) \right] < \infty$$

where $\varphi : [0, \infty[\to [0, \infty]$ is a lower semicontinuous, increasing and convex function such that $\lim_{t\to\infty} \frac{\varphi(t)}{t} = \infty$.
Then there exists a subsequence $(u_{h(n)})_{n\in\mathbb{N}}$, and a limit function $u \in \mathrm{L}^\infty(\Omega) \cap \mathrm{SBV}(\Omega)$ such that

- $u_{h(n)}$ *weakly-$*$ converges to u in* $\mathrm{BV}(\Omega)$,
- $\nabla u_{h(n)}$ *weakly converges to ∇u in* $\mathrm{L}^1(\Omega, \mathbb{R}^N)$,
- $\displaystyle\int_\Omega \varphi(|\nabla u|)dx \leq \liminf_{n\to\infty} \int_\Omega \varphi(|\nabla u_{h(n)}|)dx,$
- $\mathcal{H}^{N-1}(S_u) \leq \liminf_n \mathcal{H}^{N-1}(S_{u_{h(n)}})$.

We shall use later in a proof the notion of trace of BV functions. Let us recall the definition and a couple of important properties.

Theorem 2.12. *[Boundary trace theorem] Let u in* $\mathrm{BV}(\Omega)$*. Then, for $\mathcal{H}^{N-1}$ almost every x in $\partial\Omega$, there exists $= \mathcal{T}u(x) \in \mathbb{R}$ such that:*

$$\lim_{\rho\to 0} \frac{1}{\rho^N} \int_{\Omega \bigcap B_\rho(x)} |u(y) - \mathcal{T}u(x)|\, dy = 0$$

Moreover, $\|\mathcal{T}u\|_{L^1(\partial\Omega)} \leq C\|u\|_{\mathrm{BV}(\Omega)}$ for some constant C depending only on Ω. The extension $\tilde{u}$ of u to 0 out of Ω belongs to $\mathrm{BV}(\mathbb{R}^N)$, and viewing Du as a measure on the whole of $\mathbb{R}^N$ and concentrated on Ω, $D\tilde{u}$ is given by :

$$D\tilde{u} = Du + (\mathcal{T}u)\nu_\Omega \mathcal{H}^{N-1} \llcorner \partial\Omega$$

with ν_Ω the generalised inner normal to $\partial\Omega$.

The trace operator is not continuous with respect to the weak-$*$ convergence, but it is continuous with respect to the strict convergence.

Theorem 2.13. *[Continuity of the trace operator] The trace operator $u \mapsto Tu$ is continuous between* $\mathrm{BV}(\Omega)$*, endowed with the topology induced by the strict convergence, and* $\mathrm{L}^1(\partial\Omega, \mathcal{H}^{N-1} \llcorner \partial\Omega)$.

2.5.6 Sets of finite perimeter

Definition 2.13. Let E be a measurable subset of $\mathbb{R}^2$. Then for any open set $\Omega \subset \mathbb{R}^2$, we call perimeter of E in Ω, denoted by $P(E,\Omega)$, the total variation of $\mathbf{1}_E$ in Ω, i.e.:

$$P(E,\Omega) = \sup\left\{\int_E \operatorname{div}(\phi(x))dx / \phi \in C_c^1(\Omega;\mathbb{R}^2), \|\phi\|_{L^\infty(\Omega)} \leq 1\right\} \tag{2.87}$$

We say that E has finite perimeter if $P(E,\Omega) < \infty$.

Remark: If E has a C^1-boundary, this definition of the perimeter corresponds to the classical one. We then have:

$$P(E,\Omega) = \mathcal{H}^1(\partial E \bigcap \Omega) \tag{2.88}$$

where $\mathcal{H}^1$ stands for the 1-dimensional Hausdorff measure [5]. The result remains true when E has a Lipschitz boundary.

In the general case, if E is any open set in Ω, and if $\mathcal{H}^1(\partial E \bigcap \Omega) < +\infty$, then:

$$P(E,\Omega) \leq \mathcal{H}^1(\partial E \bigcap \Omega) \tag{2.89}$$

Definition 2.14. We denote by $\mathcal{F}E$ the reduced boundary of E.

$$\mathcal{F}E = \left\{x \in \text{ support } \left(|D_{\mathbf{1}_E}| \bigcap \Omega\right) \ / \ \nu_E = \lim_{\rho\to 0} \frac{D_{\mathbf{1}_E(B_\rho(x))}}{\left|D_{\mathbf{1}_E(B_\rho(x))}\right|} \text{ exists and verifies } |\nu_E| = 1\right\} \tag{2.90}$$

Definition 2.15. For all $t \in [0,1]$, we denote by E^t the set

$$\left\{x \in \mathbb{R}^2 \ / \ \lim_{\rho\to 0} \frac{|E \bigcap B_\rho(x)|}{|B_\rho(x)|} = t\right\} \tag{2.91}$$

of points where E is of density t, where $B_\rho(x) = \{y \ / \ \|x-y\| \leq \rho\}$. We set $\partial^* E = \mathbb{R}^2 \setminus (E^0 \bigcup E^1)$ the essential boundary of E.

Theorem 2.14. *[Federer [5]]. Let E a set with finite perimeter in Ω. Then:*

$$\mathcal{F}E \bigcap \Omega \subset E^{1/2} \subset \partial^* E \tag{2.92}$$

and

$$\mathcal{H}^1\left(\Omega \setminus \left(E^0 \bigcup \mathcal{F}E \bigcup E^1\right)\right) = 0 \tag{2.93}$$

Remark: If E is Lipschitz, then $\partial E \subset \partial^* E$. In particular, since we always have $\mathcal{F}E \subset \partial E$ (see [38]):

$$P(E,\Omega) = \mathcal{H}^1(\partial E \bigcap \Omega) = \mathcal{H}^1(\partial^* E \bigcap \Omega) = \mathcal{H}^1(\mathcal{F}E \bigcap \Omega) \tag{2.94}$$

Theorem 2.15. *[De Giorgi [5]]. Let E a Lebesgue measurable set of $\mathbb{R}^2$. Then $\mathcal{F}E$ is 1-rectifiable.*

We recall that E is 1-rectifiable if and only if there exist Lipschitz functions $f_i : \mathbb{R}^2 \to \mathbb{R}$ such that $E \subset \bigcup_{i=0}^{+\infty} f_i(\mathbb{R})$.

2.5.7 Coarea formula and applications

Theorem 2.16. Coarea formula *If u in $BV(\Omega)$, then:*

$$J(u) = \int_{-\infty}^{+\infty} P\left(\{x \in \Omega \,:\, u(x) > t\}, \Omega\right)\, dt \tag{2.95}$$

In particular, for a binary image whose gray level values are only 0 or 1, the total variation is equal to the perimeter of the object inside the image.

A straightforward consequence :

Proposition 2.19. *Let $u \in BV(\Omega)$ and $M \in \mathbb{R}$. Then $v = \inf(u, M)$ is in $BV(\Omega)$, and* $\int_\Omega |Dv| \leq \int_\Omega |Du|$

Proof:

$$\begin{aligned}
\int_\Omega |Dv| &= \int_{-\infty}^{+\infty} P\left(\{x \in \Omega \,:\, v(x) > t\}, \Omega\right)\, dt \\
&= \int_{-\infty}^{M} P\left(\{x \in \Omega \,:\, v(x) > t\}, \Omega\right)\, dt \\
&= \int_{-\infty}^{M} P\left(\{x \in \Omega \,:\, u(x) > t\}, \Omega\right)\, dt \\
&\leq \int_{-\infty}^{+\infty} P\left(\{x \in \Omega \,:\, u(x) > t\}, \Omega\right)\, dt \\
&= \int_\Omega |Du|
\end{aligned}$$

■

Example of the use of the coarea formula :
Consider the ROF model:

$$\inf_{u \in BV(\Omega)} \left(\int_\Omega |Du| + \int_\Omega (u-f)^2\, dx \right) \tag{2.96}$$

under the assumption that $u \geq 0$ (which is a reasonable assumption in image processing). Then, from the coarea forumal, we have:

$$\int_\Omega |Du| = \int_{-\infty}^{+\infty} P\left(\{u \geq t\}, \Omega\right)\, dt = \int_0^{+\infty} P\left(\{u \geq t\}, \Omega\right)\, dt \tag{2.97}$$

Let us now consider the second term. We have:

$$(u-f)^2 = \int_0^u 2(t-f)\, dt = \int_0^{+\infty} \mathbf{1}_{\{u \geq t\}}\, 2(t-f)\, dt \tag{2.98}$$

Then, using Fubini theorem:

$$\int_\Omega (u-f)^2\, dx = \int_\Omega \int_0^{+\infty} \mathbf{1}_{\{u \geq t\}}\, 2(t-f)\, dt\, dx = \int_0^{+\infty} \int_{\{u \geq t\}} 2(t-f)\, dx\, dt \tag{2.99}$$

Getting back to the original problem, we get, if we note $E_t = \{u \geq t\}$:

$$\int_\Omega |Du| + \int_\Omega (u-f)^2 \, dx = \int_0^{+\infty} \left(P\left(E_t, \Omega\right) + \int_{E_t} 2(t-f)\, dx \right) \, dt \tag{2.100}$$

We thus see that solving (2.96) is equivalent to solving for all $t \geq 0$:

$$\inf_{E_t \subset \Omega} \left(P\left(E_t, \Omega\right) + \int_{E_t} 2(t-f)\, dx \right) \tag{2.101}$$

A useful inequality :

Proposition 2.20.

$$P(E \bigcap F, \Omega) + P(E \bigcup F, \Omega) \leq P(E, \Omega) + P(F, \Omega) \tag{2.102}$$

Proof :
See [5] proposition 3.38 page 144. It suffices to take u_n and v_n in $C^\infty(\Omega)$ converging to $\mathbf{1}_E$ and $\mathbf{1}_F$ in $L^1(\Omega)$ with $0 \leq u_n, v_n \leq 1$, and $\lim_n \int_\Omega |\nabla u_n| \, dx = P(E, \Omega)$ and $\lim_n \int_\Omega |\nabla v_n| \, dx = P(F, \Omega)$. Since $u_n v_n$ converges to $\mathbf{1}_{E \bigcap F}$ and $u_n + v_n - u_n v_n$ converges to $\mathbf{1}_{E \bigcup F}$, we obtain the result by passing to the imit in the inequality (since $u_n v_n + (u_n + v_n - u_n v_n) = u_n + v_n$):

$$\int_\Omega |\nabla (u_n v_n)| \, dx + \int_\Omega |\nabla (u_n + v_n -_n v_n)| \, dx \leq \int_\Omega |\nabla u_n| \, dx + \int_\Omega |\nabla v_n| \, dx$$

■

We use here the following classical notations: $u \vee v = \sup(u, v)$, and $u \wedge v = \inf(u, v)$.

Proposition 2.21. *u and v in $BV(\Omega)$. Then:*

$$J(u \vee v) + J(u \wedge v) \leq J(u) + J(v) \tag{2.103}$$

Proof : This a direct consequence of the previous proposition and the coarea formula. Indeed, if $t \in \mathbb{R}$ if we set $E_t = \{x\,;\, u(x) \geq t\}$ and $F_t = \{x\,;\, v(x) \geq t\}$, then from the previous proposition we have:

$$P(E_t \bigcap F_t, \Omega) + P(E_t \bigcup F_t, \Omega) \leq P(E_t, \Omega) + P(F_t, \Omega) \tag{2.104}$$

We then integrate over $\mathbb{R}$ and use the coarea formula.

■

3. Energy methods

For further details, we encourage the reader to look at [9].

3.1 Introduction

In many problems in image processing, the goal is to recover an ideal image u from an observation f.

u is a perfect original image describing a real scene.

f is an observed image, which is a degraded version of u.

The simplest modelization is the following:

$$f = Au + v \tag{3.1}$$

where v is the noise,

and A is the blur, a linear operator (often a convolution).

As already seen before, the ML method leads to consider the following problem:

$$\inf_u \|f - Au\|_2^2 \tag{3.2}$$

where $\|.\|_2$ stands for the L^2 norm. This is an ill-posed problem, and it is classical to consider a regularized version:

$$\inf_u \underbrace{\|f - Au\|_2^2}_{\text{data term}} + \underbrace{L(u)}_{\text{regularization}} \tag{3.3}$$

3.2 Tychonov regularization

3.2.1 Introduction

This is probably the simplest regularization choice: $L(u) = \|\nabla u\|_2^2$.

The considered problem is the following:

$$\inf_{u \in W^{1,2}(\Omega)} \|f - Au\|_2^2 + \lambda \|\nabla u\|_2^2 \tag{3.4}$$

where A is a continuous and linear operator of $L^2(\Omega)$ such that $A(1) \neq 0$.

This is not a good regularization choise in image processing: the restored image u is much too smoothed (in particular, the edges are eroded). But we study it as an illustration of the previous sections.

We denote by:

$$F(u) = \|f - Au\|^2 + \lambda \|\nabla u\|^2 \tag{3.5}$$

Using the previous results, it is easy to show that:

(i) F is coercive on $W^{1,2}(\Omega)$.

(ii) $W^{1,2}(\Omega)$ is reflexive.

(iii) F is convex and l.s.c. on $W^{1,2}(\Omega)$.

As a consequence, the direct method of calculus of variation shows that problem (3.4) admits a solution u in $W^{1,2}(\Omega)$.

Moreover, since $A(1) \neq 0$ it is easy to show that F is strictly convex, which implies that the solution u is unique.

This solution u is characterized by it Euler-Lagrange equation. It is easy to show that the Euler-Lagrange equation associated to (3.4) is:

$$-A^* f + A^* A u - \lambda \Delta u = 0 \tag{3.6}$$

with Neumann boundary conditions $\frac{\partial u}{\partial N} = 0$ on $\partial\Omega$. We recall that A^* is the adjoint operator to A.

3.2.2 Sketch of the proof (to fix some ideas)

Computation of the Euler-Lagrange equation:

$$\begin{aligned}
\frac{1}{\alpha}(F(u+\alpha v) - F(u)) &= \frac{1}{\alpha}\|f - Au - \alpha A v\|_2^2 + \lambda\|\nabla u + \alpha \nabla v\|_2^2 - \|f - Au\|_2^2 + \lambda\|\nabla u\|_2^2 \\
&= \frac{1}{\alpha}\left(\langle \alpha A v, 2(Au - f) + \alpha A v\rangle + \lambda\langle \alpha \nabla v, 2\nabla u + \alpha \nabla v\rangle\right) \\
&= \langle v, 2A^*(Au - f)\rangle + 2\lambda\langle \nabla v, \nabla u\rangle + 0(\alpha) \\
&= 2\langle v, A^* A u - A^* f\rangle - 2\lambda\langle v, \Delta u\rangle + 0(\alpha)
\end{aligned}$$

Hence

$$F'(u) = 2\left(A^* A u - A^* f - \lambda \Delta u\right) \tag{3.7}$$

Convexity, continuity, coercivity: We have:

$$F''(u) = 2\left(A^* A. - \lambda \Delta.\right) \tag{3.8}$$

F'' is positive. Indeed: $\langle A^* A w, w\rangle = \|Aw\|_2^2 \geq 0$, and $\langle -\Delta w, w\rangle = \|\nabla w\|_2^2 \geq 0$. Hence F is convex. Moreover, since $A1 \neq 0$, F is definite positive, i.e. $\langle F''(u) w, w\rangle > 0$ for all $w \neq 0$. Hence F is strictly convex.

For the coercivity, for the sake of simplicity, we make the additional assumption that A is coercive (notice that this assumption can be dropped, the general proof requires the use of Poincaré inequality), i.e. there exists $\beta > 0$ such that $\|Ax\|_2^2 \geq \beta \|x\|_2^2$:

$$\begin{aligned}
F(u) &= \|f\|^2 + \|Au\|^2 - 2\alpha\langle f, Au\rangle + \lambda\|\nabla u\|^2 \\
&= \left(\|Au\|^2 + \lambda\|\nabla u\|^2\right) - 2\alpha\langle A^* f, u\rangle + \|f\|^2 \\
&\geq \left(\beta\|u\|^2 + \lambda\|\nabla u\|^2\right) - 2\alpha\langle A^* f, u\rangle + \|f\|^2
\end{aligned}$$

Since A is linear continuous, $u \to \|f - Au\|_2^2$ is l.s.c. Hence F is l.s.c. This conlude the proof.

■

3.2.3 General case

The considered problem is the following:

$$\inf_{u \in W^{1,2}(\Omega)} \|f - Au\|_2^2 + \lambda\|\nabla u\|_2^2 \tag{3.9}$$

where A is a continuous and linear operator of $L^2(\Omega)$ such that $A(1) \neq 0$, and $f \in L^2(\Omega)$.

Uniqueness of the solution: Let u_1 and u_2 two solutions. From the strict convexity of $\|.\|_2^2$, we get that $\nabla u_1 = \nabla u_2$, and thus $u_1 = u_2 + c$. From the strict convexity of $Au \mapsto \|f - Au\|_2^2$, we have $Au_1 = Au_2$. But $Au_1 = Au_2 + c\,A(1)$, and thus, since $A(1) \neq 0$, we deduce that $c = 0$.

Existence of a solution: Let u_n be a minimizing sequence. We thus have

$$\int |\nabla u_n|^2 \leq M \tag{3.10}$$

(M generic positive contant), and

$$\int |f - Au_n|^2 \leq M \tag{3.11}$$

By triangular inequaliy, we have: $\|Au_n\|_2 \leq \|f\|_2 + \|f - Au_n\|_2 \leq M$. *So if* A *is assumed coercive, we get* $\|u_n\|_2 \leq M$. But if there is no coercivity assumption like $\|Au\| \geq c\|u\|$, we have to work a little more.

Let us note

$$w_n = \frac{1}{|\Omega|}\int_\Omega u_n \quad \text{and} \quad v_n = u_n - w_n \tag{3.12}$$

Remark that $\int |\nabla v_n|^2 = \int |\nabla u_n|^2$. From Poincaré inequality, we get that:

$$\|v_n\|_2 \leq K\|\nabla u_n\|_2 \leq C \tag{3.13}$$

(3.11) can be rewritten in: $\int |f - Aw_n - Av_n| \leq C$. But $Aw_n = (Aw_n + Av_n - f) - (Av_n - f)$. Hence, using the triangular inequality:

$$\|Aw_n\|_{L^2(\Omega)} \leq C + \|Av_n\|_{L^2(\Omega)} + \|f\|_{L^2(\Omega)} \leq C \tag{3.14}$$

using the fact that A is continuous.

For all n, w_n is a constant function over Ω. Hence $Aw_n = w_n\,A(1)$. Moreover, we have assumed that $A(1) \neq 0$. We thus deduce from (3.14) that $|w_n|\,A(1) \leq C$, i.e.

$$\|w_n\|_2 \leq C \tag{3.15}$$

Now, using (3.13) and (3.15), we get:

$$\|u_n\|_2 = \|v_n + w_n\|_2 \leq \|v_n\|_2 + \|w_n\|_2 \leq C \tag{3.16}$$

Hence we deduce that u_n is bounded in $W^{1,2}(\Omega)$.

Since $W^{1,2}(\Omega)$ reflexive, there exists u in $W^{1,2}(\Omega)$ such that up to a subsequence, $u_n \rightharpoonup u$ in $W^{1,2}(\Omega)$ weak. By compact Sobolev embedding, we have $u_n \to u$ in $L^2(\Omega)$ strong. Since A is continuous, we have $\|f - Au_n\|_2^2 \to \|f - Au\|_2^2$. Moreover, since $\nabla u_n \rightharpoonup \nabla u$, we have $\liminf \|\nabla u_n\|_2^2 \geq \|\nabla u\|_2^2$. This conclude the proof.

■

Computation of the Euler-Lagrange equation and of the boundary conditions:

$$F(u) = \int_\Omega |f - Au|^2 \, dx + \lambda \int_\Omega |\nabla u|^2 \, dx \tag{3.17}$$

$$F(u+h) - F(u) = 2\int_\Omega h(A^*Au - A^*f) + 2\lambda \int_\Omega \nabla u.\nabla h + o(|h|) \tag{3.18}$$

We choose $h \in C_c^\infty(\Omega)$. We thus have (in $\mathcal{D}'(\Omega)$)

$$\int_\Omega \nabla u.\nabla h = -\int_\Omega h\Delta u \tag{3.19}$$

And then, if $h \in C_c^\infty(\Omega)$:

$$F(u+h) - F(u) = 2\int_\Omega h(A^*Au - A^*f + \lambda\Delta u + o(|h|) \tag{3.20}$$

Hence we deduce that:

$$\nabla F(u) = 2(A^*Au - A^*f - \lambda\Delta u) \tag{3.21}$$

in $\mathcal{D}'(\Omega)$.

Now, if u is the solution of problem (3.9), then $\nabla F(u) = 0$. Hence, for this particular u, we get that in $\mathcal{D}'(\Omega)$:

$$A^*Au - A^*f + \lambda\Delta u = 0 \tag{3.22}$$

But u belongs to $W^{1,2}(\Omega)$, A is a continuous operator of $L^2(\Omega)$, and f belongs to $L^2(\Omega)$. Hence $A^*Au - A^*f$ belongs to $L^2(\Omega)$ and thus we have Δu in $L^2(\Omega)$. We deduce that equality (3.22) holds in $L^2(\Omega)$.

We now choose $h \in C_c^\infty(\bar{\Omega})$. We have

$$\int_\Omega \nabla u.\nabla h = \int_{\partial\Omega} h\nabla u.N - \int_\Omega h\Delta u \tag{3.23}$$

and thanks to (3.22) we get:

$$F(u+h) - F(u) = 2\int_\Omega h(A^*Au - A^*f - \lambda\Delta u) + \int_{\partial\Omega} h\nabla u.N + o(|h|) \tag{3.24}$$

Now, since $\nabla F(u) = 0$ in $L^2(\Omega)$ and using (3.22), we get

$$\nabla u.N = 0 \text{ on } \partial\Omega \tag{3.25}$$

■

Convexity :

From the previous computation, we have:

$$\langle \nabla F(u), h \rangle = 2\int_\Omega h(A^*Au - A^*f + \lambda\nabla^*\nabla u) \tag{3.26}$$

i.e. :

$$\nabla F(u) = 2(A^*Au - A^*f + \lambda\nabla^*\nabla u) \tag{3.27}$$

Hence :

$$\nabla^2 F(u) = 2(A^*A. + \lambda\nabla^*\nabla.) \tag{3.28}$$

And $\nabla^2 F(u)$ is a positive operator:

$\langle \nabla^2 F(u)w, w \rangle = 2\|Au\|_2^2 + \|\nabla u\|_2^2 \geq 0$.

Moreover, since $A(1) \neq 0$ and since the kernel of ∇ is $\lambda.1$, $\lambda \in \mathbb{R}$, we deduce that $\langle \nabla^2 F(u)w, w \rangle > 0$ if $w \neq 0$. We therefore conclude that F is strictly convex.

Maximum principle (Stampacchia truncation)

Proposition 3.1. *Let f be in $L^\infty(\Omega)$ and u be a solution of the Tychonov Problem with $A = Id$. Then a maximum principle holds for u:*

$$\inf_\Omega f \leq u \leq \sup_\Omega f \tag{3.29}$$

We first multiply equation (3.22) by a function $v \in W^{1,2}(\Omega)$, and we integrate by parts (we can do it since we saw that the functions are in $L^2(\Omega)$):

$$\lambda \int_\Omega \nabla u . \nabla v + \int_\Omega v(u - f) = 0 \tag{3.30}$$

Let G be a truncature function of class C^1, such that $G(t) = 0$ on $(-\infty, 0]$, and G strictly increasing on $[0, +\infty)$, and $G^{'} \leq M$ where M is a constant.

We choose $v = G(u-k)$ where k is a constant such that $k \geq \|f\|_{L^\infty}$. We remind the reader that u belongs to $W^{1,2}(\Omega)$. Notice that thanks to the chain rule in $W^{1,2}(\Omega)$, we know that $v = G(u - k)$ also belongs to $W^{1,2}(\Omega)$, and that $\nabla v = G^{'}(u - k)\nabla u$. Equation (3.30) writes:

$$\int_\Omega |\nabla u|^2 \, G^{'}(u_\epsilon - k) \, dx + \int_\Omega (u - f) \, G(u - k) \, dx = 0 \tag{3.31}$$

from which we deduce that (using the properties of G):

$$\int_\Omega u \, G(u - k) \, dx \leq \int_\Omega f \, G(u - k) \, dx \tag{3.32}$$

And then:

$$\int_\Omega (u - k) \, G(u - k) \, dx \leq \int_\Omega (f - k) \, G(u - k) \, dx \tag{3.33}$$

We have $f - k \leq 0$ and $G(u - k) \geq 0$, hence $\int_\Omega \int_\Omega (u - k) \, G(u - k) \, dx \leq 0$.

But $t.G(t) \geq 0$ for all t, and we thus deduce that $(u - k) \, G(u - k) = 0$ a.e., hence $u \leq k$.

We get the opposite inequality by considering $-u$.

■

3.2.4 Minimization algorithms

PDE based method: (3.6) is embedded in a fixed point method:

$$\frac{\partial u}{\partial t} = \lambda \Delta u + A^* f - A^* A u \tag{3.34}$$

Figure 7 shows a numerical example in the case when $A = Id$.

Fourier transform based numerical approach In this case, a faster approach consists in using the Fourier transform.

We detail below how the model can be solved in discrete using the discrete Fourier transform (DFT).

We recall that the DFT of a discrette image is $(f(m, n))$ $(0 \leq m \leq N-1$ and $0 \leq n \leq N-1)$ is given by $(0 \leq p \leq N - 1$ and $0 \leq q \leq N - 1)$:

Figure 7: Restauration (Tychonov) par EDP

$$\mathcal{F}(f)(p,q) = F(p,q) = \sum_{m=0}^{N-1}\sum_{n=0}^{N-1} f(m,n)e^{-j(2\pi/N)pm}e^{-j(2\pi/N)qn} \tag{3.35}$$

and the inverse transform is:

$$f(m,n) = \frac{1}{N^2}\sum_{p=0}^{N-1}\sum_{q=0}^{N-1} F(p,q)e^{j(2\pi/N)pm}e^{j(2\pi/N)qn} \tag{3.36}$$

Moreover we have $\|\mathcal{F}(f)\|_X^2 = N^2\|f\|_X^2$ et $(\|\mathcal{F}(f), \|\mathcal{F}(g))_X = N^2(f,g)_X$.
It is possible to show that:

$$\|\mathcal{F}(\nabla f)\|^2 = \sum_{p,q} |\mathcal{F}(\nabla f)(p,q)|^2 = \sum_{p,q} 4\,|\mathcal{F}(f)(p,q)|^2 \left(\sin^2\frac{\pi p}{N} + \sin^2\frac{\pi q}{N}\right) \tag{3.37}$$

Using Parseval identity, it can be deduced that the solution u of (3.4) satifies:

$$\mathcal{F}(u)(p,q) = \frac{\mathcal{F}(f)(p,q)}{1 + 8\lambda\left(\sin^2\frac{\pi p}{N} + \sin^2\frac{\pi q}{N}\right)} \tag{3.38}$$

Figure 10 shows a numerical example obtained with this approach.

3.3 Rudin-Osher-Fatemi model

3.3.1 Introduction

In [65], the authors advocate the use of $L(u) = \int |Du|$ as regularization. With this choice, the recovered image can have some discontinuities (edges). The considered model is the following:

Figure 8: Image originale

Figure 9: Image bruitée à restaurer

Figure 10: Restauration (Tychonov) par TFD

$$\inf_{u\in BV(\Omega)}\left(J(u)+\frac{1}{2\lambda}\|f-Au\|_2^2\right) \tag{3.39}$$

where $J(u)=\int_\Omega |Du|$ stands for the total variation of u, and where A is a continuous and linear operator of $L^2(\Omega)$ such that $A1\neq 0$ (the case when A is compact is simpler).

The mathematical study of (3.39) is done in [25].

The proof of existence of a solution is similar to the one for the Tychonov regularization (except that now one works in $BV(\Omega)$ instead of $W^{1,2}(\Omega)$. If A is assumed to be injective, then this solution is unique.

Sketch of the proof of existence: We denote by:

$$F(u)=J(u)+\frac{1}{2\lambda}\|f-Au\|_2^2 \tag{3.40}$$

F is convex.

For the sake of simplicity, we assume that $A=Id$. See [9] for the detailed proof in the general case.

Let us consider u_n a minimizing sequence for (3.39) with $A=Id$. Hence there exists $M>0$ such that $J(u_n)\leq M$ and $\|f-u_n\|_2^2\leq M$ (M denotes a generic positive constant during the proof). Moreover, since $\|u_n\|_2\leq\|f\|_2+\|f-u_n\|_2$, we have $\|u_n\|_2\leq M$. Hence u_n is bounded in $BV(\Omega)$. By weak-* compacity, there exists therefore u in $BV(\Omega)$ such that $u_n\to u$ in $L^1(\Omega)$ strong and $Du_n\rightharpoonup Du$.

By l.s.c. of the total variation, we have $J(u)\leq\liminf J(u_n)$, and by l.s.c. of the weak norm we have $\|f-u\|^2\leq\lim\inf\|f-u_n\|^2$. Hence, up to a subsequence, we have $\lim_{n\to+\infty}\inf F(u_n)\geq F(u)$.

■

Maximum principle (truncation argument)

Proposition 3.2. *Let f be in $L^\infty(\Omega)$ and u be a solution of the ROF Problem with $A=Id$. Then a maximum principle holds for u:*

$$\inf_\Omega f\leq u\leq\sup_\Omega f \tag{3.41}$$

It is based on a standard truncation argument. We remark that $x\mapsto(x-a)^2$ is decreasing if $x\in(-\infty,a)$ and increasing if $x\in(a,+\infty)$. Therefore, if $M\geq a$, one always has:

$$(\min(x,M)-a)^2\leq(x-a)^2 \tag{3.42}$$

Hence, if we let $M=\sup_\Omega f$, we find that:

$$\int_\Omega(\min(u,M)-f)^2\,dx\leq\int_\Omega(u-f)^2\,dx \tag{3.43}$$

Moreover, it is a direct consequence of the coarea formula that $|D(\min(u,\sup f))|(\Omega)\leq|Du|(\Omega)$ which yields that the function $\min(u,\sup f)$ is a solution of the ROF Problem. We get the opposite inequality with a similar argument.

Hence we do not decrease the energy by assuming that $\inf f\leq u\leq\sup f$. We conclude by using the uniqueness of the solution of the ROF problem.

■

Comparison principle We now state a comparison principle.

Proposition 3.3. *Let f_1 and f_2 be in $L^\infty(\Omega)$. Let us assume that $f_1 \leq f_2$. We denote by u_1 (resp. u_2) a solution of the ROF problem for $f = f_1$ (resp. $f = f_2$). Then we have $u_1 \leq u_2$.*

Proof We first consider the case when $f_1 < f_2$.

We use here the following classical notations: $u \vee v = \sup(u, v)$, and $u \wedge v = \inf(u, v)$.

We have since u_i is a minimizer with data f_i:

$$J(u_1 \wedge u_2) + \int_\Omega (f_1 - u_1 \wedge u_2)^2 \geq J(u_1) + \int_\Omega (f_1 - u_1)^2 \tag{3.44}$$

and:

$$J(u_1 \vee u_2) + \int_\Omega (f_2 - u_1 \vee u_2)^2 \geq J(u_2) + \int_\Omega (f_2 - u_2)^2 \tag{3.45}$$

Adding these two inequalities, and using the fact that [45]:

$$J(u_1 \wedge u_2) + J(u_1 \vee u_2) \leq J(u_1) + J(u_2) \tag{3.46}$$

we get:

$$\int_\Omega \left((f_1 - u_1 \wedge u_2)^2 - (f_1 - u_1)^2 + (f_2 - u_1 \vee u_2)^2 - (f_2 - u_2)^2 \right) \geq 0 \tag{3.47}$$

Writing $\Omega = \{u_1 > u_2\} \cup \{u_1 \leq u_2\}$, we easily deduce that:

$$\int_{\{u_1 > u_2\}} \left((f_1 - u_2)^2 - (f_1 - u_1)^2 + (f_2 - u_1)^2 - (f_2 - u_2)^2 \right) \geq 0 \tag{3.48}$$

i.e.:

$$\int_{\{u_1 > u_2\}} 2\left(-f_1 u_2 + f_1 u_1 - f_2 u_1 + f_2 u_2 \right) \geq 0 \tag{3.49}$$

i.e.:

$$\int_{\{u_1 > u_2\}} (f_2 - f_1)(u_2 - u_1) \geq 0 \tag{3.50}$$

Since $f_1 < f_2$, we thus deduce that $\{u_1 > u_2\}$ has a zero Lebesgue measure, i.e. $u_1 \leq u_2$ a.e. in Ω.

Now, for the general case $f_1 \leq f_2$, we make use of the following lemma:

Lemma 3.1. *u_i solution of the ROF problem with data f_i. Then we have:*

$$\|u_1 - u_2\|_2 \leq \|f_1 - f_2\|_2 \tag{3.51}$$

From the above lemma, the maping $f_i \mapsto u_i$ is continuous in $L^2(\Omega)$. Hence the result

■

Proof of the lemma :

We use the Euler-Lagrange equation:

$$f_i - u_i \in \partial J(u_i) \tag{3.52}$$

We make the difference of the previous equation with $i = 1$ and then $i = 2$, and multiply by $u_1 - u_2$. We get:

$$\langle f_1 - f_2 - (u_1 - u_2), u_1 - u_2 \rangle = \langle \partial J(u_1) - \partial J(u_2), u_1 - u_2 \rangle \tag{3.53}$$

But, J being convex, it is a monotone operator, i.e. $\langle \partial J(u_1) - \partial J(u_2), u_1 - u_2 \rangle \geq 0$. Hence:

$$\langle f_1 - f_2, u_1 - u_2 \rangle \geq \|u_1 - u_2\|_2^2 \tag{3.54}$$

Using Cauchy Schwartz inequality, we get:

$$\|u_1 - u_2\|_2^2 \leq \|f_1 - f_2\| \, \|u_1 - u_2\| \tag{3.55}$$

and thus $\|f_1 - f_2, u_1 - u_2\| \leq f_1 - f_2\|$. In particular, this implies the uniqueness of u_i when f_i is fixed.

■

Euler-Lagrange equation: Formally, the associated Euler-Lagrange equation is:

$$-\text{div}\left(\frac{\nabla u}{|\nabla u|}\right) + \frac{1}{\lambda}(A^*Au - A^*f) = 0 \tag{3.56}$$

with Neuman boundary conditions.

Numerically, one uses a fixed point process:

$$\frac{\partial u}{\partial t} = \text{div}\left(\frac{\nabla u}{|\nabla u|}\right) - \frac{1}{\lambda}((A^*Au - A^*f) \tag{3.57}$$

However, in this approach, it is needed to regularize the problem, i.e. to replace $\int_\Omega |Du|$ in (3.39) by $\int_\Omega \sqrt{|\nabla u|^2 + \epsilon^2}$. The Euler-Lagrange equation is then:

$$\frac{\partial u}{\partial t} = \text{div}\left(\frac{\nabla u}{\sqrt{|\nabla u|^2 + \epsilon^2}}\right) - \frac{1}{\lambda}((A^*Au - A^*f) \tag{3.58}$$

Moreover, when working with $BV(\Omega)$, (3.56) is not true.

Sketch of the proof:

$$\phi(x) = \sqrt{\epsilon^2 + x^2} \tag{3.59}$$

We have:

$$|\nabla(u+h)|^2 = |\nabla u|^2 + 2\nabla h.\nabla u + o(|h|) \tag{3.60}$$

Thus

$$\sqrt{|\nabla(u+h)|^2} = |\nabla u|\left(1 + \frac{2\nabla h.\nabla u}{2|\nabla u|^2}\right) + o(|h|) = |\nabla u| + \frac{\nabla h.\nabla u}{|\nabla u|} + o(|h|) \tag{3.61}$$

Thus

$$\phi\left(|\nabla(u+h)|\right)=\phi\left(|\nabla u|+\frac{\nabla h.\nabla u}{|\nabla u|}+o(|h|)\right)=\phi\left(|\nabla u|\right)+\frac{\nabla h.\nabla u}{|\nabla u|}\phi'\left(|\nabla u|\right)+o(|h|) \quad (3.62)$$

Thus

$$\int_\Omega \phi(u+h)=\int_\Omega \phi(u)+\int_\Omega -h\mathrm{div}\left(\frac{\phi'\left(|\nabla u|\right)}{|\nabla u|}\nabla u\right)+o(|h|) \quad (3.63)$$

Since

$$\|A(u+h)-f\|^2=\|Au-f\|^2+2\langle h,A^*(Au-f)\rangle+o(|h|) \quad (3.64)$$

We therefore have, since $J(u+h)=J(u)+\langle h,\nabla J(u)\rangle+o(|h|)$:

$$\nabla J(u)=-\lambda\mathrm{div}\left(\frac{\phi'(|\nabla u|)}{|\nabla u|}\nabla u\right)+A^*(Au-f) \quad (3.65)$$

Notice that in the case when $\phi(x)=\sqrt{\epsilon^2+x^2}$, then $\phi'(x)=\frac{x}{\sqrt{\epsilon^2+x^2}}$, and therefore

$$\frac{\phi'(|\nabla u|)}{|\nabla u|}=\frac{1}{\sqrt{\epsilon^2+|\nabla u|^2}} \quad (3.66)$$

■

For the sake of simplicity, we assume in the following that $A=Id$.

3.3.2 Interpretation as a projection

We are therefore interested in solving:

$$\inf_{u\in BV(\Omega)}\left(J(u)+\frac{1}{2\lambda}\|f-u\|_2^2\right) \quad (3.67)$$

We consider the case $N=2$. J is extended to L^2.

Using convex analysis result, the optimality condition associated to the minimization problem (3.67) is:

$$u-f\in\lambda\partial J(u) \quad (3.68)$$

This condition is used in [24] to derive a minmization algorithm for (3.67).

Since J is homogeneous of degree one (i.e. $J(\lambda u)=\lambda J(u)\ \forall u$ and $\lambda>0$), it is standard (cf [36]) that J^* the Legendre Fenchel transform of J,

$$J^*(v)=\sup\left((u,v)_2-J(u)\right) \quad (3.69)$$

is the indicator function of a closed convex set K.

It is easy to check that K identifies with the set (using the fact that $J^{**}=J$):

$$K=\{\mathrm{div}\,(g)/g\in(L^\infty(\Omega))^2,\|g\|_\infty\leq 1\} \quad (3.70)$$

and

$$J^*(v)=\chi_K(v)=\begin{cases}0 & \text{if } v\in K\\ +\infty & \text{otherwise}\end{cases} \quad (3.71)$$

The next result is shown in [24] :

Proposition 3.4. *The solution of (3.67) is given by:*

$$u=f-P_{\lambda K}(f) \quad (3.72)$$

where P is the orthogonal projection on λK.

Proof: If $\hat{u}$ is a minimizer, then

$$0 \in (\hat{u} - f)/\lambda + \partial J(\hat{u}) \tag{3.73}$$

i.e. :

$$(f - \hat{u})/\lambda \in \partial J(\hat{u}) \tag{3.74}$$

Hence

$$\hat{u} \in \partial J^* ((f - \hat{u})/\lambda) \tag{3.75}$$

We set $\hat{w} = (f - \hat{u})$, and we get:

$$0 \in \hat{w} - f + \partial J^* (\hat{w}/\lambda) \tag{3.76}$$

We then deduce that $\hat{w}$ is the minimizer of:

$$\inf_w \left(\|w - f\|^2 + \frac{1}{2\lambda} J^* (w/\lambda) \right) \tag{3.77}$$

i.e. $\hat{w} = P_{\lambda K}(f)$, hence $\hat{u} = f - P_{\lambda K}(f)$.

■

Algorithm : [24] proposes an algorithm to compute $P_{\lambda K}(f)$ which can be written in discrete:

$$\min \left\{ \|\lambda \text{div}\,(p) - f\|_X^2 : p \ / \ |p_{i,j}| \leq 1 \ \forall i, j = 1, \ldots, N \right\} \tag{3.78}$$

(3.78) can be solved with a fixed point process:

$$p^0 = 0 \tag{3.79}$$

and

$$p_{i,j}^{n+1} = \frac{p_{i,j}^n + \tau(\nabla(\text{div}\,(p^n) - f/\lambda))_{i,j}}{1 + \tau|(\nabla(\text{div}\,(p^n) - f/\lambda))_{i,j}|} \tag{3.80}$$

And [24] gives a sufficient condition for the algorithm to converge:

Theorem 3.1. *Assume that parameter τ in (3.80) is such that $\tau \leq 1/8$. Then $\lambda div(p^n)$ converges to $P_{\lambda K}(f)$ when $n \to +\infty$.*

The solution to problem (3.67) is therefore given by:

$$u = f - \lambda \text{div}\,(p^\infty) \tag{3.81}$$

where $p^\infty = \lim_{n \to +\infty} p^n$

Figure 11: Image bruitée à restaurer

Projected gradient algorithm :

Notice that alternatively a projected gradient method can be used. Indeed, algorithm (3.80) can be rewritten as:

$$\begin{cases} v^m = \frac{f}{\lambda} + \operatorname{div} p^m \\ p_{i,j}^{m+1} = \frac{p_{i,j}^m + \tau(\nabla v^m)_{i,j}}{1+\tau|(\nabla v^m)_{i,j}|} \end{cases} \tag{3.82}$$

and λv^m converges to the solution of (3.67).

Instead of using (3.82), a simple gradient descent/retroprojection method can be used:

$$\begin{cases} v^m = \frac{f}{\lambda} + \operatorname{div} p^m \\ p_{i,j}^{m+1} = \frac{p_{i,j}^m + \tau(\nabla v^m)_{i,j}}{\max\left\{1, |p_{i,j}^n + \tau(\nabla v^m)_{i,j}|\right\}} \end{cases} \tag{3.83}$$

Such a scheme is proved to be convergent in [14].

Proposition 3.5. *If $\tau < \frac{1}{4}$, then the sequence (v^m, p^m) defined by scheme (3.83) is such that $v^m \to v$ and $p^m \to p$ with λv solution of (3.67).*

Notice that (3.83) can be written in a more compact way:

$$p^{m+1} = P_K\left(p^m + \tau\nabla\left(\frac{f}{\lambda} + \operatorname{div} p^m\right)\right) \tag{3.84}$$

Figure 12 shows an example of restoration with Chambolle's algorithm (the noisy image is displaid in Figure 9).

3.3.3 Euler-Lagrange equation for (3.67):

The optimality condition associated to (3.67) is:

$$u - f \in \lambda \partial J(u) \tag{3.85}$$

And formally, one then writes:

$$u - f = \lambda \operatorname{div}\left(\frac{\nabla u}{|\nabla u|}\right) \tag{3.86}$$

Figure 12: Image restaurée (ROF)

But the subdifferential $\partial J(u)$ cannot always be written this way.

The following result (see Proposition 1.10 in [6] for further details) gives more details about the subdifferential of the total variation.

Proposition 3.6. *Let (u,v) in $L^2(\Omega)$ with u in $BV(\Omega)$. The following assertions are equivalent:*

(i) $v \in \partial J(u)$.

(ii) Denoting by $X(\Omega)_2 = \{z \in L^\infty(\Omega, \mathbb{R}^2) \ : \ \mathrm{div}\,(z) \in L^2(\Omega)\}$, we have:

$$\int_\Omega vu\,dx = J(u) \tag{3.87}$$

and

$$\begin{array}{c} \exists z \in X(\Omega)_2 \ , \ \|z\|_\infty \leq 1 \ , z.N = 0 \ , \ \text{on } \partial\Omega \\ \text{such that } v = -\mathrm{div}\,(z) \ \text{in } \mathcal{D}'(\Omega) \end{array} \tag{3.88}$$

(iii) (3.88) holds and:

$$\int_\Omega (z, Du) = \int_\Omega |Du| \tag{3.89}$$

From this proposition, we see that (3.85) means:

$$u - f = \lambda \mathrm{div}\, z \tag{3.90}$$

with z satisfying (3.88) and (3.89). This is a rigorous way to write $u - f = \lambda \mathrm{div}\left(\frac{\nabla u}{|\nabla u|}\right)$.

3.3.4 Other regularization choices

The drawback of the total variation regularization is a *staircase* effect. There has therefore been a lot of work dealing with how to remedy to this problem. In particular, people have investigated other regularization choice of the kind $L(u) = \int_\Omega \phi(|\nabla u|)$. The functional to minimize becomes thus:

$$\inf_u \frac{1}{2}\|f - Au\|^2 + \lambda \int_\Omega \phi(\|\nabla u\|)\,dx \tag{3.91}$$

And formally, the associated Euler-Lagrange equation is:

$$-\lambda \text{div}\left(\frac{\phi^{'}(|\nabla u|)}{|\nabla u|}\nabla u\right)+A^*Au-A^*f=0 \tag{3.92}$$

with Neumann boundary conditions:

$$\frac{\phi^{'}(|\nabla u|)}{|\nabla u|}\frac{\partial u}{\partial N}=0 \text{ on } \partial\Omega. \tag{3.93}$$

We are now going to develop formally the divergence term. For each point x where $|\nabla u(x)| \neq 0$, we can define the vectors $N(x)=\frac{\nabla u(x)}{|\nabla u(x)|}$ and $T(x)$, $|T(x)|=1$, with $T(x)\perp N(x)$, respectively the normal and the tangent to the level line of u.

We can then rewrite (3.92) as:

$$A^*Au-\lambda\left(\frac{\phi^{'}(|\nabla u|)}{|\nabla u|}u_{TT}+\phi^{''}(|\nabla u|)u_{NN}\right)=A^*f \tag{3.94}$$

where we denote by u_{TT} and u_{NN} the second derivatives of u in the T direction and N direction, respectively:

$$u_{TT}=T^*\nabla^2uT=\frac{1}{|\nabla u|^2}\left(u_{x_1}^2u_{x_2x_2}+u_{x_2}^2u_{x_1x_1}-2u_{x_1}u_{x_2}u_{x_1x_2}\right) \tag{3.95}$$

$$u_{NN}=N^*\nabla^2uN=\frac{1}{|\nabla u|^2}\left(u_{x_1}^2u_{x_1x_1}+u_{x_2}^2u_{x_2x_2}+2u_{x_1}u_{x_2}u_{x_1x_2}\right) \tag{3.96}$$

This allows to see clearly the action of the function ϕ in both directions N and T.
• At the location, where the variation of the intensity are weak (low gradient), we would like to encourage smoothing, the same in all directions.

Assuming that ϕ is regular, this isotropic smoothing can be achieved by imposing:

$$\phi^{'}(0)=0\ ,\ \lim_{s\to 0^+}\frac{\phi^{'}(s)}{s}=\lim_{s\to 0^+}\phi^{''}(s)=\phi^{''}(0)>0 \tag{3.97}$$

Therefore, at points where $|\nabla u|$ is small, (3.94) becomes:

$$A^*Au-\lambda\phi^{''}(0)(\underbrace{u_{TT}+u_{NN}}_{=\Delta u})=A^*f \tag{3.98}$$

So at these points, we want to do some Tychonov regularization.
• In a neighbourhood of an edge C, the image presents a strong gradient. If we want to preserve this edge, it is preferable to diffuse along C (in the T direction) and not across it. To do so, it is sufficient in (3.94) to annihilate (for strong gradients) the coefficient of u_{NN}, and to assume that the coefficient of u_{TT} does not vanish:

$$\lim_{s\to+\infty}\phi^{''}(s)=0\ ,\ \lim_{s\to+\infty}\frac{\phi^{'}(s)}{s}=\beta>0 \tag{3.99}$$

Unfortunately, this two conditions ar not compatible. A trade-off mus be found. For instance, $\phi^{''}(s)$ and $\frac{\phi^{'}(s)}{s}$ both converge to 0 as $s\to+\infty$, but at different rates:

$$\lim_{s\to+\infty}\phi^{''}(s)=\lim_{s\to+\infty}\frac{\phi^{'}(s)}{s}=0 \text{ and } \lim_{s\to+\infty}\frac{\phi^{''}(s)}{\frac{\phi^{'}(s)}{s}}=0 \tag{3.100}$$

For instance, one may choose the hypersurface minimal function:

$$\phi(s) = \sqrt{1+s^2} \tag{3.101}$$

Of course, all these remarks are qualitative. Other conditions will arise so that the problem is mathematically well-posed.

To show the existence and uniqueness of a solution using the direct method of the calculus of variations, some minimal hypotheses are needed on ϕ:

(i) ϕ is strictly convex, nondecreasing function from $\mathbb{R}^+$ to $\mathbb{R}^+$, with $\phi(0) = 0$ (without a loss of generality).

(ii) $\lim_{s\to+\infty} \phi(s) = +\infty$.

Conditions (ii) must not be too strong, because it must not penalize strong gradients, i.e. the formation of edges (see what happens if $\phi(s) = s^2$). Hence we assume that ϕ grows at most linearly: there exist two constants $c > 0$ and $b \geq 0$ such that:

$$cs - b \leq \phi(s) \leq cs + b \text{ for all } s \geq 0 \tag{3.102}$$

With all these assumptions, it then possible to show that problem (3.91) admits a unique solution in $BV(\Omega)$ (see [9]).

Non convex ϕ function: It has been shown numerically that the choice of non convex ϕ functions can lead to very interisting results [9]. Nevertheless, in the continuous case, the direct method of calculus of variation fails to prove the existence of a solution for such regularization choices. This remains an open question. In particular, the following functions have been shown to give good restoration results:

$$\phi(s) = \frac{s^2}{1+s^2} \tag{3.103}$$

and:

$$\phi(s) = \log(1+s^2) \tag{3.104}$$

3.3.5 Half quadratic minimmization

$$J(u) = \lambda \int_\Omega \phi(|\nabla u|)\, dx + \frac{1}{2}\|f - Au\|^2 \tag{3.105}$$

with ϕ edge preserving function [9]. ϕ C^1 and convex. Typically:

$$\phi(x) = \sqrt{\epsilon^2 + x^2} \tag{3.106}$$

We have already seen before that:

$$\nabla J(u) = -\lambda \text{div}\left(\frac{\phi'(|\nabla u|)}{|\nabla u|}\nabla u\right) + A^*(Au - f) \tag{3.107}$$

Notice that in the case when $\phi(x) = \sqrt{\epsilon^2 + x^2}$, then $\phi'(x) = \frac{x}{\sqrt{\epsilon^2+x^2}}$, and therefore $\frac{\phi'(|\nabla u|)}{|\nabla u|} = \frac{1}{\sqrt{\epsilon^2+|\nabla u|^2}}$.

Fixed step gradient descent The easiest method is a fixed step gradient descent [65].

$$u_{k+1} = u_k - \delta t \nabla J(u_k) \tag{3.108}$$

Nevertheless, it is quite slow: indeed, for the algorithm to converge , δt needs to be very small. In particular, if $\phi(x) = \sqrt{\epsilon^2 + x^2}$, then a typical choice for δt is $\delta t = \epsilon/10$.

Quasi-Newton method To improve the speed of the above algorithm, it has been tried to use Newton method:

$$u_{k+1} = u_k - (\nabla^2 J(u_k))^{-1} \nabla J(u_k) \tag{3.109}$$

But in practice this is not possible.

In the case when $\phi(x) = \sqrt{\epsilon^2 + x^2}$, we have

$$\nabla J(u) = -\lambda \mathrm{div} \left(\frac{1}{\sqrt{\epsilon^2 + |\nabla u|^2}} \nabla u \right) + A^*(Au - f) \tag{3.110}$$

We are interested in solving $\nabla J(u) = 0$. Since we do not know how to invert ∇J, we want to use a fixed point process. We linearize the equation (quasi-Newton method):

$$0 = -\lambda \mathrm{div} \left(\frac{1}{\sqrt{\epsilon^2 + |\nabla u_k|^2}} \nabla u_{k+1} \right) + A^*(Au_{k+1} - f) \tag{3.111}$$

It is natural to look at a semi-explicit scheme, explicit in the non linear part and implicit in the linear part (general idea of Weiszfeld method [73]), i.e.:

$$u_{k+1} = \left(A^*A. - \lambda \mathrm{div} \left(\frac{1}{\sqrt{\epsilon^2 + |\nabla u_k|^2}} \nabla . \right) \right)^{-1} A^* f \tag{3.112}$$

It can be shown that such a scheme converges to the solution [1, 70, 34, 27, 3]. Moreover, it is much faster than a gradient descent (it is possible to show theoretically a linear convergence, but in practice a quadratic convergence is observed).

For a general ϕ C^1 function the scheme is:

$$0 = -\lambda \mathrm{div} \left(\frac{\phi'(|\nabla u_k|}{\phi(|\nabla u_k|)} \nabla u_{k+1} \right) + A^*(Au_{k+1} - f) \tag{3.113}$$

$$u_{k+1} = \left(A^*A. - \lambda \mathrm{div} \left(\frac{\phi'(|\nabla u_k|}{\phi(|\nabla u_k|)} \nabla . \right) \right)^{-1} A^* f \tag{3.114}$$

This is in fact a special case of a more general method: the *half quadratic minimization* approach [9].

Half quadratic minimization approach It is based on the following result.

Proposition 3.7. *Let $\phi : \mathbb{R}_+ \to \mathbb{R}_+$ a non decreasing function such that $\phi(\sqrt{t})$ concave. Let us define:*

$$L = \lim_{t \to +\infty} \frac{\phi'(t)}{2t} \text{ and } M = \lim_{t \to 0^+} \frac{\phi'(t)}{2t} \tag{3.115}$$

Then there exists a convex and decreasing function $\psi : (L, M] \to [\beta_1, \beta_2]$ such that:

$$\phi(s) = \inf_{L \le b \le M} (bs^2 + \psi(b)) \tag{3.116}$$

where $\beta_1 = \lim_{s \to 0^+} \phi(s)$ and $\beta_2 = \lim_{s \to +\infty} (\phi(s) - s\phi'(s)/2)$.

Moreover, the value b for which the minimum is reached is given by:

$$b = \frac{\phi'(s)}{2s} \tag{3.117}$$

The additional variable b is usually called the dual variable.

Proof: Let $\theta(s) = -\phi(\sqrt{s})$. θ is a convex function. Thus

$$\theta(s) = \theta^{**}(s) = \sup_{s^*} (ss^* - \theta^*(s^*)) \tag{3.118}$$

where $\theta^*(s^*)$ is the polar funtion of $\theta(s)$ defined by: $\theta^*(s^*) = \sup_s (ss^* - \theta(s))$. Therefore:

$$\phi(\sqrt{s}) = \inf_{s^*} (-ss^* + \theta^*(s^*)) \tag{3.119}$$

Let $b = -s^*$ and $s = \sqrt{s}$. Then ϕ can be written:

$$\phi(s) = \inf_{s^*} (bs^2 + \theta^*(-b)) \tag{3.120}$$

which gives the first part of the theorem with

$$\psi(b) = \theta^*(-b) \tag{3.121}$$

The rest of the proof is standard.

■

Remark: Let us emphasize that in the above proposition, ϕ can be non convex.

For instance, when $\phi(s) = \frac{s^2}{1+s^2}$, then $\psi(b) = (\sqrt{b} - 1)^2$, $L = 0$, and $M = 1$.

Application:

$$J(u) = \lambda \int_\Omega \phi(|\nabla u|)\, dx + \frac{1}{2}\|f - Au\|^2 \tag{3.122}$$

Assume ϕ satisfies the hypotheses of the previous proposition. Then we have:

$$J(u) = \lambda \int_\Omega \inf_{L \le b \le M} (bs^2 + \psi(b))|\nabla u|\, dx + \frac{1}{2}\|f - Au\|^2 \tag{3.123}$$

Supposing we can invert $\int_\Omega$ and the infimum (this can be justified):

$$\begin{aligned} \inf_u J(u) &= \inf_u \inf_b \left(\lambda \int_\Omega (b|\nabla u|^2 + \psi(b))\, dx + \frac{1}{2}\|f - Au\|^2 \right) \\ &= \inf_b \inf_u \left(\lambda \int_\Omega (b|\nabla u|^2 + \psi(b))\, dx + \frac{1}{2}\|f - Au\|^2 \right) \end{aligned}$$

Let us define

$$F(u,b) = \lambda \int_\Omega (b|\nabla u|^2 + \psi(b))\, dx + \frac{1}{2}\|f - Au\|^2 \tag{3.124}$$

This functional is convex in u, and for each u fixed it is convex in b (but the functional is not convex with respect to the pair (u,b)).

It can be shown that alternating minmizations with respect to u and v, then u_k converges to the minimizer u of the original functional [9, 28, 25]. Moreover, this is a much faster algorithm than the fixed step gradient algorithm.

The sequence b_k can be seen as an indicator of contours. If ϕ satisfies the edge-preserving hypotheses $\lim_{s\to+\infty} \phi'(s)/(2s) = 0$ and $\lim_{s\to+0^+} \phi'(s)/(2s) = 1$, then the following conditions are satisfied:

- If $b_k(x) = 0$, then x belongs to a contour.
- If $b_k(x) = 1$, then x belongs to a homogeneous region.

At each iteration, we define:

$$u_{k+1} = \operatorname*{argmin}_u F(u, b_k) \tag{3.125}$$

and

$$b_{k+1} = \operatorname*{argmin}_b F(u_{k+1}, b) \tag{3.126}$$

We have

$$b_{k+1} = \frac{\phi'(|\nabla u_{k+1}|)}{2|\nabla u_{k+1}|} \tag{3.127}$$

and u_{k+1} is the solution of:

$$0 = -\lambda \text{div}\ (2b_k \nabla u) + A^*(Au - f) \tag{3.128}$$

which can be solved with a conjugate gradient algorithm.

Notice that in the case when $\phi(x) = \sqrt{\epsilon^2 + x^2}$, this is in fact exactly the same scheme as above.

3.4 Wavelets

3.4.1 Besov spaces

We denote by $\{\psi_{j,k}\}$ a wavelet basis. A function f in $L^2(\mathbb{R}^2)$ can be written:

$$f = \sum_{j,k} c_{j,k} \psi_{j,k} \tag{3.129}$$

where the $c_{j,k}$ are the wavelet coefficients of f, and we have: $\|f\|_{L^2(\mathbb{R}^2)} = \sum_{j,k} c_{j,k}^2$.

Spaces weel-suited to wavelets are Besov spaces $B^s_{p,q}$ (for $0 < s < \infty$, $0 < p \le \infty$ and $0 < q \le \infty$) [57, 55, 30, 25]. $B^s_{p,q}$ correponds roughly to functions with s derivatives in $L^p(\mathbb{R}^2)$, the third parameter q being a way to adjust the regularity with precision.

Remark: if $p = q = 2$, then $B^s_{2,2}$ is the Sobolev space $W^{s,2}$, and when $s < 1$, $1 \leq p \leq \infty$, and $q = \infty$, then $B^s_{p,\infty}$ is the Lipschitz space $Lip(s, L^p(\mathbb{R}^2))$.

We can give an intrinsic definition to Besov spaces $B^s_{p,q}$ and of their norm $\|.\|_{B^s_{p,q}}$ from the regularity modulus of f [30, 25]. If we assume that the chosen wavelet ψ has at least $s+1$ vanishing moments and is of regularity at least C^{s+1}, then if $f \in B^s_{p,q}$, the norm $\|f\|_{B^s_{p,q}}$ is equivalent to:

$$\left(\sum_k \left(\sum_j 2^{skp} 2^{k(p-2)} |c_{j,k}|^p \right)^{\frac{p}{q}} \right)^{\frac{1}{q}} \tag{3.130}$$

(the constants depend on the chosen wavelet).

In what follows, we will always use the equivalent norm (3.130) for $\|f\|_{B^s_{p,q}}$.

Here, we are interested in homogeneous version of Besov spaces:

$$\dot{B^s_{p,q}} = B^s_{p,q} / \left\{ u \in B^s_{p,q} \ / \ \nabla u = 0 \right\} \tag{3.131}$$

Definition 3.1. $\dot{B}^1_{1,1}$ is the usual homogeneous Besov space (cf [57]). Let $\psi_{j,k}$ an orthonormal wavelet basis composed of regular wavelets with compact supports. $\dot{B}^1_{1,1}$ is a subspace of $L^2(\mathbb{R}^2)$ and a function f belongs to $\dot{B}^1_{1,1}$ if and only of:

$$\sum_{j \in \mathbb{Z}} \sum_{k \in \mathbb{Z}^2} |c_{j,k}| 2^{j/2} < +\infty \tag{3.132}$$

Definition 3.2. The dual space of $\dot{B}^1_{1,1}$ is the Banach space $\dot{B}^\infty_{-1,\infty}$. It is characterized by the fact that the wavelet function of a generalized function in $\dot{B}^\infty_{-1,\infty}$ are in $l^\infty(\mathbb{Z} \times \mathbb{Z}^2)$.

Remark: We have the following inclusions:

$$\dot{B}^1_{1,1} \subset \dot{BV}(\mathbb{R}^2) \subset L^2(\mathbb{R}^2) \subset \dot{B}^\infty_{-1,\infty} \tag{3.133}$$

where $\dot{BV}$ stands for the homogeneous version of BV : $\dot{BV} = BV / \{u \in BV \ / \ \nabla u = 0\}$.

3.4.2 Wavelet shrinkage

A interesting application of wavelets is image denoising. If an original image u has been degraded by some additive white gaussian noise, an efficient restoration method consists in thresholding the wavelet coefficients of the degraded image f.

We define the soft-thresholding operator as:

$$\theta_\tau(t) = \begin{cases} t - \tau \text{ if } t \geq \tau \\ 0 \text{ if } t \leq \tau \\ t + \tau \text{ si } t \leq -\tau \end{cases} \tag{3.134}$$

In an orthonormal wavelet basis, the wavelet coefficients of f denoted by $c_{j,k}(f)$ are random gaussian variables with zero mean with standard deviation σ (σ being the standard deviation of the white gaussian noise).

The wavelet soft-thresholding of f with parameter τ, denoted by $WST(f, \tau)$ (Wavelet Soft Thresholding), is the function whose wavelet coefficients are $\theta_\tau(c_{j,k}(f))$. The theoretical value proposed by Donoho is $\tau = \sigma\sqrt{2\log(N^2)}$, where N^2 stands for the size of the image (in practice, this threshold value is much too large). For further details, we refer the reader to [35, 53, 57, 55].

Figure 13: Restoration by non invariant wavelet shrinkage (Haar)

3.4.3 Variational interpretation

Let us consider the functional

$$\inf_u \|f-u\|^2 + 2\tau \|u\|_{\dot{B}^1_{1,1}} \tag{3.135}$$

The solution to (3.135) is given by:

$$u = WST(f,\tau) \tag{3.136}$$

Sketch of the proof: (see [25] for the detailed proof)

Denote by $c_{j,k}$ (resp. $d_{j,k}$) the wavelet coefficients of f (resp. u). We thus have to minimize:

$$\sum_{j,k} \left(|c_{j,k} - d_{j,k}|^2 + 2\tau |d_{j,k}| \right) \tag{3.137}$$

There is no coupling term in the equations, and we therefore just have to minimize the generic function

$$E(s) = |s-t|^2 + 2\tau|s| = |s|^2 + 2|s|(\tau - |t|) + t^2 \tag{3.138}$$

We minimize $f(x) = x^2 + 2x(\tau - |t|) + t^2$ with the constraint $x \geq 0$. We have $f'(x) = 2x + 2(\tau - |t|)$.

■

Figure 13 and 14 show examples of restoration.

Remark: In 1D, total variation minimization is equivalent to iterative wavelet shrinkage (using the Harr wavelet with one level of decomposition) [67].

Figure 14: Restoration by invariant wavelet shrinkage (Haar)

4. Advanced topics: Image decomposition

We encourage the reader to look at [57] for a nice historical introduction to the topics.

4.1 Introduction

Image restoration is an important and challenging inverse problem in image analysis. The problem consists in reconstructing an image u from a degraded data f. The most common model linking u to f is the following one: $f = Ru + v$, where R is a linear operator typically modeling blur and v is the noise. Energy minimization has demonstrated to be a powerful approach to tackle this kind of problem (see [9] and references therein for instance). Here we examine a pure denoising situation, i.e. R is the identity operator. The underlying energy is generally composed of two terms: a fidelity term to the data and a regularizing-cost function. One of the most effective method is the total variation minimization as proposed in [65]. This model relies on the assumption that $BV(\Omega)$, the space of functions with bounded variation, is a good space to study images (even if it is known that such an assumption is too restrictive [4]). In [65], the authors decompose an image f into a component u belonging to $BV(\Omega)$ and a component v in $L^2(\Omega)$. In this model v is supposed to be the noise. In such an approach, they minimize:

$$\inf_{(u,v)\in BV(\Omega)\times L^2(\Omega)/f=u+v} \left(\int |Du| + \frac{1}{2\lambda} \|v\|^2_{L^2(\Omega)} \right) \tag{4.1}$$

where $\int |Du|$ stands for the total variation of u. In practice, they compute a numerical solution of the Euler-Lagrange equation associated to (4.1). The mathematical study of (4.1) has been done in [25].

In [57], Y. Meyer shows some limitations of the model proposed in [65]. In particular, if f is the characteristic function of a bounded domain with a C^∞-boundary, then f is not preserved by the Rudin-Osher-Fatemi model (contrary to what should be expected).

Meyer model In [57], Y. Meyer suggests a new decomposition. He proposes the following model:

$$\inf_{(u,v)\in BV(\mathbb{R}^2)\times G(\mathbb{R}^2)/f=u+v}\left(\int |Du| + \alpha\|v\|_{G(\mathbb{R}^2)}\right) \tag{4.2}$$

where the Banach space $G(\mathbb{R}^2)$ contains signals with large oscillations, and thus in particular textures and noise. We give here the definition of $G(\mathbb{R}^2)$.

Definition 4.1. $G(\mathbb{R}^2)$ is the Banach space composed of distributions f which can be written

$$f = \partial_1 g_1 + \partial_2 g_2 = \mathrm{div}\,(g) \tag{4.3}$$

with g_1 and g_2 in $L^\infty(\mathbb{R}^2)$. The space $G(\mathbb{R}^2)$ is endowed with the following norm:

$$\begin{aligned} \|v\|_{G(\mathbb{R}^2)} \quad = \quad \inf\Big\{ & \|g\|_{L^\infty(\mathbb{R}^2)} = \operatorname*{ess\,sup}_{x\in\mathbb{R}^2} |g(x)| \;/ v = \mathrm{div}\,(g),\ g = (g_1, g_2), \\ & g_1 \in L^\infty(\mathbb{R}^2), g_2 \in L^\infty(\mathbb{R}^2), |g(x)| = \sqrt{(|g_1|^2 + |g_2|^2)(x)}\Big\} \end{aligned} \tag{4.4}$$

$BV(\mathbb{R}^2)$ has no simple dual space (see [5]). However, as shown by Y. Meyer [57], $G(\mathbb{R}^2)$ is the dual space of the closure in $BV(\mathbb{R}^2)$ of the Schwartz class. So it is very related to the dual space of $BV(\mathbb{R}^2)$. This is a motivation to decompose a function f on $BV(\mathbb{R}^2) + G(\mathbb{R}^2)$. This is also why the divergence operator naturally appears in the definition of $G(\mathbb{R}^2)$, since the gradient and the divergence operators are dual operators.

A function belonging to G may have large oscillations and nevertheless have a small norm. Thus the norm on G is well-adapted to capture the oscillations of a function in an energy minimization method.

4.2 A space for modeling oscillating patterns in bounded domains

4.2.1 Definition and properties

In all the sequel, we denote by Ω a bounded connected open set of $\mathbb{R}^2$ with a Lipschitz boundary. We adapt Definition 4.1 concerning the space G to the case of Ω. We are going to consider a subspace of the Banach space $W^{-1,\infty}(\Omega) = \left(W_0^{1,1}(\Omega)\right)'$ (the dual space of $W_0^{1,1}(\Omega)$).

Definition 4.2. $G(\Omega)$ is the subspace of $W^{-1,\infty}(\Omega)$ defined by:

$$G(\Omega) = \left\{ v \in L^2(\Omega) \;/\; v = \mathrm{div}\,\xi \;,\; \xi \in L^\infty(\Omega, \mathbb{R}^2) \;,\; \xi.N = 0 \text{ on } \partial\Omega \right\} \tag{4.5}$$

On $G(\Omega)$, the following norm is defined:

$$\|v\|_{G(\Omega)} = \inf\left\{ \|\xi\|_{L^\infty(\Omega,\mathbb{R}^2)} \;/ v = \mathrm{div}\,\xi \;,\; \xi.N = 0 \text{ on } \partial\Omega \right\} \tag{4.6}$$

Remark: In Definition 4.2, since $\mathrm{div}\,\xi \in L^2(\Omega)$ and $\xi \in L^\infty(\Omega, \mathbb{R}^2)$, we can define $\xi.N$ on $\partial\Omega$ (in this case, $\xi.N \in H^{-1/2}(\partial\Omega)$, see [68, 52] for further details).

The next lemma was stated in [57]. Using approximations with $C_c^\infty(\Omega)$ functions [5], the proof is straightforward:

Lemma 4.1. *Let $u \in BV(\Omega)$ and $v \in G(\Omega)$. Then: $\int_\Omega uv \leq J(u)\|v\|_{G(\Omega)}$ (where $J(u)$ is defined by (2.70)).*

We have the following simple characterization of $G(\Omega)$:

Proposition 4.1.

$$G(\Omega) = \left\{ v \in L^2(\Omega) \ / \ \int_\Omega v = 0 \right\} \tag{4.7}$$

Proof: Let us denote by $H(\Omega)$ the right-hand side of (4.7). We split the proof into two steps.
Step 1: Let v be in $G(\Omega)$. Then from(4.5) it is immediate that $\int_\Omega v = 0$, i.e. $v \in H(\Omega)$.
Step 2: Let v be in $H(\Omega)$. Then from [21] (Theorem 3') (see also [22]), there exists $\xi \in C^0(\bar\Omega, \mathbb{R}^2) \cap W^{1,2}(\Omega, \mathbb{R}^2)$ such that $v = \operatorname{div} \xi$ and $\xi = 0$ on $\partial\Omega$. In particular, we have $\xi \in L^\infty(\Omega, \mathbb{R}^2)$ and $\xi.N = 0$ on $\partial\Omega$. Thus $v \in G(\Omega)$.

■

Remark: Let us stress here how powerful the result in [22, 21] is. It deals with the limit case v in $L^q(\Omega)$, $q = 2$, when the dimension of the space is $N = 2$. The classical method for tackling the equation $\operatorname{div} \xi = v$ with $\xi.N = 0$ on $\partial\Omega$ consists in solving the problem $\Delta u = v$ with $\frac{\partial u}{\partial N} = 0$ on $\partial\Omega$, and in setting $\xi = \nabla u$. If v is in $L^q(\Omega)$ with $q > 2$ this problem admits a unique solution (up to a constant) in $W^{2,q}(\Omega)$. Moreover, thanks to standard Sobolev embeddings (see [37, 40]), $\xi = \nabla u$ belongs to $L^\infty(\Omega, \mathbb{R}^2)$. If $q = 2$, the result is not true and the classical approach does not work. So the result by Bourgain and Brezis is very sharp.

We next introduce a family of convex subsets of $G(\Omega)$. These convex sets will be useful for approximating Meyer problem.

Definition 4.3. Let $G_\mu(\Omega)$ the family of subsets defined by ($\mu > 0$):

$$G_\mu(\Omega) = \left\{ v \in G(\Omega) \ / \ \|v\|_{G(\Omega)} \leq \mu \right\} \tag{4.8}$$

Lemma 4.2. *$G_\mu(\Omega)$ is closed for the $L^2(\Omega)$-strong topology.*

Proof of Lemma 4.2 Let (v_n) be a sequence in $G_\mu(\Omega)$ such that there exists $\hat{v} \in L^2(\Omega)$ with $v_n \to \hat{v}$ in $L^2(\Omega)$-strong. We have $v_n = \operatorname{div} \xi_n$, with ξ_n such that $\|\xi_n\|_{L^\infty(\Omega,\mathbb{R}^2)} \leq \mu$ and $\xi_n.N = 0$ on $\partial\Omega$. As $\|\xi_n\|_{L^\infty(\Omega,\mathbb{R}^2)} \leq \mu$, there exists $\hat{\xi} \in L^\infty(\Omega, \mathbb{R}^2)$ such that, up to an extraction: $\xi_n \rightharpoonup \hat{\xi}$ in $L^\infty(\Omega, \mathbb{R}^2)$ weak $*$, and $\|\hat{\xi}\|_{L^\infty(\Omega,\mathbb{R}^2)} \leq \mu$.

Moreover if $\phi \in \mathcal{D}(\bar\Omega)$: $\int_\Omega v_n \phi \, dx = \int_\Omega \operatorname{div} \xi_n \phi \, dx = -\int_\Omega \xi_n \nabla\phi \, dx$. Thus as $n \to +\infty$, we get:

$$\int_\Omega \hat{v}\phi \, dx = -\int_\Omega \hat{\xi} \nabla\phi \, dx = \int_\Omega \operatorname{div} \hat{\xi} \, \phi \, dx - \int_{\partial\Omega} \hat{\xi}.N\phi \tag{4.9}$$

By choosing first a test function in $C_c^\infty(\Omega)$, we deduce from (4.9) that $\hat{v} = \operatorname{div} \hat{\xi}$ in $\mathcal{D}'(\Omega)$, and since $\hat{v} \in L^2(\Omega)$, the equality holds in $L^2(\Omega)$. Then for a general $\phi \in \mathcal{D}(\bar\Omega)$, it comes $\hat{\xi}.N = 0$ on $\partial\Omega$ (in $H^{-1/2}(\partial\Omega)$).

■

The next result is a straightforward consequence of Lemma 4.2.

Corollary 4.1. *The indicator function of $G_\mu(\Omega)$ is lsc (lower-semicontinuous) for the $L^2(\Omega)$-strong topology (and for the $L^2(\Omega)$-weak topology since G_μ is convex).*

Remarks:

1. Let us denote by $K(\Omega)$ the closure in $L^2(\Omega)$ of the set:

$$\left\{ \operatorname{div} \xi \ , \ \xi \in C_c^\infty(\Omega, \mathbb{R}^2) \ , \ \|\xi\|_{L^\infty(\Omega,\mathbb{R}^2)} \leq 1 \right\} \tag{4.10}$$

   Using Lemma 4.2 and some results in [68], one can prove that $K(\Omega) = G_1(\Omega)$.

   Moreover, one can also show in the same way that $G(\Omega)$ is the closure in $L^2(\Omega)$ of the set:

$$\left\{ \operatorname{div} \xi \ , \ \xi \in C_c^\infty(\Omega, \mathbb{R}^2) \ \right\} \tag{4.11}$$

2. From the proof of Lemma 4.2, one easily deduces that $\|.\|_G$ is lower semi continuous (lsc).

We also have the following result:

Lemma 4.3. *If $v \in G(\Omega)$, then there exists $\xi \in L^\infty(\Omega, \mathbb{R}^2)$ with $v = \operatorname{div} \xi$ and $\xi.N = 0$ on $\partial\Omega$, and such that $\|v\|_G = \|\xi\|_{L^\infty(\Omega,\mathbb{R}^2)}$.*

Proof: Let $v \in G(\Omega)$. Let us consider a sequence $\xi_n \in L^\infty(\Omega, \mathbb{R}^2)$ with $v = \operatorname{div} \xi_n$ and $\xi_n.N = 0$ on $\partial\Omega$, and such that $\|\xi_n\|_{L^\infty(\Omega)} \to \|v\|_G$. There exists $\xi \in L^\infty(\Omega, \mathbb{R}^2)$ such that, up to an extraction, $\xi_n \rightharpoonup \xi$ in $L^\infty(\Omega, \mathbb{R}^2)$ weak *. Then, as in the proof of Lemma 4.2, we can show that $\xi.N = 0$ on $\partial\Omega$ and that $v = \operatorname{div} \xi$.

■

Main property: The following lemma is due to Y. Meyer [57]. But it was stated in the case of $\Omega = \mathbb{R}^2$, and the proof relied upon harmonic analysis tools. Thanks to our definition of $G(\Omega)$, we formulate it in the case when Ω is bounded. Our proof relies upon functional analysis arguments.

Lemma 4.4. *Let Ω be a Lipschitz bounded open set, and let f_n , $n \geq 1$ be a sequence of functions in $L^q(\Omega) \bigcap G(\Omega)$ with the following two properties:*

1. *There exists $q > 2$ and $C > 0$ such that $\|f_n\|_{L^q(\Omega)} \leq C$.*

2. *The sequence f_n converges to 0 in the distributional sense (i.e. in $\mathcal{D}'(\Omega)$).*

   Then $\|f_n\|_G$ converges to 0 when n goes to infinity.

This result explains why the norm in $G(\Omega)$ is a good norm to tackle signals with strong oscillations. It will be easier with this norm to capture such signals in a minimization process than with a classical L^2-norm.

Remark: Hypotheses 1. and 2. are equivalent to the simpler one: *there exists $q > 2$ such that $f_n \rightharpoonup 0$ in $L^q(\Omega)$-weak.*

Proof of Lemma 4.4: Let us consider a sequence $f_n \in L^q(\Omega) \bigcap G(\Omega)$ satisfying assumption 1. and let us define the Neumann problem:

$$\begin{cases} \Delta u_n = f_n \ in \ \Omega \\ \frac{\partial u_n}{\partial N} = 0 \ on \ \partial\Omega \end{cases} \tag{4.12}$$

We recall that as $f_n \in G(\Omega)$, we also have $\int_\Omega f_n \, dx = 0$. We know (see [46, 59, 33]) that problem (4.12) admits a solution $u_n \in W^{2,q}(\Omega)$. From [59, 58], we also know that there exists a constant $B > 0$ such that: $\|u_n\|_{W^{2,q}(\Omega)} \leq B\|f_n\|_{L^q(\Omega)}$. And as we assume that $\|f_n\|_{L^q(\Omega)} \leq C$, we get:

$$\|u_n\|_{W^{2,q}(\Omega)} \leq BC \tag{4.13}$$

Since $q > 2$ and Ω bounded, we know (see [2]) that there exists $\theta \in (0,1)$ such that $W^{2,q}(\Omega)$ is compactly embedded in $C^{1,\theta}(\Omega)$. We denote by $g_n = \nabla u_n$. We have $\|g_n\|_{W^{1,q}(\Omega)} \leq \|u_n\|_{W^{2,q}(\Omega)} \leq BC$. And it is also standard that $W^{1,q}(\Omega)^2$ is compactly embedded in $C^{0,\theta}(\Omega)^2$.

Hence, up to an extraction, we get that there exists u and $g \in C^{0,\theta}$ such that $u_n \to u$ and $g_n \to g$ (for the $C^{0,\theta}$ topology). It is then standard to pass to the limit in (4.12) to deduce that $g_n \to 0$ uniformly (we recall that $g_n = \nabla u_n$). The previous reasonning being true for any subsequence extracted from u_n, we conclude that the whole sequence ∇u_n is such that $\nabla u_n \to 0$ as $n \to +\infty$ in $L^\infty(\Omega, \mathbb{R}^2)$-strong, i.e. $g_n = \nabla u_n \to 0$ in $L^\infty(\Omega, \mathbb{R}^2)$-strong. Since $f_n = \operatorname{div} g_n$, we easily deduce that $\|f_n\|_G \to 0$.

■

4.2.2 Study of Meyer problem

We are now in position to carry out the mathematical study of Meyer problem [57].

Let $f \in L^q(\Omega)$ (with $q \geq 2$). We recall that the considered problem is:

$$\inf_{(u,v) \in BV(\Omega) \times G(\Omega) / f = u+v} \left(J(u) + \alpha \|v\|_{G(\Omega)} \right) \tag{4.14}$$

where $J(u)$ is the total variation $|Du|$ defined by (2.70).

Remark: Since f is an image, we know that $f \in L^\infty(\Omega)$. Thus it is not restrictive to suppose $q \geq 2$.

Before considering problem (4.14), we first need to show that we can always decompose a function $f \in L^q(\Omega)$ into two components $(u, v) \in BV(\Omega) \times G(\Omega)$.

Lemma 4.5. *Let $f \in L^q(\Omega)$ (with $q \geq 2$). Then there exists $u \in BV(\Omega)$ and $v \in G(\Omega)$ such that $f = u + v$.*

Proof: Let us choose $u = \frac{1}{|\Omega|} \int_\Omega f$ and $v = f - u = f - \frac{1}{|\Omega|} \int_\Omega f$. We therefore have $u \in BV(\Omega)$ (since Ω is bounded), and $v \in L^2(\Omega)$. Moreover, since $\int_\Omega v = 0$ we deduce from Proposition 4.7 that $v \in G(\Omega)$.

■

We now show that problem (4.14) admits at least one solution.

Proposition 4.2. *Let $f \in L^q(\Omega)$ (with $q \geq 2$). Then there exists $\hat{u} \in BV(\Omega)$ and $\hat{v} \in G(\Omega)$ such that $f = \hat{u} + \hat{v}$, and:*

$$J(\hat{u}) + \alpha \|\hat{v}\|_G = \inf_{(u,v) \in BV(\Omega) \times G(\Omega)/f=u+v} \left(J(u) + \alpha \|v\|_G \right) \tag{4.15}$$

Proof: Let us first remark that the functional to minimize in (4.14) is convex with respect to its two variables. Moreover, the infimum in (4.14) is finite (thanks to Lemma 4.5).

Now, let (u_n, v_n) be a minimizing sequence for (4.14) We thus have for some constant C

$$J(u_n) \leq C \text{ and } \|v_n\|_G \leq C \tag{4.16}$$

From Poincaré inequality (see [5]), there exists a constant $B > 0$ such that: $\|u_n - \int_\Omega u_n\|_{L^2(\Omega)} \leq BJ(u_n)$. Thus from (4.16), we get $\|u_n - \int_\Omega u_n\|_{L^2(\Omega)} \leq BC$. But as $u_n + v_n = f$, we have:

$$\int_\Omega u_n + \underbrace{\int_\Omega v_n}_{=0 \text{ since } v_n \in G(\Omega)} = \int_\Omega f \tag{4.17}$$

Hence u_n is bounded in $L^2(\Omega)$. From (4.16), we deduce that u_n is bounded in $BV(\Omega)$. Thus there exists $\hat{u} \in BV(\Omega)$ such that $u_n \rightharpoonup \hat{u}$ in $BV(\Omega)$ weak *. And as $u_n + v_n = f$, we deduce that v_n is also bounded in $L^2(\Omega)$. Therefore, there exists $\hat{v} \in L^2(\Omega)$ such that, up to an extraction, $v_n \rightharpoonup \hat{v}$ in $L^2(\Omega)$ weak.

To conclude, there remains to prove that $(\hat{u}, \hat{v})$ is a minimizer of $J(u) + \alpha \|v\|_{G(\Omega)}$. And this last point comes from the fact that J is lower semi-continuous (lsc) with respect to the BV weak * topology [5], and from the fact that $\|.\|_G$ is lsc with respect to the L^2-weak topology.

■

Remark: It has been shown that Meyer problem can admit several solutions [47].

4.3 Decomposition models and algorithms

The problem of image decomposition has been a very active field of research during the last past five years. [57], was the inspiration source of many works [69, 63, 16, 8, 66, 17, 19, 26, 32, 76].

This is a *hot* topic in image processing. We refer the reader to the UCLA CAM reports web page where he can find numerous papers dealing with this subject.

4.3.1 Which space to use?

We have the following result (stated in [57]):

$$\dot{B}^1_{1,1} \subset \dot{BV} \subset L^2 \subset G \subset E = \dot{B}^\infty_{-1,\infty} \tag{4.18}$$

where $\dot{BV}$ is the homogeneous version of BV: $\dot{BV} = BV / \{u \in BV \ / \ \nabla u = 0\}$.

In Figure 15, the three images have the same L^2 norm. Table 1 presents the values of different norms. It clearly illustrates the superiority of the G norm over the L^2 norm to capture oscillating patterns in minimization processes (the G norm is much smaller for the texture

Table 1: A striking example (see Figure 15)

Images	TV	L^2	$\|.\|_{-1,2}$	G	E (Daub10)
textured image	1 000 000	9 500	33 000	360	749
geometric image	64 600	9 500	300 000	2000	355
Gaussian noise ($\sigma = 85$)	2 100 000	9 500	9 100	120	287

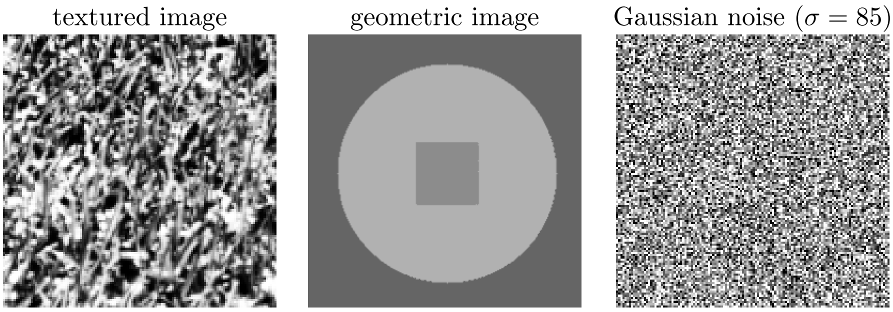

Figure 15: A striking example

image and the noise image than the geometric image), as claimed in [57]. It also illustrates why the use of the E norm is well adapted to separate the noise (the noisy image has the smallest E norm). These observations were the starting point of the decomposition algorithm by Aujol and Chambolle in [17] which split an image into three components: geometry, texture, and noise.

More generally, the choice of the functional space used in the modelling depend on the objective. The main advantage for using Besov spaces is their link with wavelet coefficients: this enables fast multi-scale algorithms. The main advantage of Sobolev spaces is their link with PDEs: an energy involving Sobolev norms can easily be minimized by solving it associated Euler-Lagrange equation. The main advantage of BV is that contrary to Sobolev or Besov spaces, it contains charcateristic functions: in particular, any piecewise regular function is in BV, which is the reason why BV is a good candidate to model the geometrical part of an image.

Wavelet based alternative to BV have been proposed in the literature. The most popular choice is $\dot{B}^1_{1,1}$. The main advantage of choosing $\dot{B}^1_{1,1}$ is that it often leads to wavelet shrinkage algorithm, and therefore very fast algorithms. And visually it gives very similar results to BV [67, 16].

Another wavelet based alternative to BV has recently been proposed in [47] with $\dot{B}^1_{1,\infty}$. This space is closer to BV then $\dot{B}^1_{1,1}$. But the drawback is that it does not lead to wavelet shrinkage algorithm, and therefore no fast algorithms have been proposed up to now.

4.3.2 Parameter tuning

When intereted in the general decomposition problem:

$$E_{Structure}(u) + \lambda E_{Texture}(v), \quad f = u + v, \tag{4.19}$$

We denote by (u_λ, v_λ) its solution (which is assumed to exist and to be unique). The problem is then to find the right regularization parameter λ. The goal is to find the right balance between the energy terms which produces a meaningful structure-texture decomposition.

For the denoising problem, one often assumes that the variance of the noise σ^2 is known *a-priori* or can be well estimated from the image. As the v part in the denoising case should contain mostly noise, a natural condition is to select λ such that the variance of v is equal to that of the noise, that is $\mathrm{var}(v) = \sigma^2$. Such a method was used in [65] in the constrained ROF model, and this principle dates back to Morozov [60] in regularization theory. Here we do not know of a good way to estimate the texture variance, also there is no performance criterion like the SNR, which can be optimized. Therefore we should resort to a different approach.

The approach follows the work of Mrazek-Navara [61], used for finding the stopping time for denoising with nonlinear diffusions. The method relies on a correlation criterion and assumes no knowledge of noise variance. As shown in [42], its performance is inferior to the SNR-based method of [42] and to an analogue of the variance condition for diffusions. For decomposition, however, the approach of [61], adopted for the variational framework, may be a good basic way for the selection of λ.

Let us define first the (empirical) notions of mean, variance and covariance in the discrete setting of $N \times N$ pixels image. The mean is

$$\bar{q} \doteq \frac{1}{N^2} \sum_{1 \le i,j \le N} q_{i,j},$$

the variance is

$$V(q) \doteq \frac{1}{N^2} \sum_{1 \le i,j \le N} (q_{i,j} - \bar{q})^2,$$

and the covariance is

$$\mathrm{covariance}(q, r) \doteq \frac{1}{N^2} \sum_{1 \le i,j \le N} (q_{i,j} - \bar{q})(r_{i,j} - \bar{r}).$$

We would like to have a measure that defines orthogonality between two signals and is not biased by the magnitude (or variance) of the signals. A standard measure in statistics is the correlation, which is the covariance normalized by the standard deviations of each signal:

$$\mathrm{correlation}(q, r) \doteq \frac{\mathrm{covariance}(q, r)}{\sqrt{V(q)V(r)}}.$$

By the Cauchy-Schwarz inequality it is not hard to see that $\mathrm{covariance}(q, r) \le \sqrt{V(q)V(r)}$ and therefore $|\mathrm{correlation}(q, r)| \le 1$. The upper bound 1 (completely *correlated*) is reached for signals which are the same, up to an additive constant and up to a positive multiplicative constant. The lower bound -1 (completely *anti-correlated*) is reached for similar signals but with a negative multiplicative constant relation. When the correlation is 0 we refer to the two signals as *not correlated*. This is a necessary condition (but not a sufficient one) for statistical independence. It often implies that the signals can be viewed as produced by different "generators" or models.

To guide the parameter selection of a decomposition we use the following assumption:

Assumption: *The texture and the structure components of an image are not correlated.*

This assumption can be relaxed by stating that the correlation of the components is very low. Let us define the pair (u_λ, v_λ) as the one minimizing (4.19) for a specific λ. As proved in [57] for the $TV - L^2$ model (and in [41] for any convex structure energy term with L^2), we have covariance$(u_\lambda, v_\lambda) \geq 0$ for any non-negative λ and therefore

$$0 \leq \text{correlation}(u_\lambda, v_\lambda) \leq 1, \ \forall \lambda \geq 0. \tag{4.20}$$

This means that one should not worry about negative correlation values. Note that positive correlation is guaranteed in the $TV - L^2$ case. In the $TV - L^1$ case we may have negative correlations, and should therefore be more careful.

Following the above assumption and the fact that the correlation is non-negative, to find the right parameter λ, we are led to consider the following problem:

$$\lambda^* = \text{argmin}_\lambda \left(\text{correlation}(u_\lambda, v_\lambda)\right). \tag{4.21}$$

In practice, one generates a scale-space using the parameter λ (in our formulation, smaller λ means more smoothing of u) and selects the parameter λ^* as the first local minimum of the correlation function between the structural part u and the oscillating part v.

This selection method can be very effective in simple cases with very clear distinction between texture and structure. In these cases correlation(u, v) behaves smoothly, reaches a minimum approximately at the point where the texture is completely smoothed out from u, and then increases, as more of the structure gets into the v part. The graphs of correlation(u, v) in the $TV - L^2$ case behave quite as expected, and the selected parameter lead to a good decomposition.

For more complicated images, there are textures and structures of different scales and the distinction between them is not obvious. In terms of correlation, there is no more a single minimum and the function may oscillate.

As a first approximation of a decomposition with a single scalar parameter, we suggest to choose λ after the first local minimum of the correlation is reached. In some cases, a sharp change in the correlation is also a good indicator: after the correlation sharply drops or before a sharp rise.

4.3.3 $TV - G$ algorithms

$$\inf_u \int_\Omega |Du| + \lambda \|f - u|_G \tag{4.22}$$

Vese-Osher model L. Vese and S. Osher were the first authors to numerically tackle Meyer program [69]. They actually solve the problem:

$$\inf_{(u,v) \in BV(\Omega) \times G(\Omega)} \left(\int |Du| + \lambda \|f - u - v\|_2^2 + \mu \|v\|_{G(\Omega)} \right) \tag{4.23}$$

where Ω is a bounded open set. To compute their solution, they replace the term $\|v\|_{G(\Omega)}$ by $\|\sqrt{g_1^2 + g_2^2}\|_p$ (where $v = \text{div}\,(g_1, g_2)$). Then they formally derive the Euler-Lagrange equations from (4.23). For numerical reasons, the authors use the value $p = 1$ (they claim they made experiments for $p = 1 \ldots 10$, and that they did not see any visual difference). They report good numerical results. See also [63] for another related model concerning the case $\lambda = +\infty$ and $p = 2$.

A^2BC model Inspired from the work by A. Chambolle [24], the authors of [16, 15] propose a relevant approach to solve Meyer problem. They consider the following functional defined on $L^2(\Omega) \times L^2(\Omega)$:

$$F_{\lambda,\mu}(u,v) = \begin{cases} \int_\Omega |Du| + \frac{1}{2\lambda}\|f-u-v\|^2_{L^2(\Omega)} & \text{if } (u,v) \in BV(\Omega) \times G_\mu(\Omega) \\ +\infty & \text{otherwise} \end{cases} \tag{4.24}$$

where $G_\mu(\Omega) = \left\{ v \in G(\Omega) \ / \ \|v\|_{G(\Omega)} \leq \mu \right\}$. And the problem to solve is:

$$\inf_{L^2(\Omega)\times L^2(\Omega)} F_{\lambda,\mu}(u,v) \tag{4.25}$$

The authors of [16] present their model in a discrete framework. They carry out a complete mathematical analysis of their discrete model, showing how it approximately solves Meyer problem.

Second order cone programming approach In [76], the authors use second order cone programming to compute the solution. In [49], a saddle point formulation is used. And in [72], general convex minimization algorithms are applied successfully to compute the solution.

4.3.4 $TV - L^1$

The use of the L^1 norm in image processing has first been proposed in [62] to remove outliers (salt and pepper noise case). The algorithm used in [62] was a relaxation algorithm (and therefore quite slow). The model in this case can be written:

$$\inf_u \int_\Omega |Du| + \lambda \|f-u\|_{L^1} \tag{4.26}$$

It was later studied from a mathematical point of view in [26], the numerical implementation being done with PDEs, but still quite slow (because of the singularity of the L^1 norm. An alternative approach was proposed in [12] with the functional:

$$\inf_{u,v} \int_\Omega |Du| + \mu\|f-u-v\|^2_{L^2} + \lambda\|v\|_{L^1} \tag{4.27}$$

By alternating minimization with respect to u and v, the solution is easily computed. Notice that minimization with respect to u amounts to classical total variation minimization, while minimization with respect to v is directly solved by thresholding v. Figure 16 shows an example of decomposition with this approach.

A fast algorithm was eventually proposed in [31]. Moreover, the authors of [31] show that (4.26) enjoys the nice property of being a contrast invariant filter.

A direct approach based on second order cone programming was proposed in [74]. Moreover, the same authors made a comparison in [75] between the classical $TV - L^2$ model (4.1), the $TV - G$ model (4.22), and $TV - L^1$ (4.26). Their conclusion is that $TV - L^1$ seems to bring better decomposition result (at least with synthetic images, where the user knows exactly what is the structure and what are the textures), although the differences are not that large with the $TV - G$ model. In any case, the classical $TV - L^2$ is worse, mainly due to its eroding effect (which implies that some of the structure always appears in the texture component).

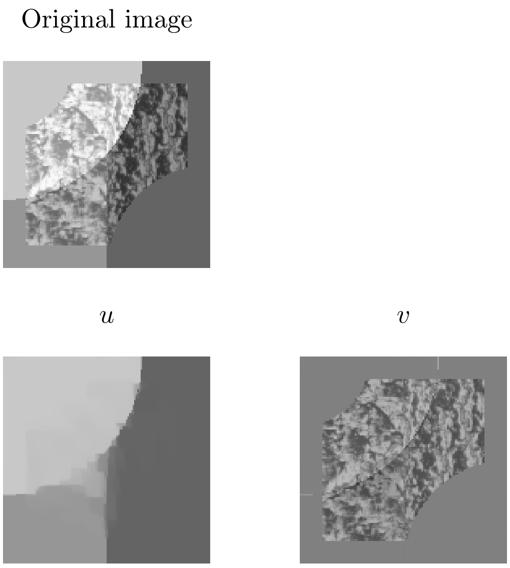

Figure 16: Approximation of the TV-L1 decomposition of non-geometric texture

.

Nevertheless, one should notice that the decomposition algorithm choice should be led by the developped application: indeed, depending whether is it a first step towards image inpainting, image compression, ..., the required properties of the algorithm can be slighty different. Moreover, all the proposed approaches assume that the user knows how to tune the parameter.

4.3.5 $TV - H^{-1}$

In [63], the authors have proposed to use H^{-1} to capture oscillating patterns. (we recall that H^{-1} is the dual space of $H_0^1 = W^{1,2}$).

The considered problem is the following:

$$\inf_u \int_\Omega |Du| + \lambda \|f - u\|_{H^{-1}}^2 \tag{4.28}$$

In [63], the solution was obtained by solving fourth order PDE. In [17], the authors proposed a modification of Chambolle's projection algorithm [24] to compute the solution (and they gave a proof of convergence). In [32], the authors replace the total variation by a $\|.\|_{B^1_{1,1}}$ regularization (with the Haar wavelet). They can then compute the solution of the problem in the frequency domain. See also [51] for other extensions.

The main advantage of using H^{-1} instead of other negative Sobolev spaces $W^{-1,p}$ (with $p \geq 1$) is that it is much easier to handle numerically. In particular, harmonic analysis tools can be applied. The main drawback of H^{-1} is that it does not lead to good decomposition results, as shown in [17] and explained in [12] (see Figure 17).

This is the reason why adaptive norms were introduced in [12].

$$\inf_u \int_\Omega |Du| + \lambda \|f - u\|_{\mathcal{H}}^2 \tag{4.29}$$

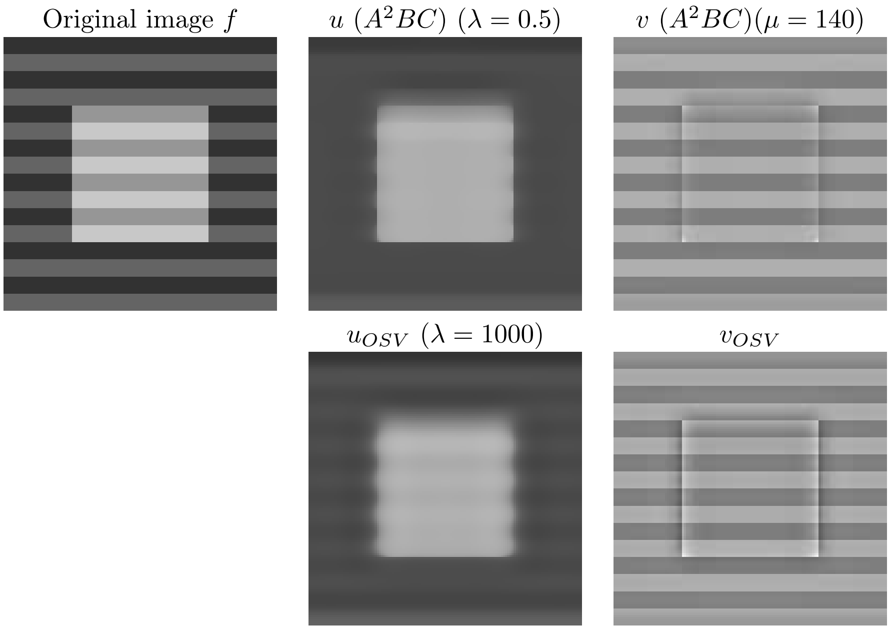

Figure 17: Decomposition (the parameters are tuned so that both v_{OSV} and the v component got with the A^2BC algorithm have the same L^2 norm)

with $\|u\|_{\mathcal{H}} = \int K\,|u|^2$, where K is a symmetric positive operator. This lead to adaptive image decomposition.

4.3.6 TV-Hilbert

The main drawback of all the proposed decomposition algorithms is their lack of adaptivity. It is obvious that in an image, the amount of texture is not uniform. A first method to incorporate spatial adaptivity has been introduced in [44], based on the local variance criterion proposed in [43]. Motivated by [65] and [63], the authors of [11] have proposed a generalization of the ROF and OSV models:

$$\inf_{(u\times v)\in BV\times\mathcal{H}/f=u+v} \left\{ \int |Du| + \lambda \|v\|_{\mathcal{H}}^2 \right\} \tag{4.30}$$

where $\mathcal{H}$ is some Hilbert space. In the case when $\mathcal{H} = L^2$, then (4.30) is the ROF model [65], and when $\mathcal{H} = H^{-1}$ then (4.30) is the OSV model [63]. By choosing suitably the Hilbert space $\mathcal{H}$, it is possible to compute a frequency and directional adaptive image decomposition, as shown on Figure 18.

More precisely, the functional to minimize is:

$$\inf_u \int_\Omega |Du| + \lambda \|\sqrt(K)(f-u)\|_{L^2}^2 \tag{4.31}$$

where K is a positive symmetric operator.

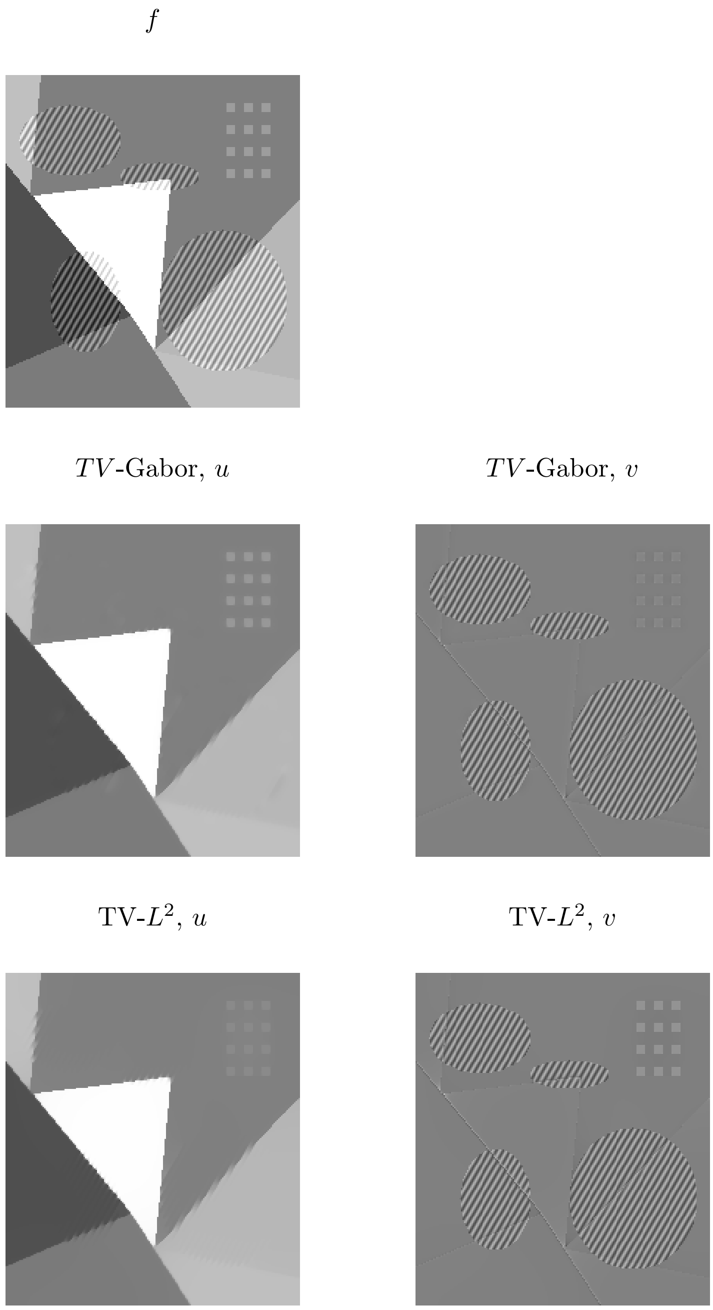

Figure 18: Decomposition of a synthetic image with textures of specific frequency and orientation by TV-Gabor and $TV - L^2$. The TV-Gabor can be more selective and reduce the inclusion in v of undesired textures / small-structures like the small blocks on the top right. Also erosion of large structures is reduced (more apparent in the brighter triangle).

4.3.7 Using negative Besov space

In [57], the author suggested to use negative Besov spaces to capture texture. This was the motivation of the work [39]. The considered functional thus becomes:

$$\inf_u \int_\Omega |Du| + \lambda \|f - u\|_{B^s_{p,q}} \tag{4.32}$$

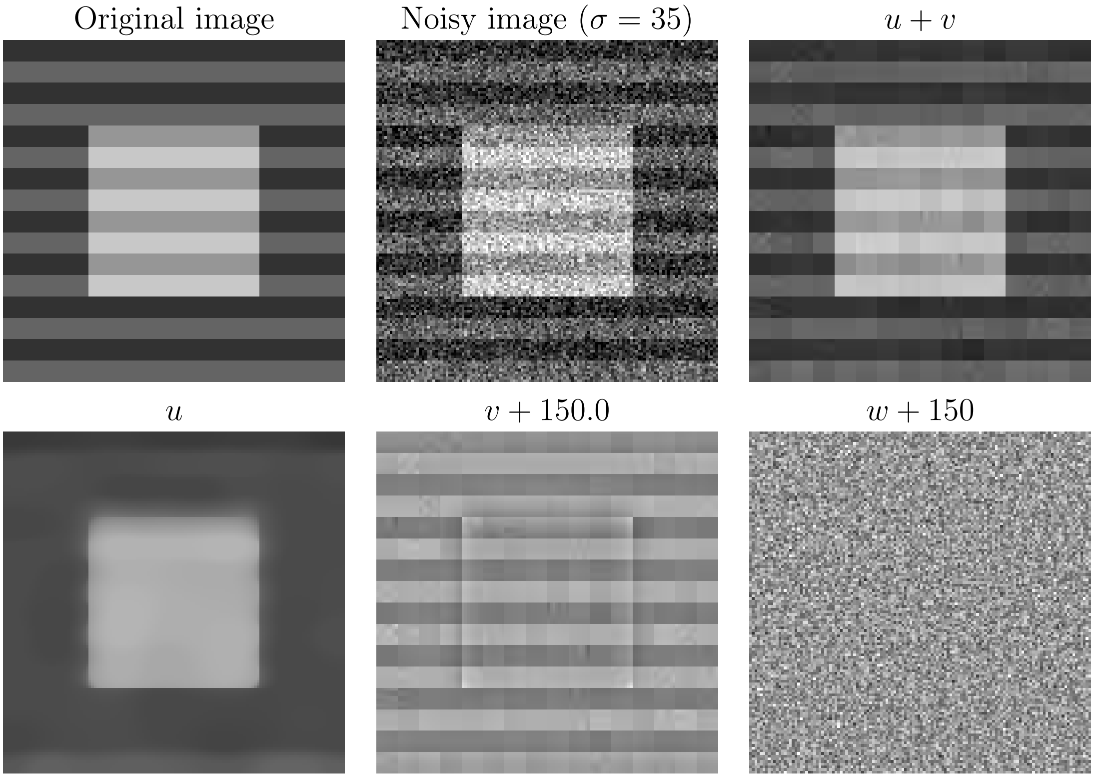

Figure 19: A simple example ($\lambda = 0.5$, $\mu = 120$, $\eta = 1.0$, Haar)

In [39], the authors use a definition of Besov spaces $B^s_{p,q}$ based on Poisson and Gaussian kernels (see [39] for further details): this enables them to compute a solution with a PDE based approach. Similar numerical results are presented in [50], where $B^s_{p,q}$ is replaced by div(BMO) (see [57, 39]).

4.3.8 Using Meyer's E space

In [17], the authors propose to use $E = \dot{B}^{\infty}_{-1,\infty}$ the dual space of $\dot{B}^1_{1,1}$. They show that such a space is particularly well suited to capture the white Gaussian noise. They introduce a model with three components to capture the geometry u, the texture v, and the noise w. Their functional is the following:

$$\inf_{u \in BV, \|v\|_G \leq \mu, \|w\| \leq \nu} \int_\Omega |Du| + \lambda \|f - u - v - w\|_2^2 \tag{4.33}$$

Minimizing this functional is done by alternating wavelet thresholding and Chambolle's projection algorithm.

A modification of this algorithm is proposed in [44]. The main novelty is the use of an adaptive weighting to locally control the balance between texture and noise. This local parameter is computed using the method proposed in [43] (depending on the local variance of the image).

A numerical example is shown on Figures 19 and 20.

In [57], the author propose a last class of functional space to model texture: BMO. We will not discuss these spaces, since from the numerical point of view it give similar results to the other functional spaces [50].

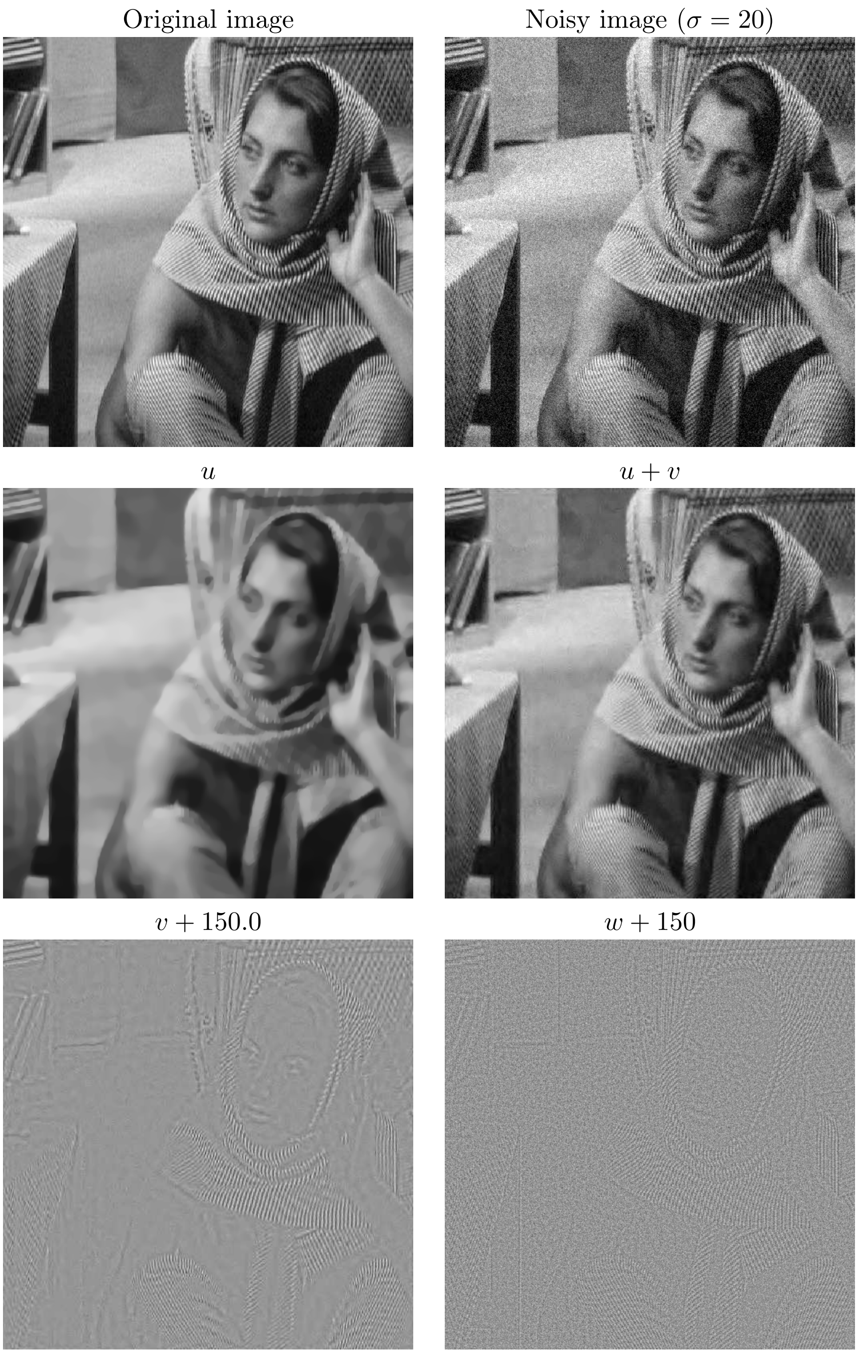

Figure 20: Barbara image ($\lambda = 1.0$, $\mu = 30$, $\eta = 0.6$, Daub8)

(a) f (b) u (c) v

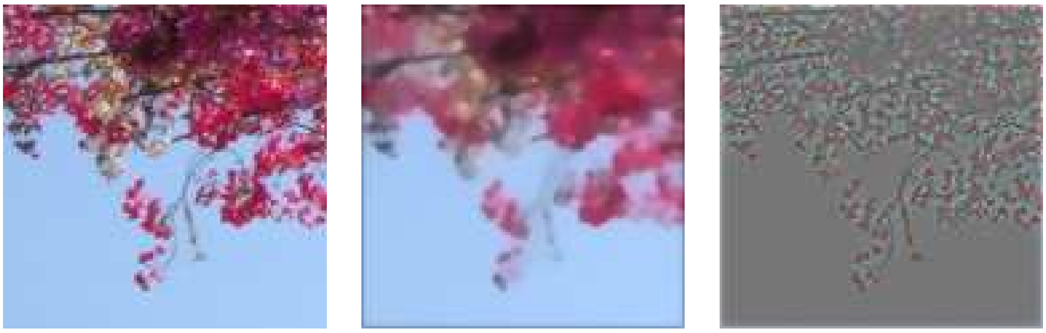

Figure 21: (a) original image f, (b) BV component u, (c) texture v component ($v+0.5$ plotted). Some of the thicker branches are in the BV part u, while the thin and narrow branches in the bottom middle are in the v component. u as well as v are both color images.

(a) f (b) u (c) v

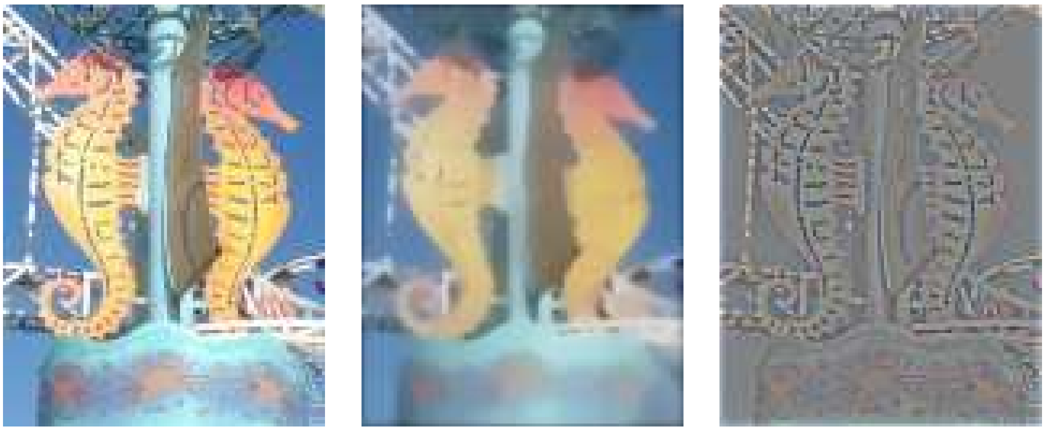

Figure 22: (a) original image f, (b) BV component u, (c) texture v component ($v+0.5$ plotted). All the details of image are in v, while the BV component is well kept in u.

4.3.9 Applications of image decomposition

The problem of image decomposition is a very interesting problem by itself. It raises both simulating numerical and mathematical issues. Moreover, it has been applied with success to some image processing problems. In [20], the authors use image decomposition to carry out image inpainting. Indeed, inpainting techniques are different depending on the type of the image. In the case of texture images, then copy paste methods are used, whereas in the case of geometric images diffusion methods give good results. In [18], image decomposition is used to improve nonlinear image interpolation results. In [10], image decomposition is applied successfully to improve image classification results. Notice also [13] where color images are considered (see Figures 21 and 22).

A . Discretization

A numerical image can be seen as vector with 2 dimensions Une image numérique, ou discrète, est un vecteur à deux dimensions de $N \times N$. We denote by X the Euclidean space $\mathbb{R}^{N \times N}$, and

$Y = X \times X$. We embed X with the inner product:

$$(u, v)_X = \sum_{1 \le i,j \le N} u_{i,j} v_{i,j} \tag{1.1}$$

and the norm:

$$\|u\|_X = \sqrt{(u, u)_X} \tag{1.2}$$

To define a discrete version of the total variation, we first introduce a discrete version of the gradient operator. If $u \in X$, the gradient ∇u is a vector in Y given by:

$$(\nabla u)_{i,j} = ((\nabla u)^1_{i,j}, (\nabla u)^2_{i,j}) \tag{1.3}$$

with

$$(\nabla u)^1_{i,j} = \begin{cases} u_{i+1,j} - u_{i,j} & \text{if } i < N \\ 0 & \text{if } i = N \end{cases} \tag{1.4}$$

and

$$(\nabla u)^2_{i,j} = \begin{cases} u_{i,j+1} - u_{i,j} & \text{if } j < N \\ 0 & \text{if } j = N \end{cases} \tag{1.5}$$

The discrete total variation of u is then given by:

$$J(u) = \sum_{1 \le i,j \le N} |(\nabla u)_{i,j}| \tag{1.6}$$

We also introduce a discrete version of the divergence operator. We define it by analogy with the continuous case:

$$\text{div } = -\nabla^* \tag{1.7}$$

where ∇^* is the adjoint operator of ∇ : i.e., for all $p \in Y$ and $u \in X$, $(-\text{div}\, p, u)_X = (p, \nabla u)_Y$. It is then easy to check:

$$(\text{div}\,(p))_{i,j} = \begin{cases} p^1_{i,j} - p^1_{i-1,j} & \text{if } 1 < i < N \\ p^1_{i,j} & \text{if i=1} \\ -p^1_{i-1,j} & \text{if i=N} \end{cases} + \begin{cases} p^2_{i,j} - p^2_{i,j-1} & \text{if } 1 < j < N \\ p^2_{i,j} & \text{if j=1} \\ -p^2_{i,j-1} & \text{if j=N} \end{cases} \tag{1.8}$$

We will use a discrete version of the Laplacien operator defined by:

$$\Delta u = \text{div}\, \nabla u \tag{1.9}$$

References

[1] R. Acar and C. Vogel. Analysis of total variation penalty methods for ill-posed problems. *Inverse Problems*, 10:1217–1229, 1994.

[2] R. Adams. *Sobolev Spaces*. Pure and applied Mathematics. Academic Press, Inc, 1975.

[3] A. Almansa, J-F. Aujol, V. Caselles, and G. Facciolo. Restoration of irregular sampled images, 2007.

[4] L. Alvarez, Y. Gousseau, and J.M. Morel. Scales in natural images and a consequence on their bounded variation norm. In *Scale-Space '99*, volume 1682 of *Lecture Notes in Computer Science*, 1999.

[5] L. Ambrosio, N. Fusco, and D. Pallara. *Functions of bounded variations and free discontinuity problems*. Oxford mathematical monographs. Oxford University Press, 2000.

[6] F. Andreu-Vaillo, V. Caselles, and J. M. Mazon. *Parabolic quasilinear equations minimizing linear growth functionals*, volume 223 of *Progress in Mathematics*. Birkhauser, 2002.

[7] H. Attouch, G. Buttazzo, and G. Michaille. *Variational Analysis in Sobolev and BV spaces. Applications to PDEs and Optimization*. MPS-SIAM Series on Optimization. SIAM, 2005.

[8] G. Aubert and J.F. Aujol. Modeling very oscillating signals. Application to image processing. *Applied Mathematics and Optimization*, 51(2), March/April 2005.

[9] G. Aubert and P. Kornprobst. *Mathematical Problems in Image Processing*, volume 147 of *Applied Mathematical Sciences*. Springer-Verlag, 2002.

[10] J-F. Aujol and T. F. Chan. Combining geometrical and textured information to perform image classification. *Journal of Visual Communication and Image representation*, 17(5):1004–1023, October 2006.

[11] J-F. Aujol and G. Gilboa. Constrained and SNR-based solutions for TV-hilbert space image denoising. *Journal of Mathematical Imaging and Vision*, 26(1-2):217–237, 2006.

[12] J-F. Aujol, G. Gilboa, T. Chan, and S. Osher. Structure-texture image decomposition – modeling, algorithms, and parameter selection. *International Journal on Computer Vision*, 67(1):111–136, 2006.

[13] J-F. Aujol and S.H. Kang. Color image decomposition and restoration. *Journal of Visual Communication and Image Representation*, 17(4):916–926, August 2006.

[14] J.F. Aujol. Some algorithms for total variation based image restoration. *CMLA Preprint 2008-05*, 2008. http://hal.archives-ouvertes.fr/hal-00260494/en/.

[15] J.F. Aujol, G. Aubert, L. Blanc-Féraud, and A. Chambolle. Decomposing an image: Application to SAR images. In *Scale-Space '03*, volume 1682 of *Lecture Notes in Computer Science*, 2003.

[16] J.F. Aujol, G. Aubert, L. Blanc-Féraud, and A. Chambolle. Image decomposition into a bounded variation component and an oscillating component. *JMIV*, 22(1):71–88, January 2005.

[17] J.F. Aujol and A. Chambolle. Dual norms and image decomposition models. *IJCV*, 63(1):85–104, June 2005.

[18] J.F. Aujol and B. Matei. Structure and texture compression, 2004. INRIA Research Report 5076, submitted, http://www.inria.fr/rrrt/rr-5076.html.

[19] J. Bect, G. Aubert, L. Blanc-Féraud, and A. Chambolle. A l^1 unified variational framwork for image restoration. In *ECCV 2004*, May 2004.

[20] M. Bertalmio, L. Vese, G. Sapiro, and S. Osher. Simultaneous structure and texture image inpainting. *IEEE Transactions on Image Processing*, 12(8):882–889, 2003.

[21] J. Bourgain and H. Brezis. On the equation div y=f and application to control of phases, October 2002. R02022, Laboratoire J.L. Lions, Université Paris 6, http://www.ann.jussieu.fr/publications/2002.php3.

[22] J. Bourgain and H. Brezis. Sur l'équation *div u=f*. *C.R.Math.Acad.Sci.Paris*, 334(11):973–976, 2002.

[23] H. Brezis. *Analyse fonctionnelle. Théorie et applications.* Mathématiques appliquées pour la maitrise. Masson, 1983.

[24] A. Chambolle. An algorithm for total variation minimization and applications. *JMIV*, 20:89–97, 2004.

[25] A. Chambolle and P.L. Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76(3):167–188, 1997.

[26] T. Chan and S. Esedoglu. Aspects of total variation regularized L^1 function approximation. *SIAM Journal on Applied Mathematics*, 65(5):1817–1837, 2005.

[27] T. Chan and P. Mulet. On the convergence of the lagged diffusity fixed point method in total variation image restoration. *SIAM Journal on Numerical Analysis*, 36(2):354–367, 1999.

[28] P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE Transactions on Image Processing*, 6(2):298–311, 1997.

[29] P.G. Ciarlet. *Introduction à l'analyse numérique matricielle et l̀'optimisation.* Mathématiques appliquées pour la maitrise. Masson, 1984.

[30] A. Cohen. *Numerical analysis of wavelet methods.* Elsevier, 2003.

[31] J. Darbon and M. Sigelle. Image restoration with discrete constrained total variation part i: Fast and exact optimization. *JMIV*, 26(3):277–291, 2006.

[32] I. Daubechies and G. Teschke. Variational image restoration by means of wavelets: simultaneous decomposition, deblurring and denoising. *Applied and Computational Harmonic Analysis*, 19(1):1–16, 2005.

[33] R. Dautray and J.L. Lions. *Analyse mathématique et calcul numérique pour les sciences et les techniques*, volume 4. Masson, 1988.

[34] D. Dobson and C. Vogel. Convergence of an iterative method for total variation denoising. *SIAM Journal on Numerical Analysis*, 34:1779–1791, 1997.

[35] D.L. Donoho and M. Johnstone. Adapting to unknown smoothness via wavelet shrinkage. *Journal of the American Statistical Association*, 90(432):1200–1224, December 1995.

[36] I. Ekeland and R. Temam. *Analyse convexe et problèmes variationnels*, volume 224 of *Grundlehren der mathematischen Wissenschaften*. Dunod, second edition, 1983.

[37] L.C. Evans. *Partial Differential Equations*, volume 19 of *Graduate Studies in Mathematics*. American Mathematical Society, 1991.

[38] L.C. Evans and R.F. Gariepy. *Measure theory and fine properties of functions*, volume 19 of *Studies in Advanced Mathematics*. CRC Press, 1992.

[39] J. Garnett, T. Le, Y. Meyer, and L. Vese. Image decompositions using bounded variation and homogeneous besov spaces, 2005. UCLA CAM Report 05-57.

[40] D. Gilbarg and N.S. Trudinger. *Elliptic Partial Differential Equations of Second Order*, volume 28 of *Princeton Mathematical Series*. Springer-Verlag, 1970.

[41] G. Gilboa, N. Sochen, and Y.Y. Zeevi. Estimation of optimal PDE-based denoising in the SNR sense. To appear in IEEE Trans. Image Processing, see http://www.math.ucla.edu/∼gilboa/.

[42] G. Gilboa, N. Sochen, and Y.Y. Zeevi. Estimation of optimal PDE-based denoising in the SNR sense, 2004. CCIT report No. 499, Technion, August, see http://www.math.ucla.edu/∼gilboa/.

[43] G. Gilboa, N. Sochen, and Y.Y. Zeevi. Variational denoising of partly-textured images by spatially varying constraints. *IEEE Transactions on Image Processing*, 15(8):2281–2289, 2006.

[44] J. Gilles. Noisy image decomposition: a new structure, textures and noise model based on local adaptivity, 2006. submitted.

[45] E. Giusti. *Minimal surfaces and functions of bounded variation*. Birkhauser, 1994.

[46] P. Grisvard. *Elliptic problems in nonsmooth domains*, volume 24 of *Monographs and studies in Mathematics*. Pitman, 1985.

[47] A. Haddad and Y. Meyer. Variational methods in image processing, 2004. UCLA CAM Report 04-52.

[48] J.B. Hiriart-Urruty and C. Lemarechal. *Convex Analysis ans Minimisation Algorithms I*, volume 305 of *Grundlehren der mathematischen Wissenschaften*. Springer-Verlag, 1993.

[49] S. Kindermann and S. Osher. Saddle point formulation for a cartoon-texture decomposition, 2005. UCLA CAM Report 05-42.

[50] T. Le and L. Vese. Image decomposition using total variation and div(bmo). *SIAM Journal on multiscale modeling and simulation*, 4(2):390–423, 2005.

[51] L. Lieu and L. Vese. Image restoration and decompostion via bounded total variation and negative hilbert-sobolev spaces, 2005. UCLA CAM Report 05-33.

[52] J.L. Lions and E. Magenes. *Problèmes aux limites non homogènes*, volume 1. Dunod, 1968.

[53] F. Malgouyres. Minimizing the total variation under a general convex constraint for image restoration. *IEEE transactions on image processing*, 11(12):1450–1456, December 2002.

[54] F. Malgouyres and F. Guichard. Edge direction preserving image zooming: a mathematical and numerical analysis. *SIAM Journal on Numerical Analysis*, 39(1):1–37, 2001.

[55] S.G. Mallat. *A Wavelet Tour of Signal Processing.* Academic Press, 1998.

[56] S. Masnou. Disocclusion: a variational approach using level lines. *IEEE Transactions on Image Processing*, 11(2):68–76, 2002.

[57] Yves Meyer. Oscillating patterns in image processing and in some nonlinear evolution equations, March 2001. The Fifteenth Dean Jacquelines B. Lewis Memorial Lectures.

[58] A.A. Miljutin. A priori estimates for solutions of second order linear elliptic equations. *Mat. Sbornik*, pages 459–474, 1960. Russian.

[59] C. Miranda. *Partial differential equations of elliptic type.* Springer-Verlag, second edition, 1970.

[60] V. A. Morosov. On the solution of functional equations by the method of regularization. *Soviet Math. Dokl.*, 7:414–417, 1966.

[61] P. Mrázek and M. Navara. Selection of optimal stopping time for nonlinear diffusion filtering. *IJCV*, 52(2/3):189–203, 2003.

[62] M. Nikolova. A variational approach to remove outliers and impulse noise. *JMIV*, 20(1-2):99–120, 2004.

[63] S.J. Osher, A. Sole, and L.A. Vese. Image decomposition and restoration using total variation minimization and the H^{-1} norm. *Multiscale Modeling and Simulation: A SIAM Interdisciplinary Journal*, 1(3):349–370, 2003.

[64] T. Rockafellar. *Convex Analysis*, volume 224 of *Grundlehren der mathematischen Wissenschaften.* Princeton University Press, second edition, 1983.

[65] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[66] J.L. Starck, M. ELad, and D.L. Donoho. Image decomposition: separation of texture from piecewise smooth content, 2003. To appear in IEEE Transactions on Image Processing.

[67] G. Steidl, J. Weickert, T. Brox, P. Mrazek, and M. Welk. On the equivalence of soft wavelet shrinkage, total variation diffusion, total variation regularization, and sides. *SIAM J. Numer. Anal.*, 42:686–658, 2004.

[68] R. Temam. *Navier Stokes equations.* Elsevier Science Publishers B.V., 1984.

[69] L.A. Vese and S.J. Osher. Modeling textures with total variation minimization and oscillating patterns in image processing. *Journal of Scientific Computing*, 19:553–572, 2003.

[70] C. Vogel and M. Oman. Iterative methods for total variation denoising. *SIAM Journal on Scientifical Computing*, 17:227–238, 1996.

[71] C.R. Vogel. *Computational Methods for Inverse Problems*, volume 23 of *Frontiers in Applied Mathematics.* SIAM, 2002.

[72] P. Weiss and G. Aubert eand L. Blanc-Feraud. Some applications of l infinite norms in image processing, 2006. INRIA Research Report 6115.

[73] E. Weiszfeld. Sur le point pour lequel la somme des distances de points donnés est minimum. *Tôhoko Mathematics Journal*, 43:355–386, 1937.

[74] W. Yin, D. Goldfarb, and S. Osher. Image cartoon-texture decomposition and feature selection using the total variation regularized l^1, 2005. UCLA CAM Report 05-47.

[75] W. Yin, D. Goldfarb, and S. Osher. A comparison of three total variation based texture extraction models, 2006. UCLA CAM Report 06-41.

[76] W. Yinn and D. Goldfarb. Second-order cone programming methods for total variation based image restoration. *SIAM J. Scientific Computing*, 27(2):622–645, 2005.