Variational and PDE approaches to imaging problems

Background reading

These lecture notes are based on the lecture notes of Carola Schönlieb from her Mathematical Tripos Part III course of "Image Processing – Variational and PDE Methods" given in Lent Term 2014. You may also want to consult:

1. Chambolle, A., Caselles, V., Cremers, D., Novaga, M., & Pock, T. (2010). *An introduction to total variation for image analysis.* Theoretical foundations and numerical methods for sparse recovery, 9(263-340), 227.
2. Chapter 3 of Aubert, G., & Kornprobst, P. (2006). *Mathematical problems in image processing: partial differential equations and the calculus of variations* (Vol. 147). Springer Science & Business Media.

Lecture 16

Recap We have discussed the representation of signals as elements of a Hilbert space, and their approximation and discretization with respect to different bases. Their approximation depends on the geometric features of the underlying signal, for example, we saw that piecewise Lipschitz-α functions have sparse wavelet representations. Furthermore, one can apply ℓ^1-regularization to exploit such sparse representations when dealing with ill-posed inverse problems.

In this final part of this course, we will look more closely at inverse problems in image analysis. Rather than considering sparsity, we will work directly with the geometric features of images as elements of some function space.

1 The variational approach

1.1 The Bayesian approach to inverse problems

We have seen how one specific variational approach, namely ℓ^1 regularization, arises from the need to recover sparse vectors. In this section, we describe the Bayesian viewpoint of variational methods.

Given $g \in \mathbb{R}^{N\times N}$, there are two components to (linear) inverse problems:

1. A data model: $g = Tu_0 + n$ where $u_0 \in \mathbb{R}^{N\times N}$ is the original image to be recovered, T is some linear transform (e.g. a blurring operator, a subsampled Fourier transform, or the identity matrix), and n is the noise. Typically, the entries in n are assumed to be Gaussian distributed with mean 0 and variance σ^2.
2. An a-priori probability density: $P(u) = e^{-p(u)}$. This represents the idea that we have of the original image.

By Bayes' rule, the posteriori probability of u knowing g is

$$P(u|g)P(g) = P(g|u)P(u),$$

where $P(g|u) = \exp\left(-\frac{1}{\sigma^2}\|g - Tu\|_2^2\right)$. So,

$$P(u|g) = \frac{\exp\left(-\frac{1}{\sigma^2}\|g - Tu\|_2^2 - p(u)\right)}{P(g)},$$

and we want to find the maximum a posteriori (MAP) reconstruction:

$$u^* \in \operatorname*{argmax}_u P(u|g).$$

Equivalently,

$$u^* \in \operatorname*{argmin}_u p(u) + \frac{1}{\sigma^2}\|g - Tu\|_2^2.$$

1.2 Regularizers

We will study the minimizers to the energies derived in the Bayesian model in the continuous setting. Given $g \in L^2(\Omega)$ where $\Omega \subset \mathbb{R}^2$ is a bounded domain (e.g. $\Omega = [0,1]^2$) and a bounded linear operator $T : L^2(\Omega) \to L^2(\Omega)$, we will consider

$$\min_{u \in L^2(\Omega)} \lambda F(u) + \frac{1}{2}\|Tu - g\|_2^2.$$

1.3 What is a good choice for F?

Tychonov regularization Standard Tychonov regularization typically consider $F(u) = \frac{1}{2}\int_\Omega u^2$ or $F(u) = \frac{1}{2}\int_\Omega |\nabla u|^2$. The intuition behind the latter regularizer is that it encourages solutions with small gradient which best fit the observation data g, so noise is removed. However, one can check that u is a minimizer if and only if

$$T^*Tu - T^*g - \lambda\Delta u = 0,$$

with Neumann boundary condition $\nabla u \cdot \eta = 0$ on $\partial\Omega$ where η is the outward unit normal to $\partial\Omega$. This leads to oversmooth reconstructions as Δ has very strong isotropic smoothing properties. Moreover, as a consequence of classical Sobolev embedding theorems, such functions cannot exhibit discontinuities across hypersurfaces. In 2D, this corresponds to no discontinuities across lines. To offer a quick (formal) justification: if $0 < s < t < 1$, $u : [0,1] \to \mathbb{R} \in W^{1,2}(0,1)$, then

$$u(t) - u(s) = \int_s^t u'(r)\mathrm{d}r \leqslant \sqrt{t-s}\sqrt{\int_s^t |u'(r)|^2\,\mathrm{d}r} \leqslant \sqrt{t-s}\|u\|_{W^{1,2}}.$$

So, u is Hölder-1/2 continuous. This is especially problematic because it is the key information about an image is encoded in its edges!

Edge preserving regularizers There were 2 types of solutions (with $T = \mathrm{Id}$) in the quest for edge preserving regularizers in the late 80's and early 90's, both of which have been hugely influential.

1. Considering the problem in finite dimensions, D. Geman and S. Geman introduced an additional variable $l = (l_{i+1/2,j}, l_{i,j+1/2})_{i,j}$. Each entry in l takes value either 0 or 1. For each i, j, $l_{i+1/2,j} = 1$

indicates an edge between pixels (i,j) and $(i+1,j)$ and $l_{i+1/2,j}=0$ otherwise. In the Bayesian context, the function p in the a-priori probability density of the true image is replaced with

$$p(u,l) = \mu \sum_{i,j} (l_{i+1/2,j} + l_{i,j+1/2})$$
$$\lambda \sum_{i,j} \left((1-l_{i+1/2,j})(u_{i+1,j}-u_{i,j})^2 + (1-l_{i,j+1/2})(u_{i,j+1}-u_{i,j})^2 \right),$$

for $\lambda,\mu>0$. In the continuous setting, $\{l=1\}$ corresponds to a 1D curve $K\subset\Omega$ and this model corresponds to the Mumford and Shah model:

$$\min_{u,K} \int_{\Omega\setminus K} |\nabla u|^2 \, \mathrm{d}x + \mu \mathrm{length}(K) + \|u-g\|_{L^2}^2$$

among 1D closed subsets K of Ω and $u\in W^{2,2}(\Omega\setminus K)$. This model has generated a lot of interesting mathematical tools over the past decades. However, one of the issues with this model is that it is mathematically hard to analyse (note that we are minimizing over two very different objects u and K). Morevoer, since it is nonconvex, except in special cases, there is no way of knowing if one is converging to a minimum.

2. Rudin, Osher and Fatemi introduced the total variation functional for image processing:

$$F(u) = \int_\Omega |\nabla u|.$$

This functional is well defined on $W^{1,1}(\Omega)$. Note that given $u\in W^{1,1}([a,b])$, one can define a continuous function $\tilde{u}(x)-\tilde{u}(a)=\int_a^x u'(t)\mathrm{d}t$ which coincides with u a.e.. So, functions in $W^{1,1}([a,b])$ cannot have discontinuities, and given $f\in W^{1,1}([a,b]^2)$, since $f(\cdot,x)\in W^{1,1}([a,b])$ for a.e. x, images cannot have jumps across vertical/horizontal boundaries. However, the key point here is that F is well defined for a more general class of functions which can have discontinuities. Furthermore, the resultant variational problem is now convex, which allows for the application of some standard numerical solvers.

Lecture 17

1.4 An example

We shall see in this example that not only can $\int|\nabla u|$ be extended to a larger class of functions where edges are permitted, it is actually necessary to do so.

Consider

$$\min_{u\in W^{1,1}([0,1])} \mathcal{E}(u), \qquad \mathcal{E}(u) = \lambda\int_0^1 |u'(t)|\,\mathrm{d}t + \int_0^1 |u(t)-g(t)|^2\,\mathrm{d}t,$$

where $g=\chi_{(1/2,1]}$. We will show that this minimization problem does not have a solution in $W^{1,1}$.

- **Maximum/minimum principles** If u is a minimizer, then $u\leqslant 1$ a.e.: Let $v\in\min\{u,1\}$. Then,
  - $v'=u'$ on $\{u<1\}$ and $v'=0$ on $\{u\geqslant 1\}$. Therefore, $\int|v'|\leqslant\int|u'|$.
  - Since $g\leqslant 1$, $\|v-g\|^2\leqslant\|u-g\|^2$.

  So, $\mathcal{E}(v)\leqslant\mathcal{E}(u)$ and this inequality is strict if $v\neq u$. Similarly, $u\geqslant 0$ a.e..
- **'Symmetry'** Note that $g(t)=1-g(1-t)$. Any minimizer must also have this property: Let $\tilde{u}=1-u(1-t)$. Then $\|\tilde{u}-g\|^2=\int|1-u(1-t)-g(t)|^2\,\mathrm{d}t=\|u-g\|^2$. Also, $\|\tilde{u}'\|_1=\|u'\|_1$. So, $\mathcal{E}(\tilde{u})=\mathcal{E}(u)$.

Also,

$$\mathcal{E}\left(\frac{1-u(1-\cdot)+u}{2}\right) \leqslant \frac{1}{2}\mathcal{E}(1-u(1-\cdot)) + \frac{1}{2}\mathcal{E}(u) = \mathcal{E}(u)$$

and by strict convexity of $\|\cdot\|_2^2$, this inequality is strict if $v = \frac{1-u(1-\cdot)+u}{2} \neq u$. Therefore, $v = u$ which implies that $u(t) = 1 - u(1-t)$.

- Let $m = \min u = u(a)$ and let $M = \max u = u(b)$. From the previous observation, $M = 1 - m$. Then, (assuming $b > a$, but the case $a \geqslant b$ can be dealt with similarly)

$$\|u'\|_1 \geqslant \int_a^b |u'(t)|\,\mathrm{d}t \geqslant \int_a^b u'(t) = M - m = 1 - 2m.$$

Also, since $m \leqslant 1 - m$, we must have $m \in [0, 1/2]$.

To summarize, we have shown that $u \in [m, 1-m]$ for some $m \in [0, 1/2]$, $u(1-t) = 1 - u(t)$, and

$$\mathcal{E}(u) \geqslant \lambda(M-m) + \int_0^{1/2} m^2 + \int_{1/2}^1 (1-M)^2 = \lambda(1-2m) + m^2.$$

The RHS is minimal when $m = \lambda$ if $\lambda \leqslant 1/2$ and $m = 1/2$ if $\lambda \geqslant 1/2$. In the latter case, we see that $u \equiv 1/2$ achieves the minimum and is the unique minimizer.

Assume now that $\lambda < 1/2$. Then for any minimizer u, $\mathcal{E}(u) \geqslant \lambda(1-\lambda)$. Let us construct a minimizing sequence: For $n \geqslant 2$, define

$$u_n(t) = \begin{cases} \lambda & t \leqslant 1/2 - 1/n, \\ \frac{1}{2} + n(t-1/2)(1/2-\lambda) & |t - 1/2| \leqslant 1/n, \\ 1 - \lambda & t \geqslant 1/2 + 1/n. \end{cases}$$

Then, $\int_0^1 |u_n'| = \int_0^1 u_n' = 1 - 2\lambda$. Also,

$$\mathcal{E}(u_n) \leqslant \lambda(1-2\lambda) + \left(1 - \frac{2}{n}\right)^2 \lambda^2 + \frac{2}{n} \to \lambda(1-\lambda), \qquad n \to \infty.$$

Hence, $\inf_u \mathcal{E}(u) = \lambda(1-\lambda)$.

If u is a minimizer, then

$$\lambda(1-\lambda) = \lambda \underbrace{\int |u'|}_{\geqslant 1-2\lambda} + \underbrace{\int |u-g|^2}_{\geqslant \lambda^2/2+\lambda^2/2=\lambda^2},$$

and we have that $\int |u'| = 1 - 2\lambda$ and $\int |u-g|^2 = \lambda^2$. From this, we see that u is nondecreasing from λ to $1-\lambda$. But this implies that $|u-g| \geqslant \lambda$ a.e. and from $\|u-g\|^2 = \lambda^2$, we have that $|u-g| = \lambda$ a.e.. Therefore, $u = \lambda\chi_{[0,1/2)} + (1-\lambda)\chi_{[1/2,1]}$. This is the L^1 limit of u_n but is not in $W^{1,1}$. Note also that since $\int |u_n'| = 1 - 2\lambda$ for all n, it is natural to assume that $\int |u'|$ makes sense. A natural extension of the functional F is to define for $u \in L^1$:

$$F(u) = \inf\left\{ \lim_{n\to\infty} \int_0^1 |u_n'(t)|\,\mathrm{d}t \;;\; u_n \to u \text{ in } L^1, \quad \lim_{n\to\infty} \int_0^1 |u_n'| < \infty \right\}.$$

As we shall see, this definition is consistent with the definition of the more standard definition of total variation.

2 Mathematical preliminaries

We recall some definitions and results from functional analysis:
Definition 1 (Strong, weak, weak-* convergence)**.** *A sequence* (x_n) *in a normed space* $(X, \|\cdot\|_X)$ *converges*

- *strongly to* $x \in X$ *if* $\lim_{n\to\infty} \|x_n - x\|_X = 0$.
- *weakly to* $x \in X$ *if* $\lim_{n\to\infty} \langle x_n, y\rangle_{X\times Y} = \langle x, y\rangle_{X\times Y}$ *for all* $y \in Y = X^*$*, the dual space of* X.
- *A sequence* $(y_n) \subset Y = X^*$ *converges weakly-* to* $y \in Y$ *if for all* $x \in X$, $\lim_{n\to\infty} \langle x, y_n\rangle_{X\times Y} = \langle x, y\rangle_{X\times Y}$.

Remark 1. Any Hilbert space $(\mathcal{H}, \|\cdot\|)$ is lower semicontinuous wrt weak topology: i.e. if $x_n, x \in \mathcal{H}$ are such that $x_n \rightharpoonup x$, then $\liminf_{n\to\infty} \|x_n\| \geqslant \|x\|$.
Definition 2. *Let* $(X, \|\cdot\|_X)$ *be a normed space. A subset* $U \subset X$ *is weak sequentially compact if every sequence in* U *has a weakly converging subsequence with limit in* U.
Theorem 1. *A normed vector space* X *is reflexive if and only if every bounded ball in* X *is weak-sequentially compact.*
Remark 2. Any Hilbert space is reflexive (so that includes L^2 and $W^{1,2}$). On the other hand, L^1 and $W^{1,1}$ are not reflexive.
Definition 3. *Let* $k \in [0, \infty]$ *and* $A \subset \mathbb{R}^d$. *The* k*-dimensional Hausdorff measure of* A *is given by*

$$\mathcal{H}^k(A) = \lim_{\rho\to 0} \left[\frac{\omega_k}{2^k} \inf \left\{ \sum_{i\in I} |\mathrm{diam}(A_i)|^k \ ; \ \mathrm{diam}(A_i) \leqslant \rho, A \subset \bigcup_{i\in I} A_i \right\} \right]$$

where $\mathrm{diam}(A_i)$ *denotes the diameter of the set* A_i *and* $\omega_k = \pi^{k/2}\Gamma(1 + k/2)$ *where* Γ *is the gamma function. Note that* ω_k *is the Lebesgue measure of the unit ball in* $\mathbb{R}^k$ *for* $k \in \mathbb{N}$.

The Hausdorff dimension of a set A *is*

$$\inf \left\{ k \geqslant 0 \ ; \ \mathcal{H}^k(A) = 0 \right\}.$$

Remark 3. For $k \in [1, N]$, given $A \subset \mathbb{R}^N$, a C^1 k-dimensional manifold in $\mathbb{R}^N$, $\mathcal{H}^k(A)$ is the classical k-dimensional area of A. For example, if A is a finite set of points, $\mathcal{H}^0(A)$ is the number of points; if A is a C^1-curve, then $\mathcal{H}^1(A)$ is the length of this curve. Moreover, $\mathcal{H}^N(A)$ of a set $A \subset \mathbb{R}^N$ is its Lebesgue measure.

We recall the Gauss-Green theorem:
Theorem 2. *If* E *is an open set with* C^1 *boundary, then for every* $\varphi \in C^1_c(\mathbb{R}^n)$,

$$\int_E \nabla\varphi(x)\, \mathrm{d}x = \int_{\partial E} \varphi\, \nu_E\, \mathrm{d}\mathcal{H}^{n-1}.$$

Equivalently, the divergence theorem also holds true:

$$\int_E \mathrm{div}\, z(x)\, \mathrm{d}x = \int_{\partial E} z \cdot \nu_E\, \mathrm{d}\mathcal{H}^{n-1}, \qquad \forall z \in C^1_c(\mathbb{R}^n; \mathbb{R}^n).$$

Definition 4. *Let* X *be a Banach space, let* $F : X \to \mathbb{R}$. *If*

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

exists, then it is called the directional derivative of F *at* u *in direction* v. *If there exists* $\overline{u} \in X'$ *(the dual space of* X*) such that* $F'(u; v) = \langle \overline{u}, v\rangle_{X'\times X}$ *for all* $v \in X$*, then we say that* F *is Gâteaux differentiable at* u *and write* $F'(u) = \overline{u}$.
Lemma 1. *If* F *is Gâteaux differentiable and* $\inf_{v\in X} F(v)$ *has a solution* u*, then* $F'(u) = 0$*. Conversely, if* F *is convex, then a solution* u *of* $F'(u) = 0$ *is a solution of the minimization problem. We call the equation* $F'(u) = 0$ *the* ***Euler Lagrange equation*** *of* F.

Proof. Indeed, if $F(u) = \min_x F(x)$ and F is Gâteaux differentiable, then $F(u+tv)-F(u) \geqslant 0$ for all $v \in X$. So, $\langle v,\, F'(u)\rangle \geqslant 0$ for all $v \in X$. By considering $-v$, we have that $\langle v,\, F'(u)\rangle = 0$ for all $v \in X$, so $F'(u) = 0$.

Conversely, if F is convex and $F'(u) = 0$, then

$$F(u+v) - F(u) \geqslant \lim_{t \to 0} \frac{F(u+tv) - F(u)}{t} = 0, \qquad \forall v \in X.$$

by monotonicity of difference quotients for convex functions. Therefore, letting $v = w - u$, $F(w) \geqslant F(u)$ for all $w \in X$ which implies that u is a minimizer.

□

Lemma 2 (Monotonicity of difference quotients)**.** *Let X be a normed space and let $F : X \to \mathbb{R}$ be a convex functional. For $u, v \in X$ and $t \geqslant 0$, let $dF(t) = \frac{F(u+tv)-F(u)}{t}$. Then, $dF(t) \leqslant dF(s)$ whenever $t < s$.*

Proof. Without loss of generality, assume that $u = 0$ and $F(0) = 0$ (this is possible by translating f appropriately). Then, by convexity of F,

$$F(tv) = F\left(\frac{t}{s}sv + (1 - \frac{t}{s})0\right) \leqslant \frac{t}{s}F(sv) + (1 - \frac{t}{s})F(0) = \frac{t}{s}F(sv),$$

as required. □

We recall also some definitions:

1. A Borel set is any set formed from open sets by countable union, intersection an relative complement.
2. A Sigma algebra is a collection of sets which are closed under union, intersection and relative complement.
3. A measure μ is said to be locally finite in Ω if given any $x \in \Omega$, there exists a neighbourhood of x, N_x such that $\mu(N_x) < \infty$.
4. A measure is μ is said to be inner regular on some Sigma algebra Σ, if given $A \in \Sigma$

$$\mu(A) = \sup\left\{\mu(K)\ ;\ K \subset A,\ K \text{ compact}\right\}.$$

5. A Radon measure is a measure on the Sigma algebra of Borel sets of a Hausdorff topological space X which is locally finite and inner regular.

3 The TV functional

Definition 5. *Given $u \in L^1(\Omega)$, the total variation of u is*

$$J(u) = \sup\left\{\langle \operatorname{div} z,\, u\rangle\ ;\ z \in C_c^\infty(\Omega, \mathbb{R}^N), \|z\|_\infty \leqslant 1\right\}.$$

Definition 6. *The space $BV(\Omega)$ of functions of bounded variation is the set of functions $u \in L^1(\Omega)$ such that $J(u) < \infty$ endowed with the norm $\|u\|_{BV} = \|u\|_{L^1} + J(u)$.*

This space is a Banach space. The following theorem shows that the space $BV(\Omega)$ is compact. In contrast, note that no such compactness result exists for $W^{1,1}(\Omega)$ since one can in fact construct bounded sequences in $W^{1,1}(\Omega)$ which converge to elements of $BV(\Omega)$.
Theorem 3 (Rellich's compactness theorem)**.** *Let $\Omega \subset \mathbb{R}^N$ be a bounded domain with Lipschitz boundary, and let $(u_n)_{n\in\mathbb{N}}$ be a sequence of functions in $BV(\Omega)$ such that $\sup_n \|u_n\|_{BV} < \infty$. Then there exists $u \in BV(\Omega)$ and a subsequence $(u_{n_k})_{k\geqslant 1}$ such that $u_{n_k} \to u$ in $L^1(\Omega)$ as $k \to \infty$.*

Theorem 4. *Let $\Omega \subset \mathbb{R}^N$ be a bounded domain with Lipschitz boundary, let $u \in BV(\Omega)$. Then, there exists a sequence (u_n) of functions in $C^\infty(\Omega) \cap W^{1,1}(\Omega)$, such that*

1. *$u_n \to u$ in L^1.*
2. *$J(u_n) = \int_\Omega |\nabla u| \to J(u) = \int_\Omega |Du|$.*

Theorem 5 (Poincaré inequality)**.** *Let $\Omega \subset \mathbb{R}^N$. For $u \in BV(\Omega)$, let $m(u) = \frac{1}{|\Omega|}\int_\Omega u(x)\mathrm{d}x$. Then there exists $C > 0$ such that*

$$\|u - m(u)\|_{L^p} \leqslant CJ(u), \qquad \forall u \in BV(\Omega),$$

for all $p \in [1, N/(N-1)]$. In particular, this holds with $p = 2$ when $N = 2$.

Lemma 3. *(i) J is lower semicontinuous wrt weak convergence in L^p for $p \in [1, \infty)$.*

(ii) J is convex.

(iii) J is one-homogeneous. i.e. $tJ(u) = J(tu)$ for all $t \geqslant 0$.

Proof. Let

$$L_\varphi : u \mapsto -\int_\Omega u(x) \operatorname{div} \varphi(x)\mathrm{d}x.$$

If $u_n \rightharpoonup u$ in $L^p(\Omega)$, then $L_\varphi u_n \to L_\varphi u$. Note however that

$$L_\varphi u = \lim_{n\to\infty} L_\varphi u_n \leqslant \liminf_{n\to\infty} J(u_n).$$

Taking the supremum over all $\varphi \in C_c^\infty(\Omega, \mathbb{R}^N)$ with $\|\varphi\|_\infty \leqslant 1$ yields

$$J(u) \leqslant \liminf_{n\to\infty} J(u_n).$$

To see that J is convex, let $u_1, u_2 \in L^p(\Omega)$ and let $t \in [0, 1]$. Then,

$$L_\varphi(tu_1 + (1-t)u_2) = tL(u_1) + (1-t)L(u_2) \leqslant tJ(u_1) + (1-t)J(u_2).$$

Taking the supremum over all $\varphi \in C_c^\infty(\Omega, \mathbb{R}^N)$ with $\|\varphi\|_\infty \leqslant 1$ yields the required result. □

Examples

- For $u \in W^{1,1}(\Omega)$, $|Du|(\Omega) = \int_\Omega |\nabla u|$.
- For $u = \mathbb{1}_C$ with $C \subset \Omega$ a bounded set with smooth boundary, $|Du|(\Omega) = |Du|(C) = \mathcal{H}^1(\partial C)$ is the perimeter of C. Here, $\mathcal{H}^1$ is the 1 dimensional Hausdorff measure.

Lecture 18

4 Total variation regularization

Let $\Omega \subset \mathbb{R}^2$ be a bounded Lipschitz domain. Given $g \in L^2(\Omega)$, we want to compute

$$\inf_{u \in L^2(\Omega)\cap BV(\Omega)} \mathcal{F}(u), \qquad \mathcal{F}(u) := \lambda J(u) + \frac{1}{2}\|Tu - g\|_{L^2}^2, \tag{1}$$

where $T : L^2(\Omega) \to L^2(\Omega)$ is a bounded linear operator. Examples of operators T include $T = k\star$, $T = \mathrm{Id}$, $T = \chi_{\Omega\setminus D}$ and the subsampled Fourier operator $T = P_S\mathcal{F}$.

4.1 Existence and uniqueness of solutions

We first consider the well-posedness of this problem.

Theorem 6. *For $g \in L^2(\Omega)$ and a bounded linear operator $T : L^2(\Omega) \to L^2(\Omega)$, there exists a minimizer $u \in BV(\Omega)$ of (1). If T is injective, then the minimizer is unique.*

Direct method of calculus of variations Let X be a Banach space and consider $\min_{u\in X} \mathcal{F}(u)$. To show that there exists a minimizer, we carry out the following steps:

1. Show that $\mathcal{F}(u)$ is bounded from below, that is $\inf \mathcal{F}(u) > -\infty$. Hence, there exists a minimizing sequence u_n such that $\mathcal{F}(u_n) < \infty$ and

$$\lim_{n\to\infty} \mathcal{F}(u_n) = \inf_{u\in L^2(\Omega)} \mathcal{F}(u).$$

2. Check that $(u_n) \subset Y \subset X$ with Y sequentially compact w.r.t. topology induced on X. Then, there exists a subsequence u_{n_k} and $u_* \in X$ such that $\lim_{k\to\infty} u_{n_k} = u_*$.
3. Check that $\mathcal{F}$ is sequentially lower semicontinuous w.r.t. topology on X. Then,

$$\inf_{u\in X} \mathcal{F}(u) \leqslant \mathcal{F}(u_*) \leqslant \liminf \mathcal{F}(u_{n_k}) = \inf_{u\in X} \mathcal{F}(u).$$

So, $u_* \in X$ is a minimizer of $\mathcal{F}$.

Proof. Let us check each of the three steps in the direct method. The first step is trivial since $\mathcal{F}(u) \geqslant 0$ for all $u \in L^2(\Omega)$. For the second step, note that given a minimizing sequence $u_n \in L^2(\Omega)$, for fixed $\varepsilon > 0$ and n sufficiently large, $\mathcal{F}(u_n) \leqslant \mathcal{F}(0) + \varepsilon =: C$. So,

$$\lambda J(u_n) \leqslant \lambda J(u_n) + \frac{1}{2}\|Tu_n - g\|_{L^2}^2 \leqslant C,$$

and $(J(u_n))_{n\in\mathbb{N}}$ is a uniformly bounded sequence. For the purpose of extracting a convergent subsequence, we first prove that $\left|\int_\Omega u_n\right|$ is uniformly bounded: Assume first that $T\chi_\Omega \neq 0$. For $n \geqslant 1$, let $w_n = \frac{\int u_n}{|\Omega|}\chi_\Omega$ and $v_n = u_n - w_n$. Then, $\int v_n = 0$ and $J(v_n) = J(u_n)$. So, by the Poincaré inequality, $\|v_n\|_2 \leqslant C'$. Observe now that

$$\sqrt{2C} \geqslant \|Tu_n - g\|_2 \geqslant \|Tw_n\|_2 - \|Tv_n - g\|_2,$$

and hence,

$$\sqrt{2C} + \|T\|_2\|v_n\|_2 + \|g\|_2 \geqslant \|Tw_n\|_2 = \left|\int u_n\right| \frac{\|T\chi_\Omega\|_2}{|\Omega|}.$$

Since $\|T\chi_\Omega\|_2 \neq 0$, it follows that $\left|\int u_n\right|$ is uniformly bounded.

Now, by the Poincaré inequality, u_n is uniformly bounded in L^2 and since the L^2 ball is weak sequentially compact, there exists a subsequence $(u_{n_k})_{k\in\mathbb{N}}$ and $u \in L^2$ such that $u_{n_k} \rightharpoonup u$ in L^2. Furthermore, since T is a bounded linear operator on $L^2(\Omega)$, we also have that $Tu_{n_{k_j}} \rightharpoonup Tu$ in L^2. Finally, the third step follows since $\|\cdot\|_{L^2}$ and J are lower semi-continuous wrt the weak topology.

Now, if $T\chi_\Omega = 0$, then $v_n = u_n - w_n$ is also a minimizing sequence and by the Poincaré inequality, $\|v_n\|_{L^2}$ is uniformly bounded and we may exact find a weak limit u which is a minimizer.

To see that the minimizer u is unique in the case where T is injective, recall from Lemma 3 that J is a convex functional. So, if u, u' are two minimizer of $\mathcal{F}$, then

$$\mathcal{F}\left(\frac{u+u'}{2}\right) = \frac{\lambda}{2}J(u) + \frac{\lambda}{2}J(u') + \frac{1}{2}\int\left|\frac{Tu+Tu'}{2} - g\right|^2 \mathrm{d}x.$$

If $u \neq u'$, so $Tu \neq Tu'$, then by strict convexity of $\|\cdot\|_2^2$,

$$\mathcal{F}\left(\frac{u+u'}{2}\right) < \frac{\lambda}{2}J(u) + \frac{\lambda}{2}J(u') + \frac{1}{4}\int |Tu-g|^2 + \frac{1}{4}\int |Tu'-g|^2 = \frac{1}{2}(\mathcal{F}(u)+\mathcal{F}(u')) = \inf \mathcal{F}.$$

This is a contradiction to the assumption that u and u' are minimizers.

□

4.2 Optimality conditions

We begin with a generalization of the gradient for convex functionals.

Definition 7. *Given a Banach space X with dual space X', the subgradient of a convex functional $F : X \to (-\infty, +\infty]$ is a set-valued map which maps $x \in X$ to*

$$\partial F(x) = \{p \in X' \; ; \; F(y) \geqslant F(x) + \langle p, \, y - x\rangle, \forall y \in X\} \, .$$

Remark 4 (Relation to directional derivatives). Note that $y \in \partial F(x)$ if and only if $F'(x, z) \geqslant \langle y, \, z\rangle_{X' \times X}$ for all $z \in X$. Moreover, if F is Gateaux differentiable to x with $F'(x) = y$, then $\partial F(x) = \{y\}$.

Remark 5. Note that u is a minimizer of $\inf_v F(v)$ if and only if $0 \in \partial F(u)$: Indeed, $0 \in \partial F(u)$ means that

$$F(v) - F(u) \geqslant 0 = \langle 0, \, v - u\rangle, \qquad \forall v \in X.$$

Lecture 19

Recall that for each $u \in L^1(\Omega)$, $J(u) = \sup_{p \in \mathcal{K}} \int_\Omega u(x)p(x)\mathrm{d}x$ where

$$\mathcal{K} = \left\{-\operatorname{div} \varphi \; ; \; \varphi \in C_c^\infty(\Omega; \mathbb{R}^N), \|\varphi\|_\infty \leqslant 1\right\}.$$

However, if $u \in L^2(\Omega)$, then we in fact have:

$$J(u) = \sup_{p \in K} \int_\Omega u(x)p(x)\mathrm{d}x,$$

where

$$K = \left\{-\operatorname{div} \varphi \; ; \; \varphi \in L^\infty(\Omega, \mathbb{R}^N), -\operatorname{div}\varphi \in L^2(\Omega), \varphi \cdot \eta_\Omega = 0\right\}.$$

In the definition of K, $-\operatorname{div}\varphi \in L^2(\Omega)$ means that there exists $\gamma \in L^2(\Omega)$ such that

$$\int_\Omega \gamma u = \int z \cdot \nabla u, \qquad \forall u \in C_c^\infty(\Omega).$$

One can show that

Lemma 4.

$$K = \left\{p \in L^2(\Omega) \; ; \; \int_\Omega p(x)u(x)\mathrm{d}x \leqslant J(u), \quad \forall u \in L^2(\Omega)\right\}.$$

Lemma 5 (Characterization of subgradient)**.** *Consider $J : L^2(\Omega) \to (-\infty, +\infty]$ and let $u \in L^2(\Omega) \cap BV(\Omega)$. Then $\partial J(u) = \left\{p \in K \; ; \; \int p(x)u(x)\mathrm{d}x = J(u)\right\}$.*

Another way of writing $\int -\operatorname{div} z \; u = J(u)$ is $z \cdot Du = |Du|$. From this, we see that z is the unit normal to the level lines of u.

Proof. Let $p \in K$ be such that $\int pu = J(u)$. Then, for all $v \in L^2(\Omega)$,

$$J(v) \geqslant \int pv = J(u) + \int pv - \int pu.$$

This implies that $p \in \partial J(u)$.

Conversely, if $p \in \partial J(u)$, then for all $t > 0$ and $v \in L^2(\Omega)$,

$$tJ(v) = J(tv) \geqslant J(u) + \int_\Omega p(tv - u).$$

By dividing by t and letting $t \to \infty$, we have that

$$J(v) \geqslant \int pv \implies p \in K.$$

Moreover, by letting $t \to 0$, we have that

$$\int pu \geqslant J(u)$$

and since $p \in K$, it follows that $J(u) = \int pu$. ☐

Theorem 7 (Optimality condition)**.** *$u \in L^2(\Omega)$ is a minimizer of $\mathcal{F}$ if and only if*

$$\frac{T^*(g - Tu)}{\lambda} \in \partial J(u).$$

Remark 6. Given the characterization of $\partial J(u)$, the equation satisfied by the minimizer u of $\mathcal{F}$ can formally be written as

$$\frac{T^*(g - Tu)}{\lambda} + \operatorname{div}\left(\frac{Du}{|Du|}\right) = 0.$$

This is called the **Euler-Lagrange equation** of $\mathcal{F}$.

Proof. Suppose that u minimizes $\mathcal{F}$. Then, for all $v \in L^2(\Omega)$, we have that

$$\begin{aligned} \lambda J(v) &\geqslant \lambda J(u) + \frac{1}{2}\int_\Omega (Tu - g)^2 - (Tv - g)^2 \mathrm{d}x \\ &= \lambda J(u) + \int_\Omega (Tu - Tv)\left(\frac{Tu + Tv}{2} - g\right)\mathrm{d}x \\ &= \lambda J(u) + \int_\Omega (Tv - Tu)(g - Tu)\mathrm{d}x - \frac{1}{2}\int_\Omega (Tu - Tv)^2 \mathrm{d}x. \end{aligned} \tag{2}$$

Let $w \in L^2(\Omega)$ and let $v = u + t(w - u)$ in the above inequality. Then,

$$\begin{aligned} \lambda(J(u + t(w - u)) - J(u)) - t\int_\Omega (Tw - Tu)(g - Tu)\mathrm{d}x \\ \geqslant -\frac{t^2}{2}\int_\Omega (Tw - Tu)^2 \mathrm{d}x. \end{aligned}$$

Dividing by t and letting $t \to 0$, we have that

$$\lim_{t \to 0} \frac{\lambda(J(u + t(w - u)) - J(u))}{t} - \int_\Omega (Tw - Tu)(g - Tu)\mathrm{d}x \geqslant 0.$$

Since J is convex, for all $t > 0$, by its difference quotient being non-increasing as $t \to 0_+$, we have that

$$\frac{\lambda(J(u + t(w - u)) - J(u))}{t} - \int_\Omega (w - u)T^*(g - Tu)\mathrm{d}x \geqslant 0.$$

For $t = 1$, we have that $\frac{T^*(g-Tu)}{\lambda} \in \partial J(u)$. Conversely, if this is true, then (2) holds and u is a minimizer of $\mathcal{F}$. ☐

4.3 Further remarks about "$-\operatorname{div} z$"

There are several works which address the question of how we should interpret $\int \operatorname{div} z \; u$ for $u \in BV$, where $\operatorname{div} z \in L^2(\Omega)$ and $z \in L^\infty(\Omega, \mathbb{R}^N)$. Anzellotti (*Pairings between measures and bounded functions and compensated compactness*, 1983) showed that for $u \in BV \cap L^2$, we can define a functional $(z, Du) : C_c^\infty(\Omega) \to \mathbb{R}$ such that

$$\langle (z, Du), \varphi \rangle = -\int u\varphi \operatorname{div} z - \int uz \cdot \nabla\varphi, \qquad \forall \varphi \in C_c^\infty(\Omega).$$

Then, (z, Du) is a Radon measure in Ω such that:

- if $u \in W^{1,1}(\Omega) \cap L^2(\Omega)$, then $(z, Du) = z \cdot \nabla u$.
- $\left|\int_B (z, Du)\right| \leqslant \int_B |(z, Du)| \leqslant \|z\|_{L^\infty(A)} \int_B |Du|$ for all Borel sets B and open sets A such that $B \subset A$.
- Given any bounded set A with Lipschitz boundary, we have the divergence formula:

$$\int_A u \; \operatorname{div} z \, \mathrm{d}x + \int_A (z, Du) = \int_{\partial A} (z \cdot \nu_{\partial A}) \, u \, \mathrm{d}\mathcal{H}^{N-1}, \qquad \forall u \in BV(\Omega)$$

4.4 Stability

Let $u_0 \in L^2(\Omega)$ and $g \in L^2(\Omega)$ be such that $\|g - Tu_0\|_{L^2} \leqslant \delta$. We saw in Theorem 7 that if u is a minimizer of (1), then there exists p such that $T^*p \in \partial J(u)$. Here, we show that the existence of such an element (in which case, one says that *the source condition* is satisfied) can allow for the derivation of stability estimates of the reconstructed image.

In the following, we assume that there exists $p \in L^2$ such that

$$v = T^*p \in \partial J(u_0).$$

The first result, Theorem 8, shows that the *Bregman distances* $J(u) - J(u_0) - \langle v, u - u_0 \rangle$ (which is always positive) can be controlled in terms of δ, $\|p\|$ and λ.

Furthermore, since $v \in \partial J(u_0)$, as mentioned in Remark 4, there exists $z \in L^\infty(\mathbb{R}^2; \mathbb{R}^2)$ such that $v = -\operatorname{div} z$ with $\|z\|_\infty \leqslant 1$. We will in Theorem 9 that one can control the total variation of u in the region U_r where for $r \in (0, 1)$

$$U_r \stackrel{\text{def.}}{=} \{x \in \Omega \; ; \; |z(x)| < r\}.$$

Theorem 8. *Let $u_0 \in L^2(\Omega)$ and $g \in L^2(\Omega)$ be such that $\|g - Tu_0\|_{L^2} \leqslant \delta$. Suppse that u is a minimizer of* (1) *and that there exists $v = T^*p \in \partial J(u_0)$. Then,*

$$J(u) - J(u_0) - \langle v, u - u_0 \rangle \leqslant \frac{\delta^2}{2\lambda} + \frac{\lambda \|p\|_{L^2}^2}{2} + \delta \|p\|_{L^2}.$$

Proof. Let $d = J(u) - J(u_0) - \langle v, u - u_0 \rangle$. Since u is a minimizer of (1),

$$\lambda J(u) + \frac{1}{2}\|Tu - g\|_{L^2}^2 \leqslant \lambda J(u_0) + \frac{1}{2}\|Tu_0 - g\|_{L^2}^2 \leqslant \lambda J(u_0) + \frac{\delta^2}{2}.$$

So,

$$\frac{1}{2}\|Tu - g\|_{L^2}^2 + \lambda d + \lambda \langle v, u - u_0 \rangle \leqslant \frac{\delta^2}{2}$$

By recalling that $v = T^*p$,

$$\frac{1}{2}\|Tu - g\|_{L^2}^2 + \lambda d + \lambda \langle p, Tu - g \rangle + \lambda \langle p, g - Tu_0 \rangle \leqslant \frac{\delta^2}{2}$$

So,

$$\frac{1}{2}\|Tu - g + \lambda p\|_{L^2}^2 + \lambda d - \frac{\lambda^2 \|p\|_{L^2}^2}{2} + \lambda \langle p,\, g - Tu_0\rangle \leqslant \frac{\delta^2}{2},$$

and by rearranging the above inequality,

$$d \leqslant \frac{\delta^2}{2\lambda} + \frac{\lambda \|p\|_{L^2}^2}{2} + \delta \|p\|_{L^2}.$$

$\square$

Remark 7. Note that we did not use any specific properties of J in the proof other than convexity. In fact, Theorem 8 can be stated more generally: it is sufficient to assume that for Banach space X and Hilbert space H, $T : X \to H$ is a bounded linear operator and $J : X \to (-\infty, +\infty]$ is a convex functional such that

(1) J is l.s.c. in a topology on X.

(2) The sublevel sets $\{u;\ J(u) \leqslant \rho\}$ are compact (wrt same topology as in (1)) and nonempty for all $\rho \geqslant 0$.

These conditions essentially ensure the existence of a minimizer via the direct method of calculus.

In particular, we can consider $J(u) = \frac{1}{2}\|u\|_{L^2}$, and

$$\min_{u \in L^2(\Omega)} \frac{\lambda}{2}\|u\|_{L^2}^2 + \frac{1}{2}\|Tu - g\|_{L^2}^2,$$

In this case, $\partial J(u_0) = u_0$ and the minimizer u satisfies

$$\frac{1}{2}\|u - u_0\|_{L^2}^2 = \frac{1}{2}\|u\|_{L^2}^2 - \frac{1}{2}\|u_0\|_{L^2}^2 - \langle u_0,\, u - u_0\rangle \leqslant \frac{\delta^2}{2\lambda} + \frac{\lambda \|u_0\|_{L^2}^2}{2} + \delta \|u_0\|_{L^2} = \mathcal{O}(\delta)$$

provided that $\lambda \sim \delta$.

Theorem 9. *Consider the setting of Theorem 8 and $v = -\operatorname{div} z$ with $\|z\|_\infty \leqslant 1$. Let*

$$U_r \overset{\text{def.}}{=} \{x \in \Omega\ ;\ |z(x)| < r\}.$$

For each $r \in (0, 1)$,

$$(1 - r)\int_{U_r} |Du| \leqslant \frac{\delta^2}{2\lambda} + \frac{\lambda \|p\|_{L^2}^2}{2} + \delta \|p\|_{L^2}.$$

Proof.

$$\begin{aligned}
d &:= J(u) - J(u_0) - \langle v,\, u - u_0\rangle \\
&= J(u) - J(u_0) + \langle \operatorname{div} z,\, u\rangle - \langle \operatorname{div} z,\, u_0\rangle \\
&= J(u) + \langle \operatorname{div} z,\, u\rangle \qquad\qquad \text{since } J(u_0) = \langle -\operatorname{div} z,\, u_0\rangle \\
&= J(u) - \int (z, Du) \\
&= J(u) - \int_{\Omega\setminus U_r} (z, Du) - \int_{U_r} (z, Du) \\
&\geqslant J(u) - \int_{\Omega\setminus U_r} |Du| - r\int_{U_r} |Du| \\
&\geqslant (1 - r)\int_{U_r} |Du|.
\end{aligned}$$

The conclusion now follows by applying the upper bound on d from Theorem 8. $\square$

Example Let us consider the case of denoising. Then, the proof of Theorem 8 actually yields

$$\int_\Omega (u-u_0)^2 + |J(u)-J(u_0)| \leqslant C\delta(\|p\|_{L^2}+1) \tag{3}$$

provided $\lambda = \delta/\|p\|_{L^2}$. Let $B_R \subset \mathbb{R}^2$ be the ball of radius R with origin 0 and let $u_0 = \chi_{B_R}$. Then let $p = -\operatorname{div}(z)$ where z is defined by

$$z(x) = \frac{q\left(\Big| |x| - R\Big|\right)}{|x|}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix}, \qquad q(s) = \max\{1 - s/\varepsilon, 0\}.$$

(In polar coordinates (r,θ), we can write $z(r,\theta) = q(|r-R|)\binom{\cos(\theta)}{\sin(\theta)})$. One can show that $\|p\| = \mathcal{O}(\varepsilon^{-1/2})$. Then, by choosing $U = \{x \in \Omega\ ;\ \operatorname{dist}(x,\partial B_R) \geqslant \varepsilon\}$, the minimizer u satisfies

$$\int_U |Du| \leqslant \mathcal{O}\left(\frac{\delta^2}{\lambda} + \frac{\lambda}{\varepsilon} + \frac{\delta}{\varepsilon}\right) = \mathcal{O}\left(\frac{\delta}{\sqrt{\varepsilon}}\right)$$

provided that $\lambda = \delta\sqrt{\varepsilon}$.

Combining with (3) yields

$$\int_{U^c} |Du| \geqslant J(u) - \int_U |Du| \geqslant 2\pi R - C\left(\delta + \frac{\delta}{\sqrt{\varepsilon}}\right).$$

Therefore, most of the total variation of u is concentrated around ∂B_R and the above theorem points to the ability of TV regularization in dampening oscillations away from the true edge ∂B_R.

Lecture 20

5 Numerical algorithms

We present some algorithms for solving (1) in the special case where $T = \mathrm{Id}$.

In this section, we consider a discretization of our minimization problem.

Let $\Omega = (0,1)^2$ and let

$$TV_h(u) = h^2 \sum_{i,j=1}^N \frac{\sqrt{|u_{i+1,j}-u_{i,j}|^2 + |u_{i,j+1}-u_{i,j}|^2}}{h}$$

where in the sum, differences are replaced by 0 when one of the points is not on the grid $\{i,j = 1,\ldots,N\}$. Moreover, $u = (u_{i,j})$ is the discrete image and $h = 1/N$ is the discretization step.

One can show that TV_h correctly discretizes the functional J:

Proposition 1. *Let $\Omega = (0,1)^2$ and let $p \in [1,\infty)$. Let $G : L^p(\Omega) \to \mathbb{R}$ be a continuous functional such that $\lim_{c\to\infty} G(c+u) = +\infty$ for any $u \in L^p(\Omega)$ (this coerciveness assumption simply ensures the existence of minimizers, other more general conditions can also be considered). Let $h = 1/N > 0$ and let $u^h = (u_{i,j})_{i,j=1}^N$, identified with the function*

$$\overline{u}^h(x) = \sum_{i,j=1}^N \chi_{((i-1)h,ih)\times((j-1)h,jh)}(x)u_{i,j},$$

be a solution of $\min_{u^h} TV_h(u^h) + G_h(u^h)$ where $G_h(u^h) = G(\overline{u}^h)$. Then, there exists $u \in L^p(\Omega)$ and some subsequence $u^{h_k} \to u$ strongly in L^1 as $k \to \infty$ such that u is a minimizer in $L^p(\Omega)$ to $J(u) + G(u)$.

For simplicity, we now let $h = 1$, and define discrete gradient operators as follow: Let $X = \mathbb{R}^{N\times N}$ and let $Y = X \times X$. $\nabla : X \to Y$ is defined by

$$(\nabla u)_{i,j} = \begin{pmatrix} (D_x^+ u)_{i,j} \\ (D_y^+ u)_{i,j} \end{pmatrix}, \quad (D_x^+ u)_{i,j} = \begin{cases} u_{i+1,j} - u_{i,j} & i < N \\ 0 & i = N, \end{cases} \quad (D_y^+ u)_{i,j} = \begin{cases} u_{i,j+1} - u_{i,j} & j < N \\ 0 & j = N. \end{cases}$$

We denote the adjoint of ∇ by $\nabla^* = -\operatorname{div} : Y \to X$. So, given $p = (p^x, p^y) \in Y$,

$$(\operatorname{div} p)_{i,j} = (p^x_{i,j} - p^x_{i-1,j}) + (p^y_{i,j} - p^y_{i,j-1})$$

where the first difference is replaced by $p^x_{i,j}$ if $i = 1$ and by $-p^x_{i-1,j}$ if $i = N$, and the second difference is replaced by $p^y_{i,j}$ if $j = 1$ and by $-p^y_{i,j-1}$ if $j = N$.

We now consider the minimization problem

$$\min_{u\in X} \lambda \|\nabla u\|_{2,1} + \frac{1}{2}\|u - g\|_2^2, \tag{4}$$

where $\|p\|_{2,1} = \sum_{i,j} \sqrt{(p^x_{i,j})^2 + (p^y_{i,j})^2}$.

This problem can also be written in the general form

$$\min F(Au) + G(u) \tag{5}$$

where $F : Y \to \mathbb{R}$ and $G : X \to \mathbb{R}$ are convex, $A : X \to Y$ is a linear operator. Here, $F = \|\cdot\|_{2,1}$, $G = \frac{1}{2}\|\cdot - g\|_2^2$ and $A = \nabla$.

5.1 Gradient descent

If we want to minimize a proper, convex, lsc functional F which is differentiable and such that ∇F has Lipschitz constant L: $|\nabla F(x) - \nabla F(y)| \leqslant L\|x - y\|$ for all $x, y \in X$, then one can apply the gradient descent algorithm:

Choose a step size $h > 0$ and any $u^0 \in X$, and let for any $n \geqslant 0$

$$u^{n+1} = u^n - h\nabla F(u^n).$$

Provided that h is sufficiently small, the sequence u^n will converge to a minimizer:

Theorem 10. *Let* $\min F = F(u^*)$. *Assume* $h \in (0, 2/L)$. *Then* $F(u^k) \to F(u^*)$ *as* $k \to \infty$. *The best rate of convergence is obtained for* $h = 1/L$ *and is*

$$F(u^k) - F(u^*) \leqslant \frac{2L\|u^0 - u^*\|^2}{k+4}.$$

The objective function in (4) is not differentiable, however, it can be approximated by

$$F_\varepsilon(u) = \lambda \sum_{i,j} \sqrt{\varepsilon^2 + |(\nabla u)_{i,j}|^2} + \frac{1}{2}\|u - g\|^2, \qquad \varepsilon > 0,$$

and one can show convergence of minimizers of this functional to minimizer of (4). Now,

$$\nabla F_\varepsilon(u) = -\lambda \operatorname{div}\left(\frac{\nabla u}{\sqrt{\varepsilon^2 + |\nabla u|^2}}\right) + u - g$$

and has Lipschitz constant $\mathcal{O}(\varepsilon^{-1})$. Therefore, the iterates of the gradient descent algorithm are

$$u^{n+1} = u^n + h\lambda \operatorname{div}\left(\frac{\nabla u^n}{\sqrt{\varepsilon^2 + |\nabla u^n|^2}}\right) - h(u^n - g),$$

with $h \sim \varepsilon^{-1}$ and we are in fact approximating the PDE

$$\partial_t u = \lambda \operatorname{div}\left(\frac{\nabla u}{\sqrt{\varepsilon^2 + |\nabla u|^2}}\right) - (u - g).$$

In practice, this algorithm is very slow as we are required to take an extremely small stepsize.

5.2 A dual formulation

This section derives a dual formulation of the problem (4).

5.2.1 Notions from convex analysis

Let us first recall some notions from convex analysis.
Definition 8. *Let $F : X \to \mathbb{R}$. The Legendre-Fenchel conjugate F^* of F for any $p \in X$ is*

$$F^*(p) = \sup_{x \in X} \langle p,\, x\rangle - F(x).$$

Note that F^* is convex and is lower semicontinuous (since it is the supremum of linear, continuous functions). It is also proper if F is convex and proper.

An example Let $G(x) = \|x - g\|_2^2$. Then,

$$G^*(y) = \sup_x \langle x,\, y\rangle - \frac{1}{2}\|x-g\|^2 = \sup_x \langle x - g,\, y\rangle - \frac{1}{2}\|x-g\|^2 + \langle g,\, y\rangle = \frac{1}{2}\|y\|^2 - \langle g,\, y\rangle = \frac{1}{2}\|y+g\|^2 - \frac{\|g\|^2}{2}.$$

Indeed, the penultimate equality follows because

- $\langle x - g,\, y\rangle - \frac{1}{2}\|x-g\|^2 \leqslant \frac{1}{2}\|y\|^2$ for all x, y, g. So, $\sup_x \langle x - g,\, y\rangle - \frac{1}{2}\|x-g\|^2 \leqslant \frac{1}{2}\|y\|^2$.
- Setting $x = y + g$, we have that $\langle x - g,\, y\rangle - \frac{1}{2}\|x-g\|^2 = \frac{1}{2}\|y\|^2$.

Theorem 11.
- *If F is proper, convex and lsc, then $F^{**} = F$.*
- *For any convex F, $p \in \partial F(x)$ if and only if $\langle p,\, x\rangle - F(x) = F^*(p)$. Furthermore, if we additionally assume that F is proper and lsc, then since $F^{**} = F$, we have that $x \in \partial F^*(p)$.*
- *Let F be convex. Then, u minimizes F if and only if $0 \in \partial F(u)$.*
- *Let F, G be convex, if $\operatorname{int}(\operatorname{dom}(G)) \cap \operatorname{dom}(F) \neq \emptyset$, then*

$$\partial(F + G) = \partial F + \partial G.$$

($\operatorname{dom}(F)$ is the set of points where $F(x) < \infty$ and note also that in general, $\partial(F + G) \supset \partial F + \partial G$.)

Proximal mappings If F is a convex proper lsc functional with bounded sublevel sets (i.e. $\{x : F(x) \leqslant \lambda\}$ is bounded for every $\lambda \in \mathbb{R}$), then

$$\min_y \delta F(y) + \frac{1}{2}\|y - x\|^2$$

always has a unique solution and this solution y satisfies

$$0 \in \delta \partial F(y) + y - x$$

We call $(I + \delta\partial F)^{-1} : X \to X$ the proximal mapping of δF. So,

$$y = (I + \delta\partial F)^{-1}(x) = \mathrm{argmin}_y\, \delta F(y) + \frac{1}{2}\|x - y\|^2$$

is a well defined mapping.

Remark 8.

- We have Moreau's identity: $x = (I+\delta\partial F)^{-1}(x) + \delta(I+\delta^{-1}\partial F^*)^{-1}(x/\delta)$. Therefore, if one can easily compute the proximal mapping F^* if we are given that of F.
- By writing $u^{n+1} \in u^n - \delta\partial F(u^{n+1})$, we have that $u^{n+1} = (I + \delta\partial F)^{-1}(u^n)$ and can consider this as an implicit gradient descent step.

Examples

- If K is a closed convex set and $H(p) = 0$ if $p \in K$ and $+\infty$ otherwise, then

  $$(I + \delta\partial H)^{-1}(x) = \mathrm{argmin}_{y\in K}\, \frac{1}{2}\|x - y\|_2^2$$

  is the Euclidean projection of x onto the set K.
- Let $f(x) = |x|$. Then

  $$(I + \delta\partial f)^{-1}(x) = \mathrm{argmin}_y\, \delta\,|y| + \frac{1}{2}\,|x - y|^2 = \begin{cases} x - \delta & \text{if } x \geqslant \delta \\ 0 & \text{if } |x| < \delta \\ x + \delta & \text{if } x \leqslant -\delta. \end{cases}$$

  This is the soft shrinkage operator, and by separability of the ℓ^1 norm,

  $$(I + \delta\partial\|\cdot\|_1)^{-1}(x) = (x - \delta\,\mathrm{sign}(x))\mathbb{1}_{\{|x|\geqslant\delta\}}.$$

5.2.2 Dual formulation

Let u be a minimizer of (4). Denote $F(u) = \|\nabla u\|_{2,1}$ Then, equivalently,

$$\frac{g-u}{\lambda} \in \partial F(u) \iff u \in \partial F^*\left(\frac{g-u}{\lambda}\right).$$

Let $v = \frac{g-u}{\lambda}$. Then

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

which is exactly the equation which characterizes the minimizers of

$$\min_v \frac{1}{2}\|v - \frac{g}{\lambda}\|^2 + \frac{1}{\lambda}F^*(v).$$

To compute F^*, by definition

$$\begin{aligned}
F(u) = \|\nabla u\|_{2,1} &= \sup\left\{\langle \xi,\, \nabla u\rangle\ ;\ \|\xi_{i,j}\| \leqslant 1,\ \forall i,j\right\} \\
&= \sup\left\{-\langle \operatorname{div}\xi,\, u\rangle\ ;\ \|\xi_{i,j}\| \leqslant 1,\ \forall i,j\right\} = \sup_{p\in X}\langle p,\, u\rangle - H(p)
\end{aligned}$$

where $K = \{-\operatorname{div}\xi\ ;\ \|\xi_{i,j}\| \leqslant 1\}$,

$$H(p) = \begin{cases} 0 & p \in K \\ \infty & p \notin K. \end{cases}$$

So, $F = H^*$. Since K is a closed and convex set, H is convex, proper, lowersemicontinuous, and we have that $F^* = H^{**} = H$. Therefore u minimizers (4) if and only if $v = (g-u)/\lambda$ minimizes $\min_{v\in K} \frac{1}{2}\|v - \frac{g}{\lambda}\|^2$. Therefore, by computing a minimizer ξ of

$$\min_{\|\xi_{i,j}\|\leqslant 1} \frac{1}{2}\|\operatorname{div}\xi + \frac{g}{\lambda}\|^2, \tag{6}$$

we recover $u = g + \lambda \operatorname{div}(\xi)$ as the solution to (4).

Lecture 21

The projected gradient descent algorithm We cannot apply the gradient descent algorithm to solve this dual problem – although the objective is differentiable with derivative $\nabla\left(\operatorname{div} p + g/\lambda\right)$ which has Lipschitz constant $L \leqslant 8$, it is minimized over a convex set. Instead, we consider the projected gradient descent algorithm of Beck and Teboulle: Consider

$$\min_{x\in X} F(x) + G(x)$$

where F is $C^{1,1}$ such that ∇F has Lipschitz constant L and $(I + h\partial G)^{-1}$ is simple to compute. The idea is to compute one step of the gradient descent of F in an explicitly, then one step of the gradient descent of G implicitly. The resulting algorithm is:

> Choose $h > 0$ and choose $u^0 \in X$. For $n \geqslant 0$, let
>
> $$u^{n+1} = (I + h\partial G)^{-1}(u^n - h\nabla F(u^n)). \tag{7}$$

Intuition behind algorithm Since ∇F is L-Lipschitz, by the fundamental theorem of calculus,

$$\begin{aligned}
F(y) &= F(x) + \int_0^1 \langle \nabla F(x + t(y-x)),\, y - x\rangle \mathrm{d}t \\
&= F(x) + \langle \nabla F(x),\, y-x\rangle + \int_0^1 \langle \nabla F(x+t(y-x)) - \nabla F(x),\, y-x\rangle \mathrm{d}t \\
&\leqslant F(x) + \langle \nabla F(x),\, y-x\rangle + \frac{L}{2}\|y-x\|^2.
\end{aligned}$$

So, the parabola defined by

$$y \mapsto Q_L(y,x) = F(x) + \langle \nabla F(x),\, y-x\rangle + \frac{L}{2}\|y-x\|^2$$

approximates F from above. Let $x = x^n$ and replace the minimization of F at step n with

$$\min_y Q_L(y, x^n).$$

Then, the minimizer y satisfies $0 = L(y - x^n) + \nabla F(x^n)$, i.e. $y = x^n - \frac{1}{L}\nabla F(x^n)$.

If we now let

$$Q_L(y, x) = F(x) + \langle \nabla F(x),\, y - x\rangle + \frac{L}{2}\|y - x\|^2 + G(y), \tag{8}$$

then

$$F(y) + G(y) \leqslant Q_L(y, x), \qquad \forall x, y \in X.$$

In particular, computing the minimizer of $\min_y Q_L(y, x^n)$ is equivalent to finding y such that

$$0 \in \nabla F(x^n) + L(y - x^n) + \partial G(y) \tag{9}$$

which is one step of the proximal gradient descent algorithm with $h = 1/L$.

Convergence

Lemma 6. *Let $x \in X$, $h > 0$ and let Q_L be defined as in* (8)*. Let $y = \operatorname{argmin} Q_{1/h}(\cdot, x)$ be such that*

$$F(y) + G(y) \leqslant Q_{1/h}(y, x).$$

Then for any $z \in X$,

$$F(z) + G(z) - (F(y) + G(y)) \geqslant \frac{1}{2h}\left(\|y - z\|^2 - \|x - z\|^2\right).$$

Proof. By assumption

$$F(z) + G(z) - (F(y) + G(y)) \geqslant F(z) + G(z) - Q_{1/h}(y, x). \tag{10}$$

Also,

$$F(z) \geqslant F(x) + \langle \nabla F(x),\, z - x\rangle \tag{11}$$

while

$$G(z) \geqslant G(y) + \langle p,\, z - y\rangle, \tag{12}$$

where

$$p = \frac{x - y - h\nabla F(x)}{h} \in \partial G(y).$$

Note that such p exists since y is a minimizer of $Q_{1/h}(\cdot, x)$ (c.f. (9)). Adding up (11) and (12) yields

$$F(z) + G(z) \geqslant F(x) + G(y) + \langle \nabla F(x),\, z - x\rangle + \langle p,\, z - y\rangle.$$

We deduce from (10) that

$$\begin{aligned}
&F(z) + G(z) - (F(y) + G(y)) \geqslant F(x) + G(y) + \langle \nabla F(x),\, z - x\rangle + \langle p,\, z - y\rangle - Q_{1/h}(y, x)\\
&= F(x) + G(y) + \langle \nabla F(x),\, z - x\rangle + \langle p,\, z - y\rangle - F(x) - G(y) - \langle \nabla F(x),\, y - x\rangle - \frac{1}{2h}\|y - x\|^2\\
&= \langle \nabla F(x) + p,\, z - y\rangle - \frac{1}{2h}\|y - x\|^2 = \frac{1}{h}\langle x - y,\, z - y\rangle - \frac{1}{2h}\|y - x\|^2 = \frac{1}{2h}\left(\|y - z\|^2 - \|x - z\|^2\right).
\end{aligned}$$

□

Theorem 12. *Let (u^n) satisfy* (7) *and let $h = 1/L$ where L is the Lipschitz constant of ∇F. Let u^* be a minimizer. Then,*

$$F(u^k) + G(u^k) - (F(u^*) + G(u^*)) \leqslant \frac{L\|u^0 - u^*\|^2}{2k}$$

for any $k \geqslant 1$.

Proof. Using Lemma 6 with $z = u^*$, $x = u^n$ and $h = 1/L$,

$$\frac{2}{L}\left(F(u^*) + G(u^*) - F(u^{n+1}) - G(u^{n+1})\right) \geqslant \|u^{n+1} - u^*\|^2 - \|u^n - u^*\|^2.$$

Summing this from $n = 0, \ldots, k-1$ yields:

$$\frac{2}{L}\left(k(F(u^*) + G(u^*)) - \sum_{n=1}^{k}(F(u^n) + G(u^n))\right) \geqslant \|u^k - u^*\|^2 - \|u^0 - u^*\|^2. \tag{13}$$

By applying Lemma 6 with $z = x = u^n$ and $h = 1/L$:

$$\frac{2}{L}\left(F(u^n) + G(u^n) - F(u^{n+1}) - G(u^{n+1})\right) \geqslant \|u^{n+1} - u^n\|^2.$$

Multiplying this by n and summing from $n = 0, \ldots, k-1$ yields

$$\frac{2}{L}\sum_{n=1}^{k-1} F(u^n) + G(u^n) - \frac{2}{L}(k-1)(F(u^k) + G(u^k)) \geqslant \sum_{n=0}^{k-1} n\|u^{n+1} - u^n\|^2.$$

Adding this inequality to (13) yields

$$\frac{2}{L}\left(k(F(u^*) + G(u^*)) - k(F(u^k) + G(u^k))\right) \geqslant -\|u^0 - u^*\|^2.$$

□

Total variation denoising To apply the projected gradient descent algorithm to solve (6), we let

$$F(p) = \frac{1}{2}\|\operatorname{div} p + g\|^2, \quad \text{and} \quad G(p) = \begin{cases} 0 & \|p_{i,j}\| \leqslant 1 \forall i,j \\ \infty & \text{otherwise.} \end{cases}$$

Note that

$$\nabla F(p) = -\nabla(\operatorname{div} p + g),$$

has Lipschitz constant at most 8. Letting P_C be the projection onto

$$C \stackrel{\text{def.}}{=} \{p \,;\, \|p_{i,j}\| \leqslant 1\}.$$

The projected gradient descent algorithm is performed by projecting back onto the set C after taking a step of the gradient descent of F, giving the following iterations:

$$p^{n+1} = P_C(p^n + h\nabla(\operatorname{div} p^n + g)) = \frac{p^n + h\nabla(\operatorname{div} p^n + g)}{\max\{1, \|p^n + h\nabla(\operatorname{div} p^n + g)\|\}},$$

with step size $h \leqslant 1/8$ (division and norm operation in the denominator is interpreted pointwise).

5.3 The primal dual algorithm

We present the primal dual alorithm of Chambolle and Pock.

The minimization problem (5) can be rewritten as

$$\min_{x \in X} \max_{y \in Y} \langle y, Ax\rangle - F^*(y) + G(x). \tag{14}$$

The idea of the primal dual algorithm is to alternate descent in x and ascent in y.

Under mild conditions ($F : \mathbb{R}^m \to (-\infty, \infty]$, and $G : \mathbb{R}^N \to (-\infty, \infty]$ are proper convex functions such that either $\mathrm{dom}(F) = \mathbb{R}^m$ or $\mathrm{dom}(G) = \mathbb{R}^N$ and there exists $x \in \mathbb{R}^N$ such that $Ax \in \mathrm{dom}(F)$,) one can swap the max and the min in the saddle point problem (14):

$$\begin{aligned}\min_{x\in X} F(Ax) + G(x) &= \min_{x\in X}\max_{y\in Y}\langle y,\, Ax\rangle - F^*(y) + G(x) = \max_{y\in Y} -\max_{x\in X} -\langle y,\, Ax\rangle + F^*(y) - G(x)\\ &= \max_{y\in Y} -G^*(-A^*y) - F^*(y).\end{aligned}$$

The primal dual gap is defined as

$$\mathcal{G}(x,y) = F(Ax) + G(x) + G^*(-A^*y) + F^*(y) \geqslant 0$$

and vanishes if and only if $(\hat{x}, \hat{y})$ solve the saddle point problem (14). In particular, for all $x, y \in X$,

$$\langle y,\, A\hat{x}\rangle - F^*(y) + G(\hat{x}) \leqslant \langle \hat{y},\, A\hat{x}\rangle - F^*(\hat{y}) + G(\hat{x}) \leqslant \langle \hat{y},\, Ax\rangle - F^*(\hat{y}) + G(x).$$

This suggest simultaneously performing approximate gradient descent in x and gradient ascent in y: choose starting points x^0 and y^0, and step sizes $\tau, \sigma > 0$. Define

$$\begin{aligned}y_{n+1} &= (I + \sigma\partial F^*)^{-1}(y_n + \sigma A x_n)\\ x_{n+1} &= (I + \tau\partial G)^{-1}(x_n - \tau A^* y_{n+1}).\end{aligned}$$

This iteration procedure was originally proposed by Zhu and Chan (2008), albeit with no proof of convergence.

A variant of this by Chambolle and Pock (2010) (and is what we shall refer to as the *primal dual algorithm*) is

Choose starting points $x^0 = \overline{x}^0$ and y^0, and step sizes $\tau, \sigma > 0$. Define

$$\begin{aligned}y_{n+1} &= (I + \sigma\partial F^*)^{-1}(y_n + \sigma A \overline{x}_n)\\ x_{n+1} &= (I + \tau\partial G)^{-1}(x_n - \tau A^* y_{n+1})\\ \overline{x}_{n+1} &= 2x^{n+1} - x^n\end{aligned}$$

and the iterates x^n and y^n converge to the solutions of the saddle point problem (14) provided that the step sizes τ and σ satisfy: $\tau\sigma\|A\|^2 < 1$.

Total variation denoising In the case of total variation denoising, we let $F = \|\cdot\|_{2,1}$, $G = \frac{1}{2}\|\cdot - g\|^2$ and $A = \nabla$. We have that

- $(I + \tau\partial G)^{-1}(u) = \frac{u+\tau g}{1+\tau}$.
- $F^* = \iota_K$ where $K = \{p\,;\, \|p_{i,j}\| \leqslant 1, \forall i,j\}$ and $\iota_K(x) = 0$ if $x \in K$ and $+\infty$ otherwise. Therefore, $(I + \sigma\partial F^*)^{-1}$ is the Euclidean projection onto K and $x = (I + \sigma\partial F^*)^{-1}(y)$ for $y \in Y$ is such that

$$x_{i,j} = \frac{y_{i,j}}{\max\{1, \|y_{i,j}\|\}}.$$

Theorem 13. *Let* $F(p) = \sum_{i,j}\sqrt{(p^x_{i,j})^2 + (p^y_{i,j})^2}$. *Then,* $(I + \tau\partial F)^{-1}(u) = ((p^x_{i,j}, p^y_{i,j}))_{i,j}$ *where for* $t \in \{x, y\}$ *and* $u_{i,j} = (u^x_{i,j}, u^y_{i,j})$,

$$p^t_{i,j} = \left(u^t_{i,j} - \tau\frac{u^t_{i,j}}{\|u_{i,j}\|}\right)\mathbb{1}_{\|u_{i,j}\| > \tau}$$

This in particular implies that the proximal mapping for the ℓ^1 norm on complex vectors is

$$(I + \tau\partial\|\cdot\|_1)^{-1}(x) = (x - \tau\,\mathrm{sign}(x))\,\mathbb{1}_{|x|\geqslant\tau}$$

Proof. (i) is straightforward. For (ii), it suffices to consider

$$v \in \mathrm{argmin}_{v\in\mathbb{R}^2} \frac{1}{2}\|\begin{pmatrix}u_1\\u_2\end{pmatrix} - \begin{pmatrix}v_1\\v_2\end{pmatrix}\|^2 + \tau\sqrt{v_1^2+v_2^2}.$$

Suppose v is such that $|v| > 0$. Then, $f(v) = \sqrt{v_1^2+v_2^2}$ is differentiable at v and v satisfies

$$\tau\frac{v_1}{\|v\|} = u_1 - v_1, \qquad \tau\frac{v_2}{\|v\|} = u_2 - v_2. \tag{15}$$

By multiplying the first equality by v_1 and the second by v_2, we see that $v_1u_2 = u_1v_2$. Therefore, v is parallel to u. Substituting this back into (15), we have that

$$v_1 = u_1 - \tau\frac{u_1}{\|u\|}, \qquad v_2 = u_2 - \tau\frac{u_2}{\|u\|}.$$

If $|u| < \tau$, then this implies that the sign of v_1 and u_1 are opposite to each other which contradicts (15). Therefore, $v = 0$ whenever $|u| < \tau$. □

6 Further remarks on TV denoising

6.1 Explicit solutions

Explicit solutions to the ROF model were studied in *"A characterization of convex calibrable sets in $\mathbb{R}^N$"* by Alter, Casselles and Chambolle (2003). In particular, they defined the set of *calibrable sets*:

Definition 9. *$C \subset \mathbb{R}^2$ is a calibrable set if there exists $\alpha > 0$ such that $\alpha\chi_C \in \partial J(\chi_C)$.*

Definition 10. *Let $C \subset \mathbb{R}^N$. The perimeter of the set C is defined to be $Per(\chi_C) := J(\chi_C)$.*

Note that if C is a calibrable set, then necessarily, $\alpha = Per(C)/|C|$ in Definition 9. Indeed, since we must have $\alpha|C| = \int \alpha\chi_C = J(\chi_C) = Per(C)$. We let $h_C := Per(C)/|C|$.

Note that if C is a calibrable set, then $\chi_C - \beta\chi_C \in \lambda\partial J(\chi_C)$, for $\beta = 1 - \lambda h_C$, which implies that $\beta\chi_C$ is the solution to the ROF model with $g = \chi_C$. Hence its edges are preserved.

Proposition 2. *Let $C \subset \mathbb{R}^2$ be a connected bounded set of finite perimeter (i.e. $Per(C) := J(\chi_C) < \infty$). Then, C is calibrable if and only if the following conditions hold:*

1. *C is a convex set;*
2. *∂C is of class $C^{1,1}$;*
3. *$\sup_{p\in\partial C}\kappa_{\partial C}(p) \leqslant P(C)/|C|$, where $\kappa_{\partial C}$ denotes the mean curvature of ∂C.*

Example of a calibrable set The above propositions says that calibrable sets are essentially convex sets with sufficiently bounded curvature. Examples include the disc, or squares with sufficiently rounded corners. In the following, we will explicitly show that the disc is a calibrable set.

To show that C is a calibrable set, it suffices to find $z \in L^\infty(\Omega;\mathbb{R}^2)$ such that $\mathrm{div}\,z(x) = h_C$ for all $x \in C$, and $\mathrm{div}\,z(x) = 0$, for all $x \notin C$.

For example, if $C = B(0;R)$, let

$$z(x) = \begin{cases} \frac{x}{R} & x \in B(0;R) \\ \frac{Rx}{|x|^2} & \text{otherwise.} \end{cases}$$

Then, $\operatorname{div} z = \frac{2}{R}\chi_C$. So, the solution to ROF with $g = \chi_C$ is $(1 - 2\lambda/R)_+\chi_C$.

Lecture 22

6.2 Gradient flow

By letting $u_0 = g$ and defining the sequence, for

$$u_{n+1} = \operatorname{argmin}_{u \in L^2(\Omega)} \tau J(u) + \frac{1}{2}\|u - u_n\|_{L^2}^2, \quad n = 0, \ldots, \lfloor T/\tau \rfloor - 1,$$

i.e.

$$\frac{u_{n+1} - u_n}{\tau} \in \partial J(u_n).$$

One can show that as $\tau \to 0$, the piecewise constant function $u^\tau(t,x) = u_{n+1}$ for $t \in (n\tau, (n+1)\tau)$ converges in $L^2((0,T) \times \Omega)$ to the function $u \in C((0,T); L^2(\Omega)) \cap L^\infty((0,T); BV(\Omega))$ which solves

$$\begin{cases} \partial_t u = \operatorname{div}\left(\frac{Du}{|Du|}\right) & (0,T) \times \Omega \\ \frac{Du}{|Du|} \cdot \nu_{\partial\Omega} = 0 & (0,T) \times \partial\Omega \\ u(0,\cdot) = g & \Omega. \end{cases}$$

This PDE is referred to as the total variation flow and can be seen as an alternative to the gradient flow derived from the gradient descent of the ROF energy.

7 PDE methods

We saw in the previous section that variational methods are often associated with a PDE. Another approach is to directly work with the PDE, without working with any energy. This can lead to different restoration processes which are not necessarily associated with any energy.

7.1 The heat equation

The heat equation is one of the oldest and most investigated equations used in image processing.

$$\begin{cases} \frac{\partial u}{\partial t}(t,x) = \Delta u(t,x) & (x,t) \in \mathbb{R}^2 \times [0,\infty) \\ u(0,x) = u_0(x) & x \in \mathbb{R}^2. \end{cases} \tag{16}$$

Gaussian linear filtering The explicit solution of (16) is

$$u(t,x) = \int_{\mathbb{R}^2} G_{\sqrt{2t}}(x-y)u_0(y)\mathrm{d}y = (G_{\sqrt{2t}} \star u_0)(x),$$

where

$$G_\sigma(x) \stackrel{\text{def.}}{=} \frac{1}{2\pi\sigma^2}\exp\left(-\frac{|x|^2}{2\sigma^2}\right)$$

is a 2D Gaussian kernel. Moreover, by denoting the Fourier transform by $\mathcal{F}f(\omega) = \int_{\mathbb{R}^2} f(x)e^{-i\omega\cdot x}\mathrm{d}x$,

$$\mathcal{F}(G_\sigma)(\omega) = \exp\left(-\frac{|\omega|^2}{2/\sigma^2}\right).$$

So,

$$\mathcal{F}\left(G_{\sqrt{2t}} \star u_0\right)(\omega) = \mathcal{F}(G_{\sqrt{2t}})(\omega)\mathcal{F}(u_0)(\omega) = \exp\left(-t\,|\omega|^2\right)\mathcal{F}(u_0)(\omega).$$

Gaussian filtering corresponds to dampening out the high Fourier frequencies and hence reduces oscillations in space. The problem with this is that the smoothing is isotropic, meaning that it is the same in all directions. Ideally, we don't want to smooth across edges.

7.2 Nonlinear Diffusion – Perona Malik

Let us consider the following equation, initially proposed by Perona and Malik (1990) as a solution to the problems caused by isotropic diffusion:

$$\begin{cases} \frac{\partial u}{\partial t} = \operatorname{div}(c(|\nabla u|^2)\nabla u) & \text{in } \Omega \times (0,T) \\ \frac{\partial u}{\partial N} = 0 & \partial\Omega \times (0,T) \\ u(0,x) = g & \text{in } \Omega, \end{cases} \tag{17}$$

where $c : [0,\infty) \to (0,\infty)$. Note that $c \equiv 1$ corresponds to the heat equation.

The idea of Perona and Malik is that one can change the rate of diffusion depending on spatial location. One would like to encourage smoothing within regions in preference to smoothing across edges. Ideally, one can set $c = 1$ in regions and $c = 0$ at edge points.

Edge enhancement Here, we describe how c can be chosen to not only diffuse, i.e. smooth out unwanted oscillations, but also enhance edges. Let us first consider the 1D case on $\Omega = \mathbb{R}$:

$$\begin{cases} \frac{\partial u}{\partial t}(t,x) = [c(u_x^2(t,x))u_x(t,x)]_x \\ u(0,x) = g. \end{cases} \tag{18}$$

By letting $b(s) = 2sc'(s) + c(s)$, the above equation becomes

$$\begin{cases} \frac{\partial u}{\partial t}(t,x) = b(u_x^2(t,x))u_{xx}(t,x) \\ u(0,x) = g. \end{cases} \tag{19}$$

To enhance edges, we would like u_x to increase over time at edge points, this leads us to consider $\partial_t(u_x)$: From (19),

$$\frac{\partial}{\partial t}(u_x) = \left(\frac{\partial u}{\partial t}\right)_x = u_{xxx}b(u_x^2) + 2u_{xx}^2 b'(u_x^2).$$

If x is an edge at time t, then $u_{xx}(t,x) \approx 0$ and $u_{xxx}(t,x) \ll 0$ (see Figure 1). So,

$$\operatorname{sign}\left(\frac{\partial u}{\partial t}\right)_x = \operatorname{sign}(-b(u_x^2)(t,x)).$$

Therefore, if $b(u_x^2) > 0$, then (18) is a forward parabolic equation and the edge is blurred; and if $b(u_x^2) < 0$, then (18) is a backward parabolic equation and the edge is enhanced.

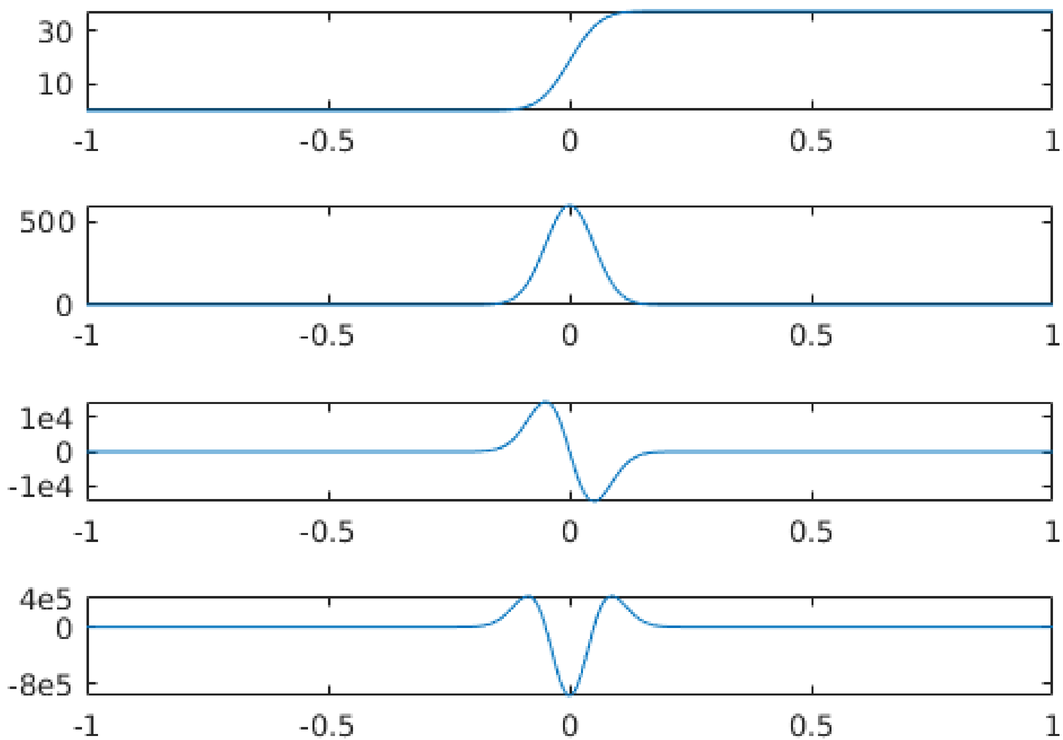

Figure 1: From top to bottom: the mollified edge u, u_x, $u_x x$ and u_{xxx}.

Let us return to the 2D Perona Malik model, recall the formulation in (23). Then, the intuition in 1D suggests that to sharpen edges, we need to impose that

$$b(s) = 2sc'(s) + c(s) < 0, \qquad \forall s \geqslant K, \tag{20}$$

were K is some threshold. If we want to smooth homogeneous regions, then we can impose that $c(0) = b(0) = 1$, which implies that (23) behaves like the heat equation for small gradients. One typical choice for c is

$$c(s) = \frac{1}{1 + s/K}.$$

What can we say about the existence of solutions to (17) under condition (20)? Hardly anything.

To understand this, consider the backward heat equation:

$$\begin{cases} \frac{\partial u}{\partial t}(t,x) = -u_{xx}(t,x) & \text{on}(0,T) \times \mathbb{R} \\ u(0,x) = g & \text{on } \mathbb{R}. \end{cases} \tag{21}$$

Let $\tau = T - t$, so if u solves this backward heat equation, then $v(\tau, x) = u(T - \tau, x)$ solves

$$\begin{cases} \frac{\partial v}{\partial \tau}(\tau,x) = v_{xx}(\tau,x) & \text{on}(0,T) \times \mathbb{R} \\ v(T,x) = g & \text{on } \mathbb{R}. \end{cases}$$

This is the heat equation with backward datum $v(T,\cdot) = g$. According to regularization properties of the heat equation, g must be infinitely differentiable. If not, then (21) cannot have a classical or weak solution.

One way of tackling this ill-posedness is to introduce regularization to ensure the existence of solutions. The idea of Catté el al (1992) is to substitute the gradient in $c(|\nabla u|^2)$ by a smoothed version $G_\sigma \star \nabla u$. Since $G_\sigma \star \nabla u = \nabla G_\sigma \star u$, the proposed regularized scheme is:

$$\begin{cases} \frac{\partial u}{\partial t} = \operatorname{div}(c(|\nabla G_\sigma \star u|^2)\nabla u) & \text{in } \Omega \times (0,T) \\ \frac{\partial u}{\partial N} = 0 & \partial\Omega \times (0,T) \\ u(0,x) = g & \text{in } \Omega, \end{cases} \tag{22}$$

Theorem 14. *Let $c : \mathbb{R}_+ \to \mathbb{R}_+$ is smooth, decreasing with $c(0) = 1$, $\lim_{s\to\infty} c(s) = 0$. If $g \in L^2(\Omega)$, then there exists a unique function $u(t,x) \in C([0,T]; L^2(\mathbb{R})) \cap L^2((0,T); W^{1,2}(\Omega))$ satisfying* (22) *in a distributional sense. Moreover, $\|u\|_{L^\infty((0,T),L^2(\Omega))} \leqslant \|g\|_{L^2}$ and $u \in C^\infty((0,T) \times \overline{\Omega})$.*

Other than well-posedness, this scheme has one other advantage: If the initial data is very noisy such that there are large oscillations in the gradient, the standard Perona-Malik scheme cannot distinguish between 'true edges' and 'false edges'. On the other hand, the smoothing kernel G_σ ensures that the diffusion is less sensitive to noise.

Despite the lack of a rigorous mathematical theory concerning the original Perona-Malik equation, it has been sucessfully used in numerical similations. This is perhaps the case the the discrete problem does not reflect the ill-posedness of the continuous problem, although this point still needs further investigation.

Directional smoothing By formally developping the divergence operator, we see that

$$\operatorname{div}(c(|\nabla u|^2)\nabla u) = \\ 2(u_x^2 u_{xx} + u_y^2 u_{yy} + 2u_x u_y u_{xy})c'(|\nabla u|^2) + c(|\nabla u|^2)(u_{xx} + u_{yy}).$$

For each x where $|\nabla u(x)| \neq 0$, define the vectors

$$N(x) = \frac{\nabla u}{|\nabla u|}$$

and $T(x)$ is such that $\langle T(x), N(x)\rangle = 0$ and $|T(x)| = 1$. Note that $N(x)$ and $T(x)$ are respectively the unit normal and unit tangent vectors to $\partial\{u \geqslant u(x)\}$ at x. Denote the Hessian of u by $\nabla^2 u$ and by $u_{TT} = \langle T, \nabla^2 u T\rangle$ and $u_{NN} = \langle N, \nabla^2 u N\rangle$ are respectively the second derivatives of u in the T-direction and N- direction. Then,

$$u_{NN} = \frac{1}{|\nabla u|^2}(u_x^2 u_{xx} + 2u_x u_y u_{xy} + u_y^2 u_{yy}),$$

and

$$u_{TT} = \Delta u - u_{NN}.$$

Therefore, writing $b(s) = c(s) + 2sc'(s)$, we have that

$$\frac{\partial u}{\partial t}(t,x) = c(|\nabla u|^2)u_{TT} + b(|\nabla u|^2)u_{NN}. \tag{23}$$

So, different choices of c will lead to different rate of diffusion along the T and N directions. Since N is normal to the edges, it would be preferable to diffuse more in the tangential direction T. So, we impose

$$\lim_{s\to\infty} \frac{b(s)}{c(s)} = 0 \implies \lim_{s\to\infty} \frac{sc'(s)}{c(s)} = -\frac{1}{2}.$$

Assuming that $c(s) = s^\beta$ for some β, this limit suggests that one should take $c(s) \approx s^{-1/2}$ as $s \to \infty$. One choice is $c(s) = (s+1)^{-1/2}$. Note that $c(s) = s^{-1/2}$ actually recovers the gradient flow of the total variation functional.

7.3 Anisotropic diffusion

Rather than simply considering the magnitude of the gradient, Weickert developped an anisotropic diffusion model which takes into account local variations of the gradient orientation.

A natural idea is to say that the preferred smoothing direction is the one that minimizes gray-value fluctuations. For $\theta \in [0, 2\pi)$, let $d(\theta) = (\cos(\theta), \sin(\theta))$. Now, $F(\theta) = \langle d(\theta), \nabla u(x)\rangle^2$ is maximal when d is parallel to ∇u (i.e. parallel to the normal direction at x), and minimal if d is orthogonal to ∇u. Maximizing/minimizing $F(\theta)$ is equivalent to maximizing/minimizing the quadratic form $d^T \nabla u \nabla u^T d$ where

$$\nabla u \nabla u^T = \begin{pmatrix} u_x^2 & u_x u_y \\ u_x u_y & u_y^2 \end{pmatrix}.$$

This is a positive semidefinite matrix with eigenvalues $\lambda_1 = |\nabla u|^2$ and $\lambda_2 = 0$. There exists orthonormal eigenvectors v_1 parallel to ∇u and v_2 orthogonal to $\nabla u^\perp$ since

$$\nabla u \nabla u^T \nabla u = |\nabla u|^2 \nabla u, \qquad \nabla u \nabla u^T \begin{pmatrix} u_y \\ -u_x \end{pmatrix} = 0.$$

It is tempting to define at x an orientation descriptor as a function of $\nabla u \nabla u^T$. The problem is that this does not take into account information in a neighborhood of x, making the descriptor sensitive to spurious variations in the image. To overcome this, Weickert introduced the use of smoothing kernels at different scales.

1. To avoid false detections due to noise, u is first convolved with a Gaussian kernel, $u_\sigma = (G_\sigma \star u)(x)$ with $\sigma > 0$.
2. Local orientation is averaged by convolving $\nabla u_\sigma \nabla u_\sigma^T$ by G_ρ, this yields the symmetric, positive semidefinite matrix (so called structure tensor)

$$J_\rho(\nabla u_\sigma) = G_\rho \star \nabla u_\sigma \nabla u_\sigma^T.$$

3. The matrix $J_\rho(\nabla u_\sigma) = (j_{ij})_{i,j=1}^2$ has orthonormal eigenvectors v_1, v_2 with v_1 parallel to

$$\begin{pmatrix} 2j_{12} \\ j_{22} - j_{11} + \sqrt{(j_{22} - j_{11})^2 - 4j_{12}^2} \end{pmatrix}.$$

The corresponding eigenvalues are

$$\mu_1 = \frac{1}{2}\left(j_{22} + j_{11} + \sqrt{(j_{22} - j_{11})^2 - 4j_{12}^2}\right)$$

and

$$\mu_2 = \frac{1}{2}\left(j_{22} + j_{11} - \sqrt{(j_{22} - j_{11})^2 - 4j_{12}^2}\right).$$

Note that $J_\rho(\nabla u_\sigma) = \mu_1 v_1 v_1^T + \mu_2 v_2 v_2^T$. Then, μ_1 and μ_2 describe the average constrast of the smoothed image function within a neighborhood of size $O(\rho)$ in each eigendirection. v_1 indicates the direction of maximal gray-value fluctuations, v_2 indicates the preferred local direction of smoothing. The eigenvalues μ_1 and μ_2 also convey the shape information in the form

$$\begin{array}{rl} \mu_1(x) \approx \mu_2(x) & \text{image has isotropic structure at } x \\ \mu_1(x) >> \mu_2(x) \approx 0 & \text{image has line-like structure at } x \\ \mu_1(x) \geqslant \mu_2(x) >> 0 & \text{image edge forms a corner at } x. \end{array}$$

The nonlinear diffusion process is now governed by a parabolic equation which can be viewed as an extension of (23):

$$\begin{cases} \frac{\partial u}{\partial t} = \operatorname{div}(D(J_\rho(\nabla u_\sigma))\nabla u) & \Omega \times (0,T) \\ u(0,x) = g(x) & \Omega \\ \langle D(J_\rho(\nabla u_\sigma))\nabla u,\, N\rangle = 0 & \partial\Omega \times (0,T), \end{cases} \tag{24}$$

where S^2 is the set of symmetric matrices, and $D \in C^\infty(S^2, S^2)$ is called the diffusion tensor which is chosen depending on the imaging task at hand. Roughly speaking, the eigenvectors of D should reflect the local image structure. Hence, a good choice is to choose them to be the orthonormal basis of eigenvectors of J_ρ. The choice of the eigenvalues depend on the desired goal.

Coherence-enhancing anisotropic diffusion Recall that $(\mu_1 - \mu_2)^2$ provides an indication of line-like structure at x. Thus, if one wants to enhance flow-like structures and close interrupted lines, one should smooth along the direction v_2 with a diffusivity which increases wrt the *coherence* $(\mu_1 - \mu_2)^2$. So, one possible choice $D(J_\rho) = \lambda_1 v_1 v_1^T + \lambda_2 v_2 v_2^T$ where the eigenvalues are

$$\begin{aligned} \lambda_1 &= \alpha, \\ \lambda_2 &= \begin{cases} \alpha & \mu_1 = \mu_2 \\ \alpha + (1-\alpha)\exp\left(\frac{-1}{(\mu_1-\mu_2)^2}\right) & \mu_1 \neq \mu_2. \end{cases} \end{aligned}$$

where the small positive parameter $\alpha \in (0,1)$ keeps the diffusion tensor uniformly positive definite.

8 Inpainting

An important task in image processing is the process of filling in missing parts of damaged images based on the information obtained from surrounding areas. It is essentially a type of interpolation and is called inpainting.

Let g represent some given image defined on an image domain Ω. Loosely speaking, the problem is to reconstruct the original image u in the (damaged) domain $D \subset \Omega$, called the inpainting domain.

Total variation inpainting was proposed by Chan and Shen where the inpainted image u is recovered as a minimizer of

$$J(u) + \frac{\lambda}{2}\|\chi_{\Omega\setminus D}(u-g)\|_{L^2}^2.$$

Note that this variational problem acts on the whole image domain Ω, instead of posing the problem on the missing domain D only. This has the advantage of simultaneous noise removal in the whole image and makes the approach independent of the number and shape of the holes in the image.

In the noise free case, that is if we assume that $g_{\Omega\setminus D}$ is completely intact, we can also formulate the following variational approach: assume that $g \in BV(\Omega)$ and see the inpainted image u that solves

$$\min\left\{J(u)\,;\; u_{\Omega\setminus D} = g_{\Omega\setminus D}\right\}. \tag{25}$$

Theorem 15. *For an original image $g \in BV(\Omega)$ and an inpainting domain $\mathcal{D}$ with Lipschitz boundary, the minimization problem (25) has a minimizer $u^* \in BV(\Omega)$.*

Proof. We can rewrite (25) as

$$\min_v J(v) + \iota_{\{u\in L^2(\Omega):\, u_{\Omega\setminus D} = g_{\Omega\setminus D}\}}(v),$$

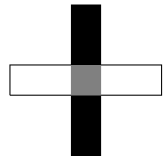

Figure 2: What colour should the gray region be?

where $\iota_S(u) = \begin{cases} 0 & u \in S \\ +\infty & \text{otherwise.} \end{cases}$ Then, we can apply the direct method of calculus of variations, noting that the indicator function is lsc and using the compactness properties of L^2. $\square$

Remark 9. Note that there is no uniqueness. However, rather than considering this an artefact of the inpainting model, one can consider this a reflection of uncertainty in human perception (see Figure 8).

8.1 The need to consider higher order methods

As we shall see shortly, minimizing the total variation of a function is equivalent to minimizing the length of its level lines. Thus, the total variation inpainting model will produce reconstructions for which the level lines from the boundary of the inpainting region will simply be connected by straight lines. This can often lead to visually unnatural reconstruction. In this section, we present a higher order method to alleviate this problem.

To begin with, we present an alternative characterization of the total variation of a function:

Definition 11. *Let $E \subset \Omega$ be a measurable set in $\mathbb{R}^2$. This set is called a set of finite perimeter if its characteristic function $\chi_E \in BV(\Omega)$. We write $Per(E;\Omega) := |D\chi_E|(\Omega)$, for the perimeter of E in Ω.*

With the notion of sets of finite perimeter we have the following theorem.

Theorem 16 (Coarea formula)**.** *Let $u \in BV(\Omega)$ and for $s \in \mathbb{R}$, the set $F_s = \{u > s\}$ is the s-level set of u. Then,*

$$|Du|(\Omega) = \int_{-\infty}^{\infty} \mathrm{Per}(F_s;\Omega)\mathrm{d}s$$

Proof. We simply outline the key steps of this proof. The first step is to establish this formula for $u \in C^\infty$. This can done rigorously by approximating smooth functions by piecewise affine functions – see Theorem 1.23 of 'Minimal surfaces and Functions of Bounded Variation' (Giusti, 1984). However, we shall simply present a simple argument where the boundaries of the level sets are differentiable.

Step 1: Assume u is continuously differentiable and that ∂F_λ is a differentiable curve $\gamma(\lambda, s)$ which is parametrized by its arc-length s. Let $T(\gamma)$ be the vector tangent to this curve and note that $\nabla u(\gamma)$ is orthogonal to $T(\gamma)$ and parallel to $N(\gamma)$, the unit normal to the curve.

Let ds and dn be the Lebesgue measures in the direction of T and N. Then,

$$|\nabla u(\gamma)| = \nabla u(\gamma) \cdot N(\gamma) = \frac{\mathrm{d}\lambda}{\mathrm{d}n}. \tag{26}$$

So,

$$J(u) = \int_\Omega |\nabla u| = \int \int_{\partial F_\lambda} |\nabla u(\gamma(\lambda, s))| \,\mathrm{d}s\mathrm{d}n.$$

Using (26), we have that

$$J(u) = \int \int_{\partial F_\lambda} \mathrm{d}s\mathrm{d}\lambda = \int \mathrm{Per}(F_\lambda)\mathrm{d}\lambda.$$

Step 2: Approximate $u \in BV(\Omega)$ by smooth functions u_n such that $\|u_n - u\|_{L^1} \to 0$ and $J(u_n) \to J(u)$. Then,

$$J(u) = \lim_{n\to\infty} J(u_n) = \lim_{n\to\infty} \int_{-\infty}^{\infty} \mathrm{Per}(\{u_n > s\})\mathrm{d}s \geqslant \int_{-\infty}^{\infty} \lim_{n\to\infty} \mathrm{Per}(\{u_n > s\})\mathrm{d}s \geqslant \int_{-\infty}^{\infty} \mathrm{Per}(\{u > s\})\mathrm{d}s.$$

Here, the first inequality follows by Fatou's lemma, and the second inequality follows by lower semicontinuity of the total variation functional (note that by letting $F_s = \{u > s\}$ and $F_s^n = \{u_n > s\}$, we have that

$$\int |u_n - u| = \int_{\mathbb{R}} |F_{n,s} \Delta F_s| \,\mathrm{d}s, \qquad F_{n,s}\Delta F_s = (F_{n,s} \setminus F_s) \cup (F_s \setminus F_{n,s}).$$

Therefore, $u_n \to u$ in L^1 implies that up to a subsequence, $\chi_{\{u_{n_k} > s\}} \to \chi_{\{u>s\}}$ in L^1 for a.e. $s \in \mathbb{R}$).

Step 3: Deriving the upper bound.

Let $u \in BV(\Omega)$ and let $\varphi \in C_c^\infty(\Omega; \mathbb{R}^2)$ be such that $\|\varphi\|_{L^\infty} \leqslant 1$.

If $u \geqslant 0$ a.e., then $u(x) = \int_0^{u(x)} \mathrm{d}s = \int_0^\infty \chi_{u>s}(x)\mathrm{d}s$. So, by Fubini,

$$\int_\Omega u(x) \operatorname{div} \varphi(x)\mathrm{d}x = \int_0^\infty \int_\Omega \chi_{u>s}(x) \operatorname{div} \varphi(x)\mathrm{d}x\mathrm{d}s. \tag{27}$$

On the other hand, if $u \leqslant 0$ a.e., then $u(x) = -\int_{u(x)}^0 \mathrm{d}s = \int_{-\infty}^0 -\chi_{u<s}(x)\mathrm{d}s = \int_{-\infty}^0 (\chi_{u>s}(x) - 1)\mathrm{d}s$. Note that $\int_\Omega \operatorname{div} \varphi(x) = \int_{\partial\Omega} \varphi \cdot \nu_{\partial\Omega} = 0$ since $\varphi \in C_c^\infty(\Omega)$. Therefore, by this observation and by Fubini,

$$\int_\Omega u(x) \operatorname{div} \varphi(x)\mathrm{d}x = \int_{-\infty}^0 \int_\Omega \chi_{u>s}(x) \operatorname{div} \varphi(x)\mathrm{d}x\mathrm{d}s. \tag{28}$$

So, given u, we can split u into its positive and negative parts $u_+ = u\chi_{u\geqslant 0}$, $u_- = u\chi_{u<0}$, and combining (27) and (28), we have that

$$\int_\Omega u(x) \operatorname{div} \varphi(x)\mathrm{d}x = \int_{-\infty}^\infty \int_\Omega \chi_{u>s} \operatorname{div} \varphi(x)\mathrm{d}x\mathrm{d}s.$$

Since $\int_\Omega \chi_{u>s} \operatorname{div} \varphi(x)\mathrm{d}x \leqslant \mathrm{Per}(\{u > s\})$, it follows that

$$\int_\Omega u \operatorname{div} \varphi \leqslant \int_{\mathbb{R}} \mathrm{Per}(\{u > s\})\mathrm{d}s.$$

Taking the supremum over all $\varphi \in C_c^\infty(\Omega; \mathbb{R}^2)$ with $\|\varphi\|_{L^\infty} \leqslant 1$ yields the required upper bound.

□

8.2 Interlude: curvature

Let $x(p) = (x_1(p), x_2(p))$ be a continuously differentiable curve in $\mathbb{R}^2$ with $p \in [0, 1]$. Then

- $T(p) = x'(p) = (x_1'(p), x_2'(p))$ is the tangent vector at $x(p)$,
- $N(p) = (-x_2'(p), x_1'(p))$ is the normal vector at $x(p)$,
- $s(p) = \int_0^p \sqrt{(x_1'(r)^2 + x_2'(r)^2} \mathrm{d}r$ is the arc length.

If we parametrize the curve x by the arc length s instead, then,

$$T(s) = \frac{\mathrm{d}x}{\mathrm{d}s}(s)$$

is such that $|T(s)| = 1$, since $\frac{\mathrm{d}p}{\mathrm{d}s} = |x'(p)|^{-1}$ and $\frac{\mathrm{d}x}{\mathrm{d}s} = x'(p)\frac{\mathrm{d}p}{\mathrm{d}s}$ (this yields a standardized parametrization). The *curvature tensor* is

$$\frac{\mathrm{d}T}{\mathrm{d}s}(s) = \frac{\mathrm{d}^2 x}{\mathrm{d}s^2}(s)$$

and is parallel to $N(s)/|N(s)|$, i.e.

$$\frac{\mathrm{d}T}{\mathrm{d}s}(s) = \kappa(s)\frac{N(s)}{|N(s)|}$$

where $\kappa(s)$ is the *curvature.*

Note that for any parametrization p,

$$\kappa(p) = \frac{x_1'(p)x_2''(p) - x_2'(p)x_1''(p)}{|x'(p)|^3}.$$

Curves as isolevels of a function Let $k \in \mathbb{R}$, $u : \mathbb{R}^2 \to \mathbb{R}$ and let

$$x(s) = \{(x_1(s), x_2(s))\ ;\ u(x(s)) = k\}$$

be parametrized by its arc-length. By differentiating $u(x(s)) = k$ with respect to s,

$$\langle x'(s),\, \nabla u(x(s))\rangle = 0. \tag{29}$$

So, $x'(s)$ is parallel to $(-u_{x_2}(x(s)), u_{x_1}(x(s)))$ and there exists λ such that

$$x_1'(s) = -\lambda u_{x_2}, \quad x_2'(s) = \lambda u_{x_1}.$$

Differentiating (29), we get

$$(x_1'(s))^2 u_{x_1x_1} + (x_2'(s))^2 u_{x_2x_2} + 2x_1'(s)x_2'(s)u_{x_1x_2} + x_1''(s)u_{x_1} + x_2''(s)u_{x_2} = 0,$$

and

$$(\lambda u_{x_2})^2 u_{x_1x_1} + (\lambda u_{x_1})^2 u_{x_2x_2} - 2\lambda^2 u_{x_1}u_{x_2}u_{x_1x_2} + \frac{1}{\lambda}(x_1''(s)x_2'(s) - x_2''(s)x_1'(s)) = 0$$

Since $|x'(s)| = 1$, we get that $\lambda^2 = |\nabla u|^{-2}$, and one can rearrange to observe that

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

8.3 The Euler's Elastica inpainting model

In the TV inpainting model, if g is smooth and $\Gamma_s = \{g = s\}$ intersects ∂D at p_1 and p_2, then due to the coarea formula, we are essentially searching for

$$\min_{\gamma_s \in \mathcal{A}} \mathcal{H}_1(\gamma_s \cap D)$$

over curves γ_s for which $\gamma_s(p_1) = \Gamma_s(p_1)$ and $\gamma_s(p_2) = \Gamma_s(p_2)$. This encourages straight line interpolations between $\Gamma_s(p_1)$ and $\Gamma_s(p_2)$ and potentially leading to sharp corners. An alternative approach to alleviate these problems would be to consider the curvature of the level lines.

In 1998, Masnous and Morel introduced Euler's elastica energy (which originates from Euler in 1744) for the interpolation of level lines: To connect $\Gamma_s(p_1)$ with $\Gamma_s(p_2)$ with $p_1, p_1 \in \partial D$, find a curve which minimizes

$$\min_{\gamma_s \in \mathcal{A}} \int_{\gamma_s} (\alpha + \beta\kappa^2) \mathrm{d}t,$$

where κ is the curvature of γ_s, α, β are positive weighting parameters and

$$\mathcal{A} = \{\gamma_s \; ; \; \gamma_s(p_i) = \Gamma_s(p_i), \quad i = 1, 2\}.$$

For all grey values $s \in [0, 1]$, one aims at minimizing

$$\mathcal{E}(\mathcal{F}) = \int_0^1 \int_{\gamma_s} (\alpha + \beta\kappa^2) \mathrm{d}t\mathrm{d}s,$$

over $\mathcal{F} = \{\gamma_s \; ; \; s \in [0, 1], \quad \text{with appropriate boundary conditions}\}$. The problem with this formulation is that the curves may intersect (and hence would not define a function), it is also difficult numerically to deal directly with this formulation.

Functionalized formulation In 2002, Chan, Kang and Shen (*Euler's Elastica and curvature-based inpainting, SIAM J. Appl. Math.*) introduced a functionalized formlation of the Euler's elastica model for inpainting: Suppose for now that u is smooth so that its curvature is well defined. Let $\gamma_s = \{u = s\}$. Recall that $\kappa = \operatorname{div}(\nabla u / |\nabla u|)$. Furthermore, if $\mathrm{d}n$ is the change in the normal direction (where u has maximal ascent) and s denotes the intensity of u, then

$$\frac{\mathrm{d}s}{\mathrm{d}n} = |\nabla u|, \quad \text{i.e. } \mathrm{d}s = |\nabla u| \, \mathrm{d}n.$$

Plugging in these observations yields,

$$\begin{aligned} \int_0^1 \int_{\gamma_s} (\alpha + \beta\kappa^2) \mathrm{d}t\mathrm{d}s &= \int_0^1 \int_{\gamma_s} \left(\alpha + \beta \left(\operatorname{div} \left(\frac{\nabla u}{|\nabla u|} \right) \right)^2 \right) |\nabla u| \, \mathrm{d}t\mathrm{d}n \\ &= \int_D \left(\alpha + \beta \left(\operatorname{div} \left(\frac{\nabla u}{|\nabla u|} \right) \right)^2 \right) |\nabla u| \, \mathrm{d}x. \end{aligned}$$

The inpainting model is now: Minimize

$$J_2(u) = \int_{\mathcal{D}} \left(\alpha + \beta \left(\operatorname{div} \left(\frac{Du}{|Du|} \right) \right)^2 \right) |Du| \, \mathrm{d}x$$

with conditions

$$u\chi_{\Omega \setminus D} = g\chi_{\Omega \setminus \mathcal{D}}, \quad \int_{\partial\mathcal{D}} |Du| = 0, \quad |\kappa(p)| < \infty \quad \textit{a.e.} \quad \partial\mathcal{D}.$$

Remark 10. Note that we implicitly assume that the input data $g \in BV(\Omega)$ is such that $\int_{\partial\mathcal{D}} |Dg| = 0$. The trace of BV functions is well defined and letting g^+ and g^- denote the exterior and interior trace of g along $\partial\mathcal{D}$,

$$\int_{\partial\mathcal{D}} |Dg| = \int_{\partial\mathcal{D}} \left| g^+ - g^- \right| \mathrm{d}\mathcal{H}^1$$

which means that $g^+ = g^-$ a.e. (w.r.t. to $\mathcal{H}^1$) along $\partial\mathcal{D}$. In other words, we assume that g does not jump across the boundary of $\mathcal{D}$, i.e. there is no essential overlap between the boundary of the missing domain $\mathcal{D}$ and image edges. Note that this is a natural restriction, since this is the assumption that entire image features are not missing: imagine an image of a face with clearly outlined lips. If the inpainting domain is created by cutting along the outer edges of the upper and lower lips, then an inpainting scheme will, in the absence of additional information, inpaint this region in accordance to the surrounding pixel values. Thus creating a face without a mouth.

Thus the second constraint is natural since a low-level inpainting scheme is not expect to create new objects, but to complete objects based on information outside the inpainting domain.

Curvature for BV functions? Suppose that $u \in BV(\mathcal{D})$, then $|Du|$ is a Radon measure on $\mathcal{D}$. Let $\mathrm{Supp}(|Du|)$ denote the support of this measure. Then, for any $p \in \mathrm{Supp}(|Du|))$, on any of its small neighbourhoods N_p,

$$|Du|\,(N_p) = \int_{N_p} |Du| > 0.$$

Let ρ be a fixed radially symmetric non-negative mollifier with compact support and unit total integral. Let

$$\rho_\sigma = \frac{1}{\sigma^2}\rho\left(\frac{x}{\sigma}\right), \quad \text{and} \quad u_\sigma = \rho_\sigma \star u.$$

Recall the fact that that $f \in C_c^\infty$ and $g \in L^1_{loc}$ implies that $f \star g \in C^\infty$. So:

Definition 12. *We define the weak absolute curvature $\tilde{\kappa}(p)$ of u at p by*

$$\tilde{\kappa}(p) = \limsup_{\sigma \to 0} \left| \mathrm{div}\left(\frac{\nabla u_\sigma}{|\nabla u_\sigma|}\right)(p)\right|,$$

where for those σ for which $|\nabla u_\sigma| = 0$, we define $\mathrm{div}\left(\frac{\nabla u_\sigma}{|\nabla u_\sigma|}\right) = \infty$. *Finally, outside* $\mathrm{Supp}(|Du|)$*, since u is constant a.e., we assign 0 to $\tilde{\kappa}(p)$.*

The generalized Euler's elastics inpainting model of Chan, Kang and Shen is therefore: Minimize over $u \in BV(\Omega)$

$$J_2(u) = \int_{\mathcal{D}} \left(\alpha + \beta\tilde{\kappa}^2\right)|Du|\,\mathrm{d}x \tag{30}$$

with conditions

$$u\chi_{\Omega\setminus D} = g\chi_{\Omega\setminus\mathcal{D}}, \quad \int_{\partial\mathcal{D}} |Du| = 0, \quad |\tilde{\kappa}(p)| < \infty \quad a.e. \quad \partial\mathcal{D}.$$

In the presence of noise, one may solve:

$$J_2^\lambda(u) = \int_\Omega \left(\alpha + \beta\tilde{\kappa}^2\right)|Du|\,\mathrm{d}x + \frac{\lambda}{2}\|\chi_{\Omega\setminus\mathcal{D}}(u-g)\|^2_{L^2}. \tag{31}$$

8.3.1 Euler Lagrange equation

Note that the Euler Elastica energy is nonconvex (unless $\beta = 0$). Moreover, due to the presence of the curvature (which has no linear structure, e.g. one cannot say much about κ_{u+v} given κ_u and κ_v), the direct method is difficult to emply and it is not clear how to prove establish existence of minimizers to the Euler Elastica variational problem. In such situations, one common approach is to formally derive the Euler-Lagrange equation. The resulting PDE formulation often allows us to handle geometry more explicitly than the variation formulation.

We now provide a formal derivation of the optimality conditions associated with the generalized Euler's elastica inpainting model.

In the following, we assume that we are dealing with sufficiently smooth functions, say in $W^{2,1}$ and the curvature is well-defined.

Theorem 17. *Let $\varphi \in C^1(\mathbb{R}, (0,\infty))$ be a given function and*

$$R(u) = \int_\Omega \varphi(\kappa)\, |\nabla u|$$

for $u \in W^{2,1}(\Omega)$. Then, its first variation over $C_c^\infty(\Omega)$ is given by

$$\nabla_u R = -\operatorname{div} V$$

where

$$V = \varphi(\kappa)\cdot N - \frac{T}{|\nabla u|}\frac{\partial(\varphi'(\kappa)\,|\nabla u|)}{\partial T}.$$

Here, $N = \nabla u/|\nabla u|$ is the unit normal vector, T is the unit tangent vector ($\langle T, N\rangle = 0$). By first variation, we mean that

$$\int_\Omega \nabla_u R \cdot v = \left(\frac{\mathrm{d}}{\mathrm{d}\tau} R(u+\tau v)\right)\bigg|_{\tau=0}, \qquad \forall v \in C_c^\infty.$$

Proof. We first compute

$$\frac{\mathrm{d}}{\mathrm{d}\tau} R(u+\tau v) = \int_\Omega \varphi(\kappa_{u+\tau v})\frac{\nabla(u+\tau v)}{|\nabla(u+\tau v)|}\cdot\nabla v + |\nabla(u+\tau v)|\,\varphi'(\kappa_{u+\tau v})\frac{\mathrm{d}}{\mathrm{d}\tau}\kappa_{u+\tau v}. \tag{32}$$

To deal with the second term, recall that $\kappa_{u+\tau v} = \operatorname{div}(\nabla(u+\tau v)/|\nabla(u+\tau v)|)$. So,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}\tau}\kappa_{u+\tau v} &= \operatorname{div}\left(\frac{\mathrm{d}}{\mathrm{d}\tau}\frac{\nabla(u+\tau v)}{|\nabla(u+\tau v)|}\right)\\
&= \operatorname{div}\left(\frac{\nabla v}{|\nabla(u+\tau v)|} - \frac{\nabla(u+\tau v)(\nabla(u+\tau v)\cdot\nabla v)}{|\nabla(u+\tau v)|^3}\right)\\
&= \operatorname{div}\left(\frac{1}{|\nabla(u+\tau v)|}\left[\mathrm{Id} - \frac{\nabla(u+\tau v)}{|\nabla(u+\tau v)|}\otimes\frac{\nabla(u+\tau v)}{|\nabla(u+\tau v)|}\right]\nabla v\right)
\end{aligned}$$

Plugging this computation back into (32), we obtain

$$\frac{\mathrm{d}}{\mathrm{d}\tau} R(u+\tau v)\bigg|_{\tau=0} = \int_\Omega \varphi(\kappa_u) N\cdot\nabla v + |\nabla u|\,\varphi'(\kappa_u)\operatorname{div}\left(\frac{1}{|\nabla u|}[\mathrm{Id} - N\otimes N]\nabla v\right), \tag{33}$$

where we have written $N = \nabla u/|\nabla u|$. By integration by parts, the second term on the right can be simplified to

$$\begin{aligned}
-\int_\Omega \nabla(|\nabla u|\,\varphi'(\kappa_u))\cdot\left(\frac{1}{|\nabla u|}[\mathrm{Id}-N\otimes N]\nabla v\right) &= -\int_\Omega \nabla v\cdot\left(\frac{1}{|\nabla u|}[\mathrm{Id}-N\otimes N]\nabla(|\nabla u|\,\varphi'(\kappa_u))\right)\\
&= \int_\Omega v\cdot\operatorname{div}\left(\frac{1}{|\nabla u|}[\mathrm{Id}-N\otimes N]\nabla(|\nabla u|\,\varphi'(\kappa_u))\right).
\end{aligned}$$

For the second equality above, we have used the fact that $[\mathrm{Id}-N\otimes N]$ is symmetric. Moreover, by integration by parts, the first term of (33) can be simplified as

$$-\int_\Omega v\operatorname{div}(\varphi(\kappa_u)N).$$

So, (33) can be rewritten as

$$\frac{\mathrm{d}}{\mathrm{d}\tau} R(u+\tau v)\bigg|_{\tau=0} = -\int_\Omega v\operatorname{div}\left(\varphi(\kappa_u)N - \frac{1}{|\nabla u|}[\mathrm{Id}-N\otimes N]\nabla(|\nabla u|\,\varphi'(\kappa_u))\right),$$

Note that $(N \otimes N)(w) = \langle N,\, w \rangle N$ is the orthogonal projection of w into the normal direction, and that $I = N \otimes N + T \otimes T$. It thus follows that

$$\nabla_u R = -\operatorname{div}\left(\varphi(\kappa_u) N - \frac{T}{|\nabla u|} \langle \nabla \left(|\nabla u|\, \varphi'(\kappa_u) \right),\, T \rangle \right) = -\operatorname{div}\left(\varphi(\kappa_u) N - \frac{T}{|\nabla u|} \frac{\partial \left(|\nabla u|\, \varphi'(\kappa_u) \right)}{\partial T} \right).$$

□

Corollary 1. *For the elastica painting model* (31)*, the first variation is*

$$\nabla_u J_2^\lambda(u) = -\operatorname{div} V + \lambda(u-g)\chi_{\Omega \setminus \mathcal{D}}$$

where

$$V = (\alpha + \beta \kappa^2) N - \frac{2\beta}{|\nabla u|} \frac{\partial (\kappa\, |\nabla u|)}{\partial T} T,$$

and the associated steepest descent PDE is

$$\frac{\partial u}{\partial t} = \operatorname{div} V - \lambda(u-g)\chi_{\Omega \setminus \mathcal{D}},$$

on Ω.

8.4 An interpretation in terms of transport and diffusion

Chan, Kang and Shen offered an interpretation of their derived Euler elastica PDE in terms of transportation along isophotes and diffusion across isophotes.

We first recall two PDE-based inpainting schemes proposed prior to the functionalized Euler Elastica formulation of Chan, Kang and Shen.

Curvature Driven Diffusion (CDD) In order to circumvent the inability of total variation inpainting to continue level lines across large disconnected regions, Chan and Shen introduced the CDD inpainting model:

$$\frac{\partial u}{\partial t} = \operatorname{div}\left(\frac{c(|\kappa|)}{|\nabla u|} \nabla u \right),$$

where $c : [0, \infty) \to [0, \infty)$ is a continuous and increasing function satisfying $c(0) = 0$ and $c(\pm\infty) = +\infty$. If $c \equiv 1$, then this is simply the total variation flow. Note also the resemblence of this to the Perona Malik model. The function c essentially controls the amount of diffusion in the normal direction. By the prescribed properties of c, diffusion will be stronger where a level curve of u has larger curvature, and the amount of diffusion goes to 0 as the level curves stretch out. This discourages the formation of corner which is characteristic of TV inpainted images.

Bertalmio's Transport Inpainting In contrast to the diffusion models that we have seen so far, the PDE inpainting model of Bertalmio is based on the 'transportation of smoothness' along level curves:

$$\frac{\partial u}{\partial t} = \nabla^\perp u \cdot \nabla L(u)$$

where $\nabla^\perp u = (-u_y, u_x) = |\nabla u|\, T$ where T is the unit tangent vector to the level curve of u, and $L(u)$ is any smoothness measure of the u. In the original work of Beltamio, $L(u) = \Delta u$. As the evolution approaches its equilibrium state,

$$T \cdot \nabla L(u) = 0, \qquad \text{i.e. } \frac{\partial L(u)}{\partial T} = 0,$$

this means that along level lines, the smoothness measure is constant. Thus, given boundary data, boundary smoothness is transported along level lines into the missing domain.

Back to the Euler's Elastical model... We saw that CDD diffuses across level lines, whereas the Beltamio model transports along level lines. The Euler's elastica model can be seen as an interpolation of these two approaches. We saw that the PDE associated with Euler's elastica model is written as

$$\frac{\partial u}{\partial t} = \operatorname{div} V$$

where V consists of two components: the normal part

$$V_N = \varphi(\kappa) \cdot N,$$

and the tangent part

$$V_T = -\frac{1}{|\nabla u|} \frac{\partial(\varphi'(\kappa)\,|\nabla u|)}{\partial T} T.$$

Now, the normal part V_N corresponds to the CDD scheme with

$$g(\kappa) = \varphi(\kappa).$$

Moreover, the tangent part V_T can be written as

$$V_T = -\left(\frac{1}{|\nabla u|^2} \frac{\partial(\varphi'(\kappa)\,|\nabla u|)}{\partial T}\right) \nabla^{\perp} u.$$

and since $\operatorname{div}(\nabla^{\perp} u) = 0$,

$$\operatorname{div}(V_T) = \nabla \left(\frac{-1}{|\nabla u|^2} \frac{\partial(\varphi'(\kappa)\,|\nabla u|)}{\partial T}\right) \cdot \nabla^{\perp} u.$$

This therefore corresponds to Bertalmio's scheme with smoothness measure

$$L(u) = \frac{-1}{|\nabla u|^2} \frac{\partial(\varphi'(\kappa)\,|\nabla u|)}{\partial T}.$$

In the case of $\varphi(s) = |s|$, we get

$$L(u) = \frac{\pm 1}{|\nabla u|} [\nabla^2 u](N, T),$$

where $A(N, T) = \langle AN,\, T \rangle$. This resembles Bertalmio's choice of the Laplacian as

$$\Delta u = [\nabla^2 u](N, N) + [\nabla^2 u](T, T).$$