Approximating the Total Variation with Finite Differences or Finite Elements

Antonin Chambolle, Thomas Pock

Approximating the Total Variation with Finite Differences or Finite Elements

Antonin Chambolle[*,1] **and Thomas Pock**[†]

[*]*CMAP, CNRS, Ecole Polytechnique, Institut Polytechnique de Paris, Palaiseau, France,* [†]*Institute for Computer Graphics and Vision, Graz University of Technology, 8010 Graz*

ABSTRACT

We present and compare various types of discretizations which have been proposed to approximate the total variation (mostly, of a grey-level image in two dimensions). We discuss the properties of finite differences and finite elements based approach and compare their merits, in particular in terms of error estimates and quality of the reconstruction.

Keywords: Image processing, numerical analysis, total variation, finite differences, finite elements, error bounds.

MSC (2020): 49Q06 49Q20 49M29 26A45 65K10 65N30

1 INTRODUCTION

1.1 The total variation

In mathematical and computational imaging, images are represented by (possibly multi-channel) intensity functions and typical inverse problems aim at recovering such functions from measurements. Such problems are usually ill-posed and require some a priori knowledge of the class of signals one wishes to recover, this is usually expressed by means of a "regularizer", which measures how likely an image could be. The total variation has been proposed as a regularizer for imaging in the 90' [68]. It was introduced as a substitute to more common quadratic penalizations of the gradient, which prevent large jumps in the solution, and therefore are unsuited for recovering intensity functions representing images: indeed, these are discontinuous as they jump across the edges and boundaries of the objects. This regularizer is not so widely used nowadays as its performances are below many state-of-the art reconstruction methods (in particular, based on deep learning or other non-local techniques), yet, being convex and relatively

1. Now with CEREMADE, CNRS and Université de Paris-Dauphine PSL, France

simple, it is still useful and popular for large scale inverse problems (typically, medical imaging applications) or in low noise regimes. We refer to [25, 23] for a quick introduction to total variation for imaging and to [32] for more numerical details and algorithms. In fact, these notes are not only focused on imaging applications, as one might need to minimize the total variation for other goals, such as computing minimal surfaces, equilibrium configurations of mixtures with surface tension, etc. Our purpose is rather to discuss, from a numerical analysis point of view, the discretization of the total variation functional. For simplicity, we will mostly stick to the two-dimensional case, yet most of the results we state are not limited to this case. We will particularly focus on the quality (in terms of error bounds, when available) and the isotropic or anisotropic behaviour of various possible schemes. Error estimates are not so standard nor easy to compute for problems involving the total variation, as usual proofs rely on regularity properties of the continous solutions, while the minimizers of the problems we will be interested in are not regular and not even, in general, continuous. We will try to mention as exhaustively as possible what is known in this context.

1.2 Definition, basic properties

We consider throughout this note $\Omega \subset \mathbb{R}^n$ a bounded open set. In most of the practical applications, Ω will be a planar rectangle. Classically, the total variation is defined for a function $u \in L^1(\Omega)$ by duality [2, 44, 46] as follows:

$$TV(u;\Omega) = \sup\left\{-\int_\Omega u \operatorname{div}\varphi\, dx : \varphi \in C_c^\infty(\Omega;\mathbb{R}^n), \|\varphi(x)\| \le 1\ \forall x \in \Omega\right\}. \quad (1)$$

If u is a smooth function (or more generally, in the class $W^{1,1}(\Omega)$), one sees that after integration by part this is nothing else as $\int_\Omega |\nabla u| dx$; however it is also finite if $u = \chi_E$ is the characteristic of a smooth set. Indeed, using Green's formula the integral is $\int_{\partial E} \varphi \cdot \nu_E d\mathcal{H}^{n-1}$ where here, ν_E is an inner normal to ∂E and $\mathcal{H}^{n-1}$ the $(n-1)$-dimensional surface ("Hausdorff") measure, see for instance [2, 44] for a definition. Then the supremum (which is reached when $\varphi \approx \nu_E$) yields $TV(u;\Omega) = \mathcal{H}^{n-1}(\partial E \cap \Omega)$ (the *perimeter* of E in Ω), and indeed (1) is used, for $u = \chi_E$, as a defining a generalization of the perimeter for a measurable set [46]. Such a set is then called a "set with finite perimeter" in Ω (or "Caccioppoli set") as soon as $TV(\chi_E;\Omega) < +\infty$.

In general, it is clear (by Riesz' theorem) that $TV(u;\Omega) < +\infty$ if and only if Du, the distributional derivative of u, is a bounded Radon measure; one then has

$$TV(u;\Omega) = |Du|(\Omega) = \int_\Omega |Du|$$

the total variation of the vectorial measure Du. The space of functions $u \in L^1(\Omega)$ with bounded variation is denoted $BV(\Omega)$, with the norm $\|u\|_{BV} := \|u\|_{L^1(\Omega)} + |Du|(\Omega)$, it is a Banach space.

For $u \in BV(\Omega)$, it is shown (see for instance [2, 44, 74, 45]) that the measure Du decomposes as

$$Du = \nabla u\, dx + (u^+ - u^-)\nu_u \mathcal{H}^{n-1} \llcorner J_u + Cu \tag{2}$$

where: ∇u is the absolutely continuous part of Du with respect to the n-dimensional measure dx, J_u is the "jump set" (a $(n-1)$-dimensional subset), on which $\mathcal{H}^{n-1}$-a.e. one can define a normal vector $\nu_u(x)$, an upper value $u^+(x)$ and a lower value $u^-(x)$ (the simplest way is by blow up, defining J_u as the set of points x where the functions $y \mapsto u(x+\rho y)$, $\rho > 0$, converge in $L^1(B_1)$ as $\rho \to 0$ to $y \mapsto u^+(x)\chi_{\{y\cdot\nu_u(x)>0\}} + u^-(x)\chi_{\{y\cdot\nu_u(x)<0\}}$). The remaining part, Cu, or *Cantor part*, can be seen as made of infinitely many infinitesimal jumps. One has then obviously

$$TV(u;\Omega) = \int_\Omega |\nabla u(x)|dx + \int_{J_u} |u^+(x) - u^-(x)|d\mathcal{H}^{n-1}(x) + |Cu|(\Omega). \tag{3}$$

One sees from the remarks above that this functional (which thanks to (1) is trivially convex, lower-semicontinuous, with respect to the distributional convergence for u) is finite for functions which are possibly discontinuous, which is the main reason for its success in imaging. Indeed, if u represents the intensity values of an image (in 2D or 3D), using the total variation as a prior or regularizer for inverse problems will possibly promote images u with discontinuities, as expected across the edges or boundaries of objects (we refer to [25, 24, 23] for a mathematical study of the properties of the total variation as a regularizer).

From the numerical point of view, this raises important difficulties. How easy is it to precisely discretize a discontinuity and faithfully estimate its total variation? This is the main issue we will address in this small note, where we will review a few options proposed in the literature and discuss their merits and drawbacks.

1.3 Variational problems involving the total variation

To simplify, in order to assess the quality of a discretization we will concentrate on two simple reconstruction problems: the so-called "Rudin-Osher-Fatemi" or

"ROF" denoising problem and the "inpainting" problem. The first was proposed in [68] as an approach for image denoising (from the application's point of view, it is however a bit too elementary to be useful beyond the case of very low noise): assuming an observation $g \in L^\infty(\Omega)$ is the sum of a "clean" image $u(x)$ and an oscillatory signal $n(x)$ (in the discrete setting, a random Gaussian noise with known variance σ^2), the idea is to find u as the function with minimal total variation $TV(u;\Omega)$ among functions with $\int_\Omega (u-g)^2 dx = \sigma^2$. In practice, it is shown that the problem can be written (for an appropriate, and not obvious choice of parameter $\lambda > 0$, see for instance [29]) as follows:

$$\min_{u \in BV(\Omega)} \int_\Omega |Du| + \frac{1}{2\lambda} \int (u-g)^2 dx. \tag{4}$$

Even if this problem is elementary, one good reason for focusing on it is that its solution corresponds to evaluating the "proximity operator" of the total variation at g, and can be used as a basic brick in many other minimization algorithms involving this functional. Additionally, since we will focus mostly on the discretization of the first term in this problem, our study will apply with little or no change to many other second terms, such as appearing in more complex inverse problems. A typical denoising result obtained by (4) is shown in Fig. 1.

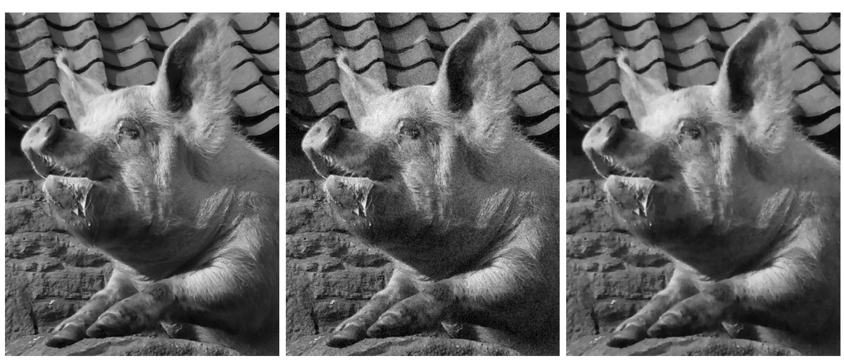


FIGURE 1 Left: original image, middle: noisy (Gaussian noise, $\sigma = 5\%$), right: denoised with (4)

We will make use also of the convex dual problem to (4), which is easily shown to be:

$$\sup \left\{ \frac{1}{2\lambda} \int_\Omega g^2 - (g + \lambda \operatorname{div} z)^2 dx : \right.$$
$$\left. z \in L^\infty(\Omega; \mathbb{R}^n), |z| \le 1 \text{ a.e.}, \operatorname{div} z \in L^2(\Omega), z \cdot \nu = 0 \text{ on } \partial\Omega \right\} \quad (5)$$

where here the condition $z \cdot \nu = 0$ on the boundary is understood in a weak sense, namely div $(z\chi_\Omega) \in L^2(\mathbb{R}^n)$. The solution u to (4) and a solution z to (5) are linked through the Euler-Lagrange equation:

$$\begin{cases} -\lambda \operatorname{div} z + u = g & \text{a.e. in } \Omega, \\ |z| \le 1 & \text{a.e. in } \Omega, \\ z \cdot Du = |Du|. \end{cases} \quad (6)$$

The last equation in (6) is in the sense of Anzelotti [3]: in case u is differentiable it implies that $z(x) = \nabla u(x)/|\nabla u(x)|$ where $\nabla u(x) \neq 0$. See [27] for more precise results in dimension 2 and 3.

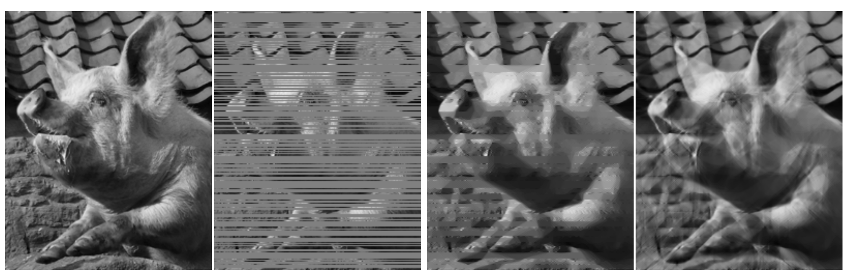


FIGURE 2 Original image, same image with 65% missing lines, total variation-based inpainting, elastica-based inpainting [33]

Variational "inpainting" is a bit more tricky to introduce. Formally, the problem consists in recovering missing data in an image by minimizing some functional, with a constraint that the image is known outside of some mask. Using the total variation for solving this problem is a terrible idea, as one can easily see that the best choice will consist in minimizing the length of each level line[2] inside the reconstructed region, which is not very wise: see Fig. 2 for an

2. thanks to the celebrated "co-area" formula which relates the total variation of a function and the perimeter of its level sets: $\int_\Omega |Du| = \int_{-\infty}^{+\infty} \left(\int_\Omega |D\chi_{\{u \ge s\}}| \right) ds$.

example, where we compare TV-inpainting with a variant penalizing also the *curvature* of the level sets, as proposed in [63, 57, 17, 33]. A simplified view of the TV-inpainting problem consists in minimizing the total variation with a Dirichlet boundary condition, formally:

$$\min_{u \in BV(\Omega)} \int_\Omega |Du| : \quad u = g \text{ on } \partial\Omega$$

for $g \in L^1(\partial\Omega)$. In fact, it is well known that this need to be relaxed (as a minimizing sequence (u_n) with $u_n = g$ for all n on $\partial\Omega$ could converge to a limit u with a different trace), and should rather be written (assuming here that Ω is for instance a Lipschitz domain — in practice we will consider rectangles or disks —, as things would be more complicated in the presence of interior boundaries):

$$\min_{u \in BV(\Omega)} \int_\Omega |Du| + \int_{\partial\Omega} |u(x) - g(x)| d\mathcal{H}^{n-1}(x). \tag{7}$$

Here, $u(x)$ in the second integral is the trace of u on the boundary (see for instance [46, Chap. 2]). The energy in (7) coincides with the total variation if the trace of u is equal to g on $\partial\Omega$, while if not, it counts in addition, as part of the energy and according to the penalization of the jump part in (3), the jump from the trace $u(x)$ to the value $g(x)$ on the boundary.

The dual formulation is a bit different from the definition in (1) as it is natural, in that case, to consider dual fields which do not vanish near the boundary $\partial\Omega$: formally, one sees that $\int_\Omega -u \operatorname{div} \varphi \, dx = -\int_{\partial\Omega} u\varphi \cdot \nu_\Omega d\mathcal{H}^{n-1} + \int_\Omega \varphi \cdot Du$ (where ν_Ω is the outer normal to Ω) and one can show that the energy in (7) can be defined as

$$TV_g(u; \Omega) = \sup \left\{ -\int_\Omega u \operatorname{div} \varphi \, dx + \int_{\partial\Omega} g\varphi \cdot \nu_\Omega d\mathcal{H}^{n-1} : \right.$$
$$\left. \varphi \in C^\infty(\overline{\Omega}; \mathbb{R}^n), \|\varphi(x)\| \leq 1 \ \forall x \in \overline{\Omega} \right\}. \tag{8}$$

In order to assert the quality of a discretization we will particularly be interested in (7) whenever $\Omega \subset \mathbb{R}^2$ is a square and $g = \chi_{\{x \cdot \vec{n} \geq a\}}$ for some direction $\vec{n} \in \mathbb{S}^{n-1}$ and $a \in \mathbb{R}$. In this case, it is easily seen that the (unique) solution is $u = g$ (that is, a straight line is the shortest path between two points...)

The current note aims at reviewing a few of the discretization schemes which have been proposed in the literature to numerically solve problems such as (4) and (7). We will consider scalar (grey-level valued) 2 dimensional images. Most of what we will discuss can be extended without pain to vectorial (multispectral or color) images although then, the metric which is used to measure distances in the data space starts to be highly relevant, hence also the norm used for measuring the gradient tensor in (4), see for instance [47, 42]. Also, most of what we will describe is trivially extended to three or higher dimensional problems. What we will not address here, or minimally, on the other hand, are

- the algorithmic issues, except to discuss the relative complexity of various discretizations from the optimization point of view (see for instance [32] for an overview);
- graph based approaches, which in general can be very efficiently optimized by combinatorial optimization techniques [52, 26], yet approximate, in the continuous limit, "crystalline" total variations (such as, for instance, $\int_\Omega |\partial_1 u| + |\partial_2 u|$ in 2D). From the numerical analysis point of view, theses approaches yield much better error estimates (see in particular [11] for a study), however the behaviour of the limiting functional can be very different from what is expected with the isotropic total variation (1) (in particular, the set of solutions for a crystalline version of (7) can be very large, contrarily to the isotropic case where generic uniqueness holds). The use of such approaches for approximating isotropic perimeters has been studied by many authors, starting from [14, 15] and comparison with "continuous" approaches are found for instance in [53, 56].

2 FINITE DIFFERENCES DISCRETIZATIONS OF THE TOTAL VARIATION

2.1 Standard discretization

In what follows, we consider to simplify a rectangular or square domain Ω and address the issue of discretizing problems such as (4). In most of the inverse problems and imaging literature, the isotropic total variation (1) is simply discretized as

$$TV^h(u) := h \sum_{i,j} \sqrt{(u_{i+1,j} - u_{i,j})^2 + (u_{i,j} - u_{i,j+1})^2}. \tag{9}$$

Here $h > 0$ represents the size of the pixels, and $u = (u_{i,j})_{i=1,\dots,M;j=1,\dots N}$ is a matrix representing the values of a (rectangular) image. We denote

$$S^h_{i-\frac{1}{2},j-\frac{1}{2}} = [(i-1)h, ih) \times [(j-1)h, jh)$$

the square "pixel" of size h^2 centered in $(ih - h/2, jh - h/2)$, for i, j integers indices. The standard convention in (9) on the boundary is to replace with zero the differences which involve indices which are out of range (e.g., $u_{M+1,j} - u_{M,j} = 0$ for all j, etc), however for addressing the Dirichlet case (8) one should rather replace the values of u out of range with the discretization of the boundary condition g.

Assume one views u as the discretization of an image defined in $\Omega = (0, Mh)\times(0, Nh)$: if $u \in C^1(\Omega)$, then clearly,[3] letting $u_{i,j} = u((i-\frac{1}{2})h, (j-\frac{1}{2})h)$ one has that

$$u_{i+1,j} - u_{i,j} = \int_{(i-\frac{1}{2})h}^{(i+\frac{1}{2})h} \partial_1 u(s, (j-\tfrac{1}{2})h)ds$$

and it is easy to see that $TV^h(u) \approx \int_\Omega |\nabla u| dx$: however typical minimizers of (4) are never C^1 (at best, Lipschitz, and often discontinuous, see for instance [22, 25]) and one cannot hope such a strong result for discontinuous functions. Worse than that: simple examples show that pointwise convergence obviously fails for basic discontinuous functions, such as $\chi_{\{(ih,jh)\cdot\vec{n}\geq a\}}$ for well chosen directions $\vec{n}$, see for instance [34, Fig. 1].

The convergence of TV^h to (1) is in fact quite weaker and needs to be understood in a variational sense (that is, for minimization problems), and more precisely in the sense of "Γ-convergence":

Proposition 2.1. *Let $h = 1/M = 1/N$, $\Omega = (0,1)^2$, and consider, for $u \in L^1(\Omega)$, the functional given by* (9)*, if*

$$u = \sum_{i=1}^{M} \sum_{j=1}^{N} u_{i,j} \chi_{S^h_{i-\frac{1}{2},j-\frac{1}{2}}}, \tag{10}$$

and $+\infty$ else. Then, as $h \to 0$, TV^h Γ-converges to:

$$TV(u;\Omega) = \begin{cases} \displaystyle\int_\Omega |Du| & \text{if } u \in BV(\Omega) \cap L^1(\Omega), \\ +\infty & \text{if } u \in L^1(\Omega) \setminus BV(\Omega); \end{cases} \tag{11}$$

The proof of this result is easy and relies in part on a duality argument, as developed in [36, Appendix] and [19] for variants. We refer to [38, 16] for the definitions and main properties of Γ-convergence. It means in practice that

3. We apologize for the small inconsistency in the notation, yet denoting $u_{i-\frac{1}{2},j-\frac{1}{2}}$ the pixel's value would have been a bit heavy and non standard.

minimizers of discrete energies such as

$$TV^h(u) + \frac{h^2}{2\lambda}\sum_{i=1}^{M}\sum_{j=1}^{N}(u_{i,j} - g_{i,j}^h)^2, \tag{12}$$

where $g_{i,j}^h$ is any consistent discretization of the function $g \in L^\infty(\Omega)$, such as

$$g_{i,j}^h = \frac{1}{h^2}\int_{S^h_{i-\frac{1}{2},j-\frac{1}{2}}} g(x)dx,$$

will converge to the minimizer of (4) as $h \to 0$. This numerical analysis result is a bit weak as it does not really specify how fast the discretization converges to the continuous limit. It turns out that one cannot expect a very fast convergence, as typical solutions to (4) have discontinuities [22]. In fact, if $u \in BV(\Omega)$ and one considers its piecewise constant projection

$$P_h u = \sum_{i,j} u_{i,j}^h \chi_{S^h_{i-\frac{1}{2},j-\frac{1}{2}}}, \quad u_{i,j}^h = \frac{1}{h^2}\int_{S^h_{i-\frac{1}{2},j-\frac{1}{2}}} u(x)dx \tag{13}$$

then it is standard that $\|u - P_h u\|_{L^1} \le Ch|Du|(\Omega)$, where C depends on the dimension. If in addition u is bounded (as the solution to (4) is, since one easily shows that $\|u\|_\infty \le \|g\|_\infty$), one has $\|u - P_h u\|_{L^2} \le C\sqrt{\|u\|_\infty |Du|(\Omega)}\sqrt{h}$. For u with discontinuities, one cannot hope for a better result, indeed if for instance $u = \chi_{\{x\cdot\vec{n}\ge a\}}$, a simple computation shows that $\|u - P_h u\|_{L^2} \approx C\sqrt{h}$, unless $\vec{n} = (1,0)$ or $(0,1)$ in which case one may have the error reduced to 0 (if the discontinuity is precisely between two pixels), or of lower order.

Error bounds for (12) are significantly larger than this theoretical limit. The most precise study is found in [71, 54, 55]. The main result of Wang and Lucier, [71, Thm 4.2], shows that if u^h is the minimizer of (12) and u the minimizer of (4), then (here u^h is identified with the piecewise constant function with value $u_{i,j}^h$ on the pixel $S^h_{i-\frac{1}{2},j-\frac{1}{2}}$):

$$\|u^h - u\|_{L^2} \le \|g\|_{\mathrm{Lip}(\alpha, L^2(\Omega))} h^{\frac{\alpha}{2(\alpha+1)}} \tag{14}$$

where $\alpha > 0$, and

$$\|g\|_{\mathrm{Lip}(\alpha,L^2(\Omega))} = \|g\|_{L^2(\Omega)} + \sup_{t>0} t^{-\alpha} \sup_{|z|\le t} \|g(x+z) - g(x)\|_{L^2(\Omega\cap(\Omega - z))}$$

is a measure of smoothness of the function g. In particular, if $g \in BV(\Omega) \cap L^\infty(\Omega)$, then $g \in \mathrm{Lip}(\frac{1}{2}, L^2(\Omega))$ and the rate of convergence (14) is $O(h^{1/6})$.

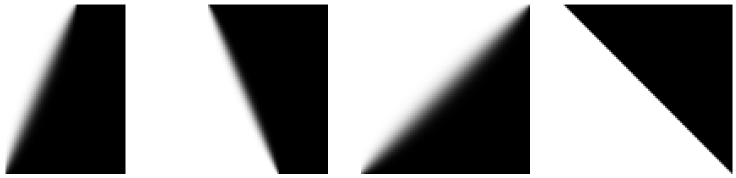


FIGURE 3 "Inpainting" with discretization (9): when solving (7) for $g(x) = \chi_{\{x\cdot\vec{n}\geq a\}}$ for different angles of the normal vector $\vec{n}$, the quality of the result strongly depends on the orientation.

The proof in [71] relies on a refined implementation of a relatively standard method for the numerical analysis of nonsmooth problems: the solution u is smoothed at a certain scale and then discretized; thanks to the smoothness, precise error rates can be derived for the discretization of the energy. Then, one selects the optimal scale for the smoothing, which depends on h and α, to obtain the bound (14). The same proof extends to many similar discretizations of the total variation, in particular, in [71], the result also applies to symmetrized variants of (9), as well as to an "upwind" variant introduced in [28], see Section 2.2 for details.

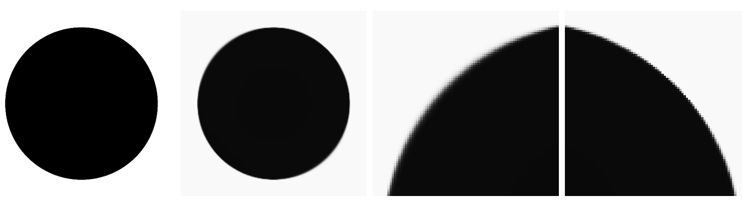


FIGURE 4 Original g: the characteristic of a disk. Solution u of (12). Details showing the anisotropic behaviour. The direction in which (9) approximates best the total variation of the discontinuity is better resolved (right).

On the other hand, it is known that (9) suffers important anisotropy issues and is not very precise. This is illustrated in Fig. 3 where a discrete approximation of the inpainting problem (7) is solved with this approximation in several directions: in some of them, the output is very blurry. This is due to the fact that a sharper result would have a very high discrete energy (and related to the fact that the convergence of TV^h to TV is not pointwise). Also Problem (4) is affected by this issue. In Fig. 4, g is simply the characteristic of a disk, and so is u with a slight loss of contrast (*cf* for instance [58]). However, one sees that the boundary of the disc is sharper in some directions than other. The anisotropy is even more striking if g is the characteristic of a square, see Fig. 5, and leads to quite bad results.

A precise study of this issue is presented in [20]. The following experiment is studied: one considers a domain Ω which is a periodic strip, either of the form $\{|x_1 - x_2| \le L\}$ or $\{|x_1 + x_2| \le L\}$. In each case, g is the characteristic of a half-strip (respectively, $\chi_{\{x_1 \ge x_2\}}$ and $\chi_{\{x_1 \ge -x_2\}}$). The idea is that in both cases, by translational invariance, both problems (4) and (12) boil down to a 1D total variation regularization problem which is easy to analyse. It is then shown that while in the second case, the recovery is perfect (the unique solution u^h is precisely $P_h u$), in the first, the difference of between the discrete and continuous energies is of order $h^{2/3}$, and while it is yet not proven that $\|u^h - u\|_{L^2} \approx Ch^{1/3}$ as $h \to 0$, an almost optimal signal which achieves this rate is built (see [20, Sec. 3]).

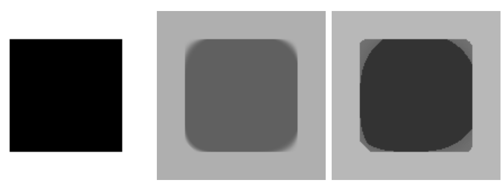


FIGURE 5 Original g: the characteristic of a square. Here the anisotropic character of (9) is striking (right, equalized version of the solution u of (12) shown in the middle).

While the anisotropy issue is easily corrected, as we will see in the next Section 2.2, it is yet not clear that this error can be improved by considering simple variants of this discrete energy. Actually, the best reason for using the discretization (9) is that it is simple and easy to implement. In particular, Problem (4) can be easily solved by many standard proximal algorithms for nonsmooth convex optimization, including with acceleration, such as the accelerated primal-dual method proposed by Zhu and Chan [73] (see also [43]) and its variant (with convergence guarantees and rates) [30, Alg. 2], or an accelerated proximal gradient descent on the dual and in particular the "FISTA" algorithm [13], see also [60, 62] or the variant in [61]. A slight smoothing of the total variation (replacing TV^h with a "Huber" total variation, quadratic when the differences are nearly zero) makes the problem even more regular and the above algorithms converging with a very efficient linear rate (*cf* [60, (2.2.19) and Thm 2.2.3]), see [32] for a general description.

2.2 Improving the standard discretization

It is not hard to reduce the anisotropic character of discretization (9) by relatively simple techniques, usually at the expense of (slightly) increasing the complexity of the discretization. We mention in this section a few straightforward techniques. All produce energies which are invariant by a 90° rotation, however without clearly improving the quality of the approximation in terms of precision (all variants in this section could be analysed by the methods in [71] and enjoy similar approximation properties).

The first variant, introduced and studied in [71, 54], consists simply in averaging the four possible rotations of (9), by defining $TV^h(u)$ as:

$$\frac{h}{4}\sum_{i,j}\sqrt{(u_{i+1,j}-u_{i,j})^2+(u_{i,j}-u_{i,j+1})^2}+\sqrt{(u_{i-1,j}-u_{i,j})^2+(u_{i,j}-u_{i,j+1})^2}$$
$$+\sqrt{(u_{i+1,j}-u_{i,j})^2+(u_{i,j}-u_{i,j-1})^2}+\sqrt{(u_{i-1,j}-u_{i,j})^2+(u_{i,j}-u_{i,j-1})^2}. \tag{15}$$

As before, the differences involving indices out of range are set to zero. Notes that this cannot solve the issue of the precision and smoothing of discontinuities, since in this case, *all* slanted sharp discontinuities are overestimated by the energy. This is visible in both Figures 6 and 7 (left).

The "upwind" variant in [28] is defined by

$$TV_u^h(u) = h\sum_{i,j}\Big([(u_{i,j}-u_{i+1,j})^+]^2+[(u_{i,j}-u_{i-1,j})^+]^2$$
$$+[(u_{i,j}-u_{i,j+1})^+]^2+[(u_{i,j}-u_{i,j-1})^+]^2\Big)^{1/2}. \tag{16}$$

Here, the notation x^+ stands for the positive part $\max\{x,0\}$. It is experimentally demonstrated in [28] that this greatly improves the anisotropy issues raised by (9). The error analysis in [71] is also carried on for this variant, with the same result. In addition, one can see that that convergence to $\int|Du|$ does hold pointwise for u the characteristic of a half-plane bounded by a horizontal, vertical or ±45° diagonal line, so that one could expect a better behaviour in terms of precision (no smoothing is required to approximate such a discontinuity, and probably less in many directions than with (9)) and the analysis in [71] might be suboptimal for this approach, yet a more precise analysis is not available. A strange feature of (16), though, is that in general, $TV_u^h(-u)\neq TV_u^h(u)$. In Figure 6 we see that using this latter version effectively improves the inpainting result of Fig. 3 in the directions where (9) fails, while using the previous variant (15) gives a very smooth reconstruction in all directions.

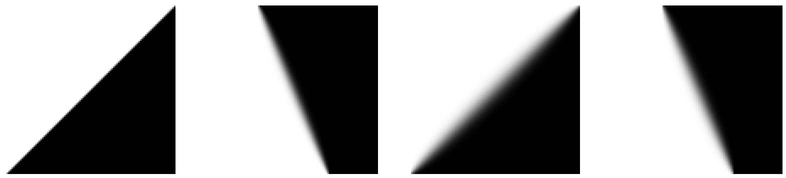


FIGURE 6 Inpainting with (16) (left) and (15) (right).

We mention a last discretization also enjoying this isotropy property, introduced in [31, 36] and defined as follows:

$$
\begin{aligned}
TV_4^h(u) = h \sum_{i,j \text{ even}} & \sqrt{2}\Big((u_{i+1,j} - u_{i,j})^2 + (u_{i+1,j+1} - u_{i+1,j})^2 \\
& \qquad + (u_{i,j+1} - u_{i+1,j+1})^2 + (u_{i,j} - u_{i,j+1})^2 \Big)^{1/2} \\
+ h \sum_{i,j \text{ odd}} & \sqrt{2}\Big((u_{i+1,j} - u_{i,j})^2 + (u_{i+1,j+1} - u_{i+1,j})^2 \\
& \qquad + (u_{i,j+1} - u_{i+1,j+1})^2 + (u_{i,j} - u_{i,j+1})^2 \Big)^{1/2}.
\end{aligned}
\tag{17}
$$

As before, it does not converge pointwise for oblique discontinuity and it is unclear that one could hope for better error estimates than obtained in [71]. The only result proved for sure is again its Γ-convergence to the total variation (1), see [36, Appendix]. This approach is introduced for optimization purposes, indeed, dualizing (12) with TV_4^h yields independent problems on "even" and "odd" squares (indexed by indices i, j both even or both odd), which can be solved in parallel, and an accelerated alternating descent method yields extremely fast results on parallel hardware.

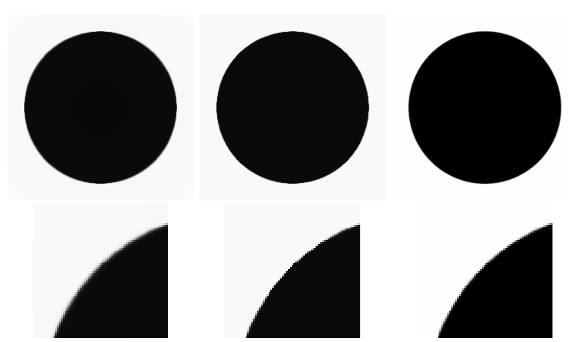


FIGURE 7 Same as Fig. 4 using (15) (left), (16) (center), (17) (right) (bottom: detail).

Figure 7 compares these variants in the setting of Figure 4. In particular, the last two approaches yield sharper results and are to be preferred.

2.3 The "Shannon" total variation [1]

While the (more) isotropic variants of (9) rely on heuristic arguments to make the approximation more isotropic, a few authors have tried to address this problem using a somewhat deeper analysis. An interesting attempt is found in [1] (see also [59]), introduced as a "Shannon" total variation. The idea is to view an image as a low frequency representation of a continuous signal and try to estimate the total variation of the latter. This needs to be approximated: a possibility is to over-sample a $M \times N$ image $(u_{i,j})$ as a $nM \times nN$ image, by "zero-padding" the discrete Fourier coefficients (that is, considering the Fourier transform $\hat{u}_{k,l}$, $-M/2 \leq k \leq M/2$, $-N/2 \leq l \leq N/2$ as the low frequencies of the larger signal, the other frequencies being set to 0). A striking property of such a definition is that images with low total variation can be interpolated via zero-padding without creating further artefacts (this is clearly the main reason to do so, see [1] for many examples). On the other hand (and, in some sense, as a consequence), sharp discontinuities cannot be represented by sharply discontinuous discrete functions (whose interpolation will necessarily be altered by an important "ringing" phenomenon).

We describe rapidly the construction in [1], using the notation in that paper. Assuming $u_{k,l}$[4] is defined for $(k,l) \in I_M \times I_N$ where $I_M = \{0, \ldots, M-1\}$, the authors of [1] first introduced the "Shannon" interpolate of u by considering the given data as a low-frequency image in $\mathbb{R}^2$: first letting

$$\hat{u}(\alpha, \beta) = \sum_{(k,l) \in I_M \times I_N} u_{k,l} e^{-2i\pi\left(\frac{\alpha k}{M} + \frac{\beta l}{N}\right)}$$

for $(\alpha, \beta) \in \mathbb{Z}^2$, the Shannon interpolate is then given, for $(x, y) \in \mathbb{Z}^2$ (actually it is a $[0, M] \times [0, N]$-periodic function), by

$$U(x,y) = \frac{1}{MN} \sum_{\substack{-\frac{M}{2} \leq \alpha \leq \frac{M}{2} \\ -\frac{N}{2} \leq \beta \leq \frac{N}{2}}} \varepsilon_\alpha^M \varepsilon_\beta^N \hat{u}(\alpha, \beta) e^{2i\pi\left(\frac{\alpha x}{M} + \frac{\beta y}{N}\right)}. \tag{18}$$

Here α, β are integers and the coefficients ε_α^M, ε_β^N, given by $\varepsilon_\alpha^M = 1$ if $|\alpha| < M/2$ and $1/2$ else, etc, are introduced for symmetry reasons.

4. Here, the notation i is preserved to denote $\sqrt{-1}$.

Then, one can compute the (true) gradient of U by simple derivation under the $\sum$ sign in (18), and given a $n \geq 2$ an oversampling factor, one lets for $(k, l) \in I_{nM} \times I_{nN}$

$$\nabla_n u_{k,l} = \nabla U\left(\frac{k}{n}, \frac{l}{n}\right)$$

and:

$$STV_n(u) = \frac{1}{n^2} \sum_{(k,l) \in I_{nM} \times I_{nN}} |\nabla_n u(k, l)|.$$

The paper [1] provide formulas for efficiently computing ∇_n and its adjoint $-\mathrm{div}_n$ (see [1, Prop. 8]), by means of two FFTs, and these can be used to implement standard proximal algorithms for solving problems such as the discretizations of (4) or (7).

In Fig. 13e we show a result of denoising obtained with this approach (for $n = 3$). While the result is nice and does not seem to favor any direction, it is a bit less sharp than with other discretizations of the total variation (*cf.* Figs. 13–14). We refer to [1] for more applications (in particular to zooming, which really is, by construction, where this approach is superior to many others). On the other hand, this clearly is not adapted to inpainting very sharp edges, as such objects can precisely not be "Shannon"-interpolated without a strong ringing effect, and experiments will always produce smooth edges, in all directions.

3 FINITE ELEMENTS BASED APPROXIMATIONS

In order to introduce discretization with better approximation properties and smaller error bounds, it is tempting to rely on the machinery of finite elements. We now discuss this point of view and show that it can be actually fruitful, provided one does not stick to the most straigthforward approximation spaces.

3.1 P1 discretizations

The most natural approaches rely on conforming "P1" representation: in this case, the domain is discretized with a simplicial mesh (triangles in 2D) and the functions are assumed to be piecewise affine, continuous, and defined by the values at the vertices of the triangulation. The natural basis for this finite element space is of course made of the functions which are 1 on one node and zero on all other nodes of the mesh. To simplify, we assume the dimension is 2 (although this does not really matter in this section). Assuming $\mathcal{T}_h = \{T_1, \ldots, T_N\}$ is a triangulation[5] of Ω (which for simplicity we assume is a polygon, typically a square or rectangle), with triangles with maximal edge of size $\leq h$, we denote $\mathrm{P1}(\mathcal{T}_h)$ the space of such P1 functions. This approximation is "conforming" and the the total variation at scale h is simply given by the total variation itself:

$$TV^h(u) = \int_\Omega |\nabla u| dx = \sum_{T \in \mathcal{T}_h} |T||\nabla_T u| \tag{19}$$

where $\nabla_T u$ denotes the constant gradient of u on the triangle T.

This representation enjoys slightly better error estimates than the finite-differences based approximations, as shown in particular in [10] (see also [4]). We consider as before the minimization of the energy:

$$TV^h(u_h) + \frac{1}{2} \int_\Omega (u_h - g)^2 dx \tag{20}$$

for $u_h \in \mathrm{P1}(\mathcal{T}_h)$, with TV^h given by (19). Then, we have the following result:

Proposition 3.1 ([10, Thm 7.1])**.** *Let u_h be the minimizer of* (20) *and as before, u the minimizer of* (4)*, and assume Ω is star-shaped. Then one has:*

$$\|u_h - u\|_{L^2} \leq c h^{1/4} |Du|(\Omega).$$

Here the assumption that Ω is star-shaped is used to show that an appropriate smoothing u_ε of u at scale $\varepsilon > 0$ will satisfy $|Du_\varepsilon|(\Omega) \leq |Du|(\Omega) + c\varepsilon$. The statement follows comparing the energy of u_h with the energy of the projection of u_ε on $\mathrm{P1}(\mathcal{T}_h)$, for the choice $\varepsilon = h^{1/2}$. It is also observed that if TV^h were replaced with a "TV-diminishing" approximation of the total variation (ie., with $TV^h(u_h) \leq TV(u; \Omega)$ for u_h the projection of u onto $\mathrm{P1}(\mathcal{T}_h)$), then the estimate could be improved to $h^{1/2}$: such an approximation is however not known in this context, see [11] for a diminishing interpolation for the anisotropic ℓ_1-based total variation.

Numerical methods for solving the discrete problems have been studied and compared in particular in [8, 9, 69] and strategies for mesh refinement based on a posteriori error estimates are found in [5, 9].

One must observe that P1 discretization suffers more or less the same drawbacks as the finite differences based approaches mentioned before. The analysis is very simple: consider a finite-perimeter set $E \subset \Omega \subset \mathbb{R}^2$ (that is a measurable set with $|D\chi_E|(\Omega) < \infty$, but to simplify we even consider a smooth set).

5. That is, we assume $\Omega \subset \bigcup_{i=1}^N T_i$, the closed triangles T_i are non overlapping and share only common edges and/or common vertices.

Consider also triangulations $\mathcal{T}_h$ with maximal diameter $h > 0$, and the basic projection u_h of $u = \chi_E$ onto $\mathrm{P1}(\mathcal{T}_h)$ consisting in letting $u_h(x) = \chi_E(x)$ for any vertex, assuming that no vertex is on the boundary (which, up to an infinitesimal translation, can always be assumed). Then, given a triangle $T \in \mathcal{T}_h$ crossing the boundary, one will have two vertices with value 0 and one with value 1, or the opposite. It follows that the gradient $Du_h(T)$ will be a vector orthogonal to the edge E of T with equal values, and magnitude $1/h_E$ the corresponding height. In particular, $|T||Du_h|(T) = |T|/h_E$ will be precisely the length of the segment parallel to E joining the middle of the two other edges, and in the end, the total variation $\int_\Omega |Du_h|$ will be precisely the perimeter of the polygonal set $\hat{E}$ bounded by these segments, see Fig. 8.

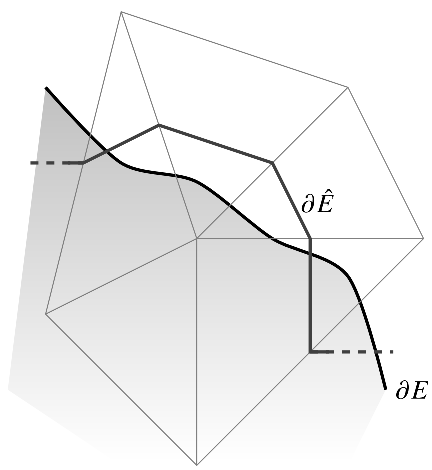


FIGURE 8 With a P1 finite elements approximation, the perimeter $\int |D\chi_E|$ of E is approximated with the perimeter of the polygonal set $\hat{E}$ bounded by the jagged line passing through the middle of the edges which are crossed by ∂E.

In general (and in the limit $h \to 0$), the polygonal line $\partial\hat{E}$ is much longer than ∂E, so that one cannot expect a pointwise convergence of TV^h to TV. This means again that, as with finite differences, a characteristic function (and in general any discontinuity) needs to be smoothed in order to have its energy well approximated at the discrete level. This is visible in the example in Fig. 9 which shows the solution u_h of (20) with g the characteristic of the unit disk, computed using `FreeFem++` [48].

In the next sections we describe recent approaches for tackling this issue. The first series is based on non-conforming finite elements representations of

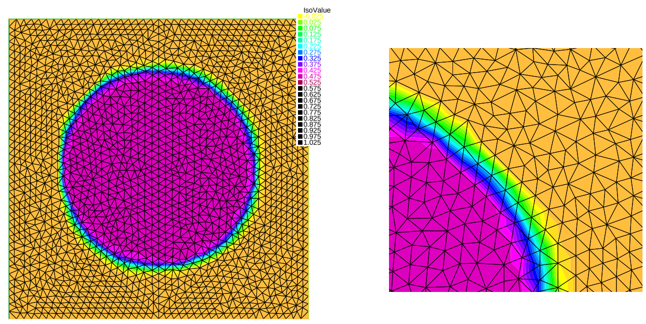


FIGURE 9 "ROF" denoising of the characteristic of a disk with P1 finite elements (right: detail).

the functions. The second is related, but can be seen as a careful discretization of the dual problem.

3.2 Non conforming FEM discretization

3.2.1 Definition. Denoising with non-conforming FEM.

In a recent attempt to improve over the standard P1 discretization, it has been suggested in [34] to consider rather nonconforming P1 discretizations of BV functions, following similar attempts for various nonlinear problems [49, 72, 41, 65, 64, 21]. A survey of such elements is found in [18]. The nonconforming-P1 or Crouzeix-Raviart (CR) functions over a mesh $\mathcal{T}_h$ of a domain Ω are functions u_h with are affine inside each $T \in \mathcal{T}_h$, but required to be continuous only at the center of each facet of ∂T. We will denote $\mathrm{CR}(\mathcal{T}_h)$ the (finite dimensional) vector space of such functions.

The natural projection u_h of a function $u \in W^{1,1}(\Omega)$ onto CR functions is obtained by averaging u on each facet of the mesh, and assigning the value to the center of the corresponding facet. It is easily shown, by Green's formula, that it produces a piecewise affine function (not continuous) which is such that on each $T \in \mathcal{T}_h$, denoting $\nabla_T u_h$ the (constant) gradient of u_h on T, one has

$$\nabla_T u_h = \frac{1}{|T|} \int_T \nabla u(x) dx.$$

In other words, projecting u onto CR functions is equivalent to considering a "P0" (piecewise constant) approximation of ∇u.

The situation is almost the same for a BV function u. One just needs to be a bit cautious, as in case the jump J_u of u has positive measure on a facet, then "averaging" u on this facet does not really make sense (for instance one should average the trace of u from either one side or the other, which might differ).

The simplest way to deal with this (technical) issue is to define more precisely the elements $T \in \mathcal{T}_h$ as possibly semi closed / semi open simplices (and not just, as usual, open simplices or closed simplices). Precisely, to each pair of elements T, T' with $\mathcal{H}^{n-1}(\partial T \cap \partial T') > 0$ (n being as before the dimension of the space), we associate (arbitrarily) to one and *only one* of the elements T or T', say T, the common $(n-1)$-dimensional facet $F = \partial T \cap \partial T'$, that is we assume that $F \subset T$ and $T' \cap F = \emptyset$.

Then, the average of u on the facet F is understood as the average of the trace of $u_{|T'}$ restricted to T'. In this sense, if $\mathcal{H}^{n-1}(J_u \cap F) > 0$, so that in particular $|Du|(F) > 0$, the value $Du(F)$ measures (thanks to the decomposition (2)) the possible discrepancy between the average of the trace of $u_{|T\setminus F}$ and of the trace of $u_{|T'}$, one obtains as before the formula

$$\nabla_T u_h = \frac{1}{|T|} Du(T),$$

using that $Du(T) = Du(T \setminus F) + Du(F)$.

Now, the interesting point about this construction, is that it obviously follows (from Jensen's inequality), that given $\Phi : \mathbb{R}^n \to \mathbb{R} \cup \{+\infty\}$ any convex potential,

$$\sum_{T \in \mathcal{T}_h} |T| \Phi(\nabla_T u_h) = \sum_{T \in \mathcal{T}_h} |T| \Phi\left(\frac{1}{|T|} Du(T)\right) \leq \sum_{T \in \mathcal{T}_h} \Phi(Du(T)) = \int_\Omega \Phi(Du)$$

and in particular, this holds for the standard total variation which corresponds to $\Phi = |\cdot|$. It makes interesting to define, for such Φ, the discrete functional

$$J^h_\Phi(u_h) = \sum_{T \in \mathcal{T}_h} |T| \Phi(\nabla_T u_h), \tag{21}$$

over all CR functions on the mesh $\mathcal{T}_h$. Indeed, as we just observed, one has $J^h_\Phi(u_h) \leq \int_\Omega \Phi(Du)$ if u_h is the projection of u onto CR functions. This is interesting because of the following result: assuming Φ in addition is coercive: $\lim_{|p|\to\infty} \Phi(p) = +\infty$, one has

Proposition 3.2. *As $h \to 0$, J^h_Φ Γ-converges in $L^1(\Omega)$ to*

$$J_\Phi(u) = \begin{cases} \int_\Omega \Phi(Du) & \text{if } u \in BV(\Omega), \\ +\infty & \text{else.} \end{cases}$$

The proof of this result is identical to the proof of [34, Prop. 3.1]. A corollary (or part of the proof) is the following pointwise convergence:

Corollary 3.3. *Given $u \in BV(\Omega)$, denote u_h its projection onto CR functions over the mesh $\mathcal{T}_h$, defined as above. Then* $\lim_{h\to 0} J_\Phi^h(u_h) = J_\Phi(u)$.

This is interesting, as even if it does not guarantee any precise error estimate, it makes more likely that the solution of a discretized problem is close, or equal to, the solution of the corresponding continuous problem.

We apply this construction to $\Phi = |\cdot|$ (so that $J_\Phi = TV(\cdot;\Omega)$), defining in this way a *Crouzeix-Raviart total variation* $TV_{CR}^h = J_\Phi^h$. A remark is that for $u \in BV(\Omega)$, the CR projection u_h also satisfies $u_h \in BV(\Omega)$, so one could consider CR functions as "conforming" in the BV setting. We claim however that the approximation TV_{CR}^h is still "non-conforming" in the sense that it does not consider the jumps of u_h, carried by the facets of the mesh, as part of its derivative, as done for instance in [50, 67] where the total variation is approximated by the full functional $\int_\Omega |Du_h|$ evaluated on the approximate CR functions. In the latter case, one has to expect a mesh dependency phenomenon similar to what is illustrated in Fig. 8 for the approximation of the total variation and should therefore also expect a smoothing of the discontinuity as observed in Fig. 9.

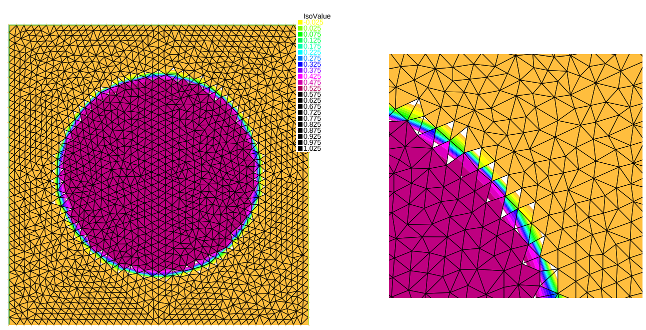


FIGURE 10 "ROF" denoising of the characteristic of the unit disk with Crouzeix-Raviart finite elements (right: detail). While in Fig. 9 the computed energy estimates a perimeter of 6.37, in this experiments perimeter ≈ 6.29.

The difference between using (20) and the same problem with the variant TV_{CR}^h of the discrete total variation is striking in Fig. 10. Now, the orientation of the discontinuity is better rendered and the transition region is sharper (in our particular example, about 2 elements, while in Fig. 9 the discontinuity is smoothed over about 6-7 elements). On the other hand, as a drawback, the values of the function are not well recovered at the discontinuity. However in general, and particularly for this example, the precision of the algorithm is improved in this framework. One finds in particular in [34] the following result:

Proposition 3.4 ([34, Sec. 5.1.1]). *Assume $g \in L^\infty(\Omega)$ is such that problem* (5) *admits as a solution a Lipschitz-continuous dual field. Then, if u_h is the minimizer of* (20) *with TV^h replaced with TV^h_{CR} and u as before the minimizer of* (4)*, one has*

$$\|u_h^0 - u\|_{L^2} \leq c h^{1/2} |Du|(\Omega).$$

where u_h^0 is the P0 average of u_h and for a constant c depending on $\|g\|_\infty$ and the Lipschitz constant of the dual field.

Here, the P0 average is simply the piecewise constant function obtained by averaging u_h over each element, or equivalently since u_h is affine on each element, by assigning to each element the value of u_h in the center of the element. Unfortunately, it is unclear when one can expect (5) to have a Lipschitz continuous solution. Even in 2D, assuming that $g \in L^\infty(\Omega)$ is not sufficient, yet it is standard that it is the case in the setting of Fig. 10. The proof of Prop. 3.4 is based on standard convex duality, and makes use in particular of an interesting discrete duality property enjoyed by the Crouzeix-Raviart total variation and that we rapidly mention in the next Section 3.2.2.

Other discontinuous Galerkin (DG) type approximations have been introduced recently and less recently for tackling such problems, see in particular [7] for a very general framework, and further estimates which generalize Prop 3.4. We also mention [50] where is studied a *conforming* approximation which is a special limiting case of [7], closer to the classical work of Repin [67] (see also [9]), and for which we should not expect an error bound as good as in Prop. 3.4. An interesting variant also based on a DG representation is described [70], where the functions are sampled after convolution with a mask, and mesh adaption is used, leading to an expression which enjoys also strong isotropic properties — and could share similarities with the approximations presented later on (see Sec. 4.2 and 4.3). However, for the latter approaches, it is quite unclear right now how to obtain sound error estimates.

3.2.2 Duality

Let us first introduce some additional notation and definitions. In what follows, given a mesh $\mathcal{T}_h$ and $u \in L^1(\Omega)$ we denote by $\Pi_h^0(u)$ the (L^2) projection of u on

P0 functions, that is functions which are constant on each element $T \in \mathcal{T}_h$. It is defined as usual by

$$\Pi_h^0(u) = \sum_{T \in \mathcal{T}_h} \left(\frac{1}{|T|} \int_T u(y)dy \right) \chi_T .$$

We denote $\mathrm{P0}(\mathcal{T}_h)$ the vector space of the P0 functions over $\mathcal{T}_h$.

We also define the space of "lowest order Raviart-Thomas" vector field subject to the mesh $\mathcal{T}_h$, denoted $\mathrm{RT0}(\mathcal{T}^h)$. The basis of this space is composed of the affine vector fields which vanish on all but two elements T, T' (or one boundary element), have flux $|F|$ through the common facet $F = \partial T \cap \partial T'$ (or the boundary facet in case of a boundary element) and zero through the other facets. If z (resp., z') is the vertex of T (resp., T') opposite to F, this vector field (up to switching the sign and/or exchanging T, T'), is given by

$$\phi_{T,T'}(x) = \begin{cases} \frac{x-z}{h_T} & \text{if } x \in T, \\ -\frac{x-z'}{h_{T'}} & \text{if } x \in T', \\ 0 & \text{else} \end{cases}$$

where here, $h_T = n|T|/|F|$ is the height of the element T over the facet F, etc. One sees that $\phi_{T,T'}$ is not continuous, however its normal flux through F is one ($|F|$ after integration on the facet) and does not jump, so that

$$\operatorname{div} \phi_{T,T'} = \frac{n}{h_T}\chi_T - \frac{n}{h_{T'}}\chi_T = |F| \left(\frac{\chi_T}{|T|} - \frac{\chi_{T'}}{|T'|} \right)$$

in the sense of distributions. Hence, a generic vector field $\phi^h : \Omega \to \mathbb{R}^n$ is of the form $\sum_{T,T'} f_F \phi_{T,T'} + \sum_T f_F \phi_T$ where the first sum is over elements T, T' such that $F = \partial T \cap \partial T' \neq \emptyset$, and the second over elements T such that $F = \partial T \cap \partial\Omega \neq \emptyset$ (and ϕ_T is defined as above, but on T only). Here $f_F \in \mathbb{R}$ are the pointwise fluxes from T to T' across the common facet F (oriented in an arbitrary but fixed direction). Such a field ϕ^h has a true divergence (in the weak sense). We also define the subspace $\mathrm{RT0}_0(\mathcal{T}^h)$ of fields $\phi^h \in \mathrm{RT0}(\mathcal{T}^h)$ with $\phi^h \cdot \nu = 0$ on $\partial\Omega$, that is fields with only the sum over T, T' above, and for which $f_F = 0$ for $F \subset \partial\Omega$.

Defining now, for $u_h^0 \in \mathrm{P0}(\mathcal{T}_h)$, the discrete total variation:

$$TV_0^h(u_h^0) = \min \left\{ TV_{CR}^h(v) : v \in \mathrm{CR}(\mathcal{T}_h),\ \Pi_h^0 v = u_h^0 \right\},$$

one has the following result:

Proposition 3.5 ([34, Thm. 1])**.** *For any P0 function u_h^0,*

$$TV_0^h(u_h^0) = \sup\left\{-\int_\Omega u \operatorname{div}\phi^h : \phi_h \in \mathrm{RT0}_0(\mathcal{T}^h), |\phi^h(c_T)| \le 1\ \forall T \in \mathcal{T}_h\right\}. \tag{22}$$

We must observe a slight difference between this result and the result as stated in [34], where u_h^0 is assumed to be already the projection of a CR function. This is due to the fact, recently proved in [12, Cor. 3.2], that the operator Π_h^0 maps $\mathrm{CR}(\mathcal{T}_h)$ onto $\mathrm{P0}(\mathcal{T}_h)$, so that it does not make any difference to assume u_h^0 is *any* P0 function. The above mentioned paper [12] studies more deeply the connections between CR and RT spaces. This duality property has interesting extensions studied by S. Bartels in [6] in the framework [7] of more general DG approximations (see also [34, App. A] for an elementary generalization).

3.3 Inpainting with Crouzeix-Raviart finite elements

It turns out that the Crouzeix-Raviart total variation enjoys also a remarkable isoptropic behaviour for inpainting problems. We consider an arbitrary (polygonal) domain Ω and problem (7) with $g(x) = \chi_{\{x\cdot\vec{n}\ge t\}}$ for some direction $\vec{n} \in \mathbb{S}^{n-1}$ and $t \in \mathbb{R}$, that is, the characteristic function of a half-space. Then, one has the following striking result:

Proposition 3.6 ([34, Prop. 3.3.])**.** *Let $g_h \in \mathrm{CR}(\mathcal{T}_h)$ be the projection of g onto CR functions. Then for any $u_h \in \mathrm{CR}(\mathcal{T}_h)$ with $u_h = g_h$ on the middle of the boundary facets (on $\partial\Omega$), one has: $J_h(u_h) \ge J_h(g_h)$.*

In other words, the discrete projection of the exact solution of the inpainting problem solves the discrete inpainting problem, for any direction $\vec{n}$. Unfortunately, what the Proposition does not say, is that in general there could be many other solutions and it is very unlikely to numerically find g_h by solving the discrete inpainting problem. This issue is studied in details in [34, Sec. 3.4] and we refer to that paper for a numerical illustration. We will see later on that Prop. 3.6 relies on general properties which can be shared by other discretization, see Sec. 4.1.

4 DUAL DISCRETIZATIONS

We now turn to discrete total variations defined rather by a dual formulation and appropriate constraints on dual fields, that is, by an appropriate discretization of (1), rather than by a primal energy. The CR approximation, which enjoys (22), could fall into this category, however for the formulations we will now describe, there is no obvious primal formulation such as (21) (for $\Phi = |\cdot|$).

4.1 Raviart-Thomas duals

The first and most natural way to define a discrete variation by a dual formula is to let, given the mesh $\mathcal{T}_h$ and $u_h \in \mathrm{P0}(\mathcal{T}_h)$,

$$TV^h_{RT}(u_h) = \sup\left\{\int_\Omega u_h \operatorname{div}\phi_h\,dx : \phi_h \in \mathrm{RT0}_0(\mathcal{T}_h), |\phi_h(x)| \le 1\ \forall x \in \Omega\right\} \tag{23}$$

Observe, actually, that this formula also makes sense if u_h is replaced with $u \in L^1(\Omega)$, and depends only on $\Pi^0_h(u)$. In particular, it is easy to show that (comparing with (1)) one will have, for any $u \in L^1(\Omega)$,

$$TV^h_{RT}(\Pi^0_h(u)) \le TV(u;\Omega).$$

It turns out that with this property, and the fact that if $z(x)$ is a Lipschitz dual field for problem (5), then one can easily "project" it as an admissible dual field for the definition (23) (up to some small error), one can reproduce almost verbatim the proof of Prop. 3.4 and get the same estimate for this variant, see [20] for details.

In the same way, one can show also a variant of Prop. 3.6, valid for this approximation. To make the statement precise, we assume Ω is a rectangle (or a polygon) and we consider a mesh $\mathcal{T}_h$ over Ω, made of simplices (triangles in 2D). Then, we introduce the discrete version of (7)–(8), given $g \in L^1(\partial\Omega)$. We let:

$$\begin{aligned} TV_{g,RT}(u;\Omega) = \sup\Bigg\{ -\int_\Omega u \operatorname{div}\phi_h dx + \int_{\partial\Omega} g\phi_h\cdot\nu_\Omega d\mathcal{H}^{n-1} : \\ \phi_h \in \mathrm{RT0}(\mathcal{T}_h), |\phi_h(x)| \le 1 \ \text{ for a.e. } x \in \overline{\Omega}\Bigg\}, \end{aligned} \tag{24}$$

which of course, since $\operatorname{div}\phi_h \in \mathrm{P0}(\mathcal{T}_h)$, depends only on the $P0$ projection of u and defines, in particular, a Dirichlet discrete total variation on $\mathrm{P0}(\mathcal{T}_h)$.

Integrating by parts, the expression in the sup is the same as

$$\int_\Omega \phi_h \cdot Du + \int_{\partial\Omega} (g(x) - u(x))\phi_h \cdot \nu_\Omega(x) d\mathcal{H}^{n-1}(x) \\ \leq \int_\Omega |Du| + \int_{\partial\Omega} |u(x) - g(x)| d\mathcal{H}^{n-1}(x)$$

which is the expression in (7), and equality is reached whenever $\phi_h \cdot Du = |Du|$ (in the sense of measures) and $\phi_h \cdot \nu_\Omega(x) = \operatorname{sign}(g(x) - u(x))$ a.e. on the boundary (which is of course ensured when $g(x) = u(x)$). But it turns out that $\phi_h \cdot \nu_\Omega$ is constant on each facet $\partial T \cap \partial\Omega =: E$ of a boundary element $T \in \mathcal{T}_h$, so that the boundary integral $\int_E (g - u)\phi_h \cdot \nu_h d\mathcal{H}^{n-1}$ also vanishes if on average on E, $u = g$.

It follows that if one considers, as in Prop. 3.6, $g(x) = \chi_{\{x \cdot \vec{n} \geq t\}}$ for some $\vec{n} \in \mathbb{S}^{n-1}$ and $t \in \mathbb{R}$, and one lets $\bar{g}_h$ be the piecewise constant function on $\partial\Omega$ given, on each boundary element, by the average of g on the facet, and $g_h = \Pi_h^0(g)$, then, on the one hand, one has for any other $u \in BV(\Omega)$, considering the constant vector field $\phi \equiv \vec{n}$ and using (8):

$$TV_g(u; \Omega) \geq -\int_\Omega u \operatorname{div} \phi dx + \int_{\partial\Omega} g\phi \cdot \nu_\Omega d\mathcal{H}^{n-1} = \int_{\partial\Omega} g\phi \cdot \nu_\Omega \\ = \int_\Omega \phi \cdot Dg = \int_\Omega |Dg| = TV_g(g; \Omega)$$

that is, g is minimizing the total variation for its own boundary condition (this is of course well known!). On the other hand, since $\phi \equiv \vec{n} \in \mathrm{RT0}(\mathcal{T}_h)$ and has norm less than 1, one also has for $u_h \in \mathrm{P0}(\mathcal{T}_h)$ that:

$$TV_{\bar{g}_h, RT}(u_h; \Omega) \geq \int_{\partial\Omega} g_h \phi \cdot \nu_\Omega d\mathcal{H}^{n-1} = \int_{\partial\Omega} g\phi \cdot \nu_\Omega d\mathcal{H}^{n-1} = TV_g(g; \Omega)$$

while $TV_{\bar{g}_h, RT}(g_h; \Omega)$ is the sup over admissible ϕ_h of:

$$-\int_\Omega g_h \operatorname{div} \phi_h dx + \int_{\partial\Omega} \bar{g}_h \phi_h \cdot \nu_\Omega d\mathcal{H}^{n-1} \\ = -\int_\Omega g \operatorname{div} \phi_h dx + \int_{\partial\Omega} g\phi_h \cdot \nu_\Omega d\mathcal{H}^{n-1} = \int_\Omega \phi_h \cdot Dg$$

which is clearly reached for $\phi_h \equiv \vec{n}$ and has value $TV_h(g, \Omega)$. We have proved the following:

Proposition 4.1. *For any $\vec{n} \in \mathbb{S}^{n-1}$, $t \in \mathbb{R}$, for g, $\bar{g}_h$, g_h defined as above, then g_h is a solution of*

$$\min_{u \in L^1(\Omega)} TV_{\bar{g}_h, RT}(u; \Omega).$$

Moreover, the value of this problem is the same as the value $TV_g(g;\Omega)$ of the minimization problem $\min_{u\in BV(\Omega)} TV_g(u;\Omega)$.

Hence again, in theory, the inpainting problem should be solved perfectly, whatever the direction of the edge, by such a discretization. In practice, it seems this is not necessarily the case, as there could be again (as shown in [34] for Crouzeix-Raviart discretizations) other, more diffuse solutions to the same problem, and for strange reasons this approach is not good at recovering missing edges. This is illustrated later on (Fig. 11), for the variant we now introduce.

Cubic meshes

On the other hand, the implementation of (23) is not straightforward and optimization might be more difficult than with the previous formulations. An easier variant (which is the one actually studied in [20] and was first introduced and analysed, in a slightly different form, in [39, 40]) consists in replacing the simplicial mesh $\mathcal{T}_h$ with a cubic (square in 2D) mesh based on the pixel representation of the images.

We briefly introduce this framework, which we describe only in 2D (extension to higher dimension is obvious). Here to simplify, Ω is a rectangle (and to simplify further the square $\Omega = (0,1)^2$), we let $N \geq 1$ and $h := 1/N$. As in Section 2.1, the pixels correspond to the little squares $S^h_{i-\frac12,j-\frac12} = (ih-h, ih) \times (jh-h, jh)$ for $1 \leq i,j \leq N$, with value $u_{i,j}$. That is, the mesh $\mathcal{T}_h$ of the previous section is replaced with the cubic mesh $\{S^h_{i-\frac12,j-\frac12} : 1 \leq i,j \leq N\}$, and we can associate to the matrix $(u_{i,j})_{i,j=1}^N$ the "Q0" (the cubic equivalent of P0) piecewise constant function defined by (10). We introduce a "gradient operator" as the simple difference operator $\boldsymbol{D}u = (D^1u, D^2u)$ where:

$$\begin{cases} (D^1u)_{i+\frac12,j} = u_{i+1,j} - u_{i,j} & i = 1,\dots,N-1,\ j = 1,\dots,N\ ; \\ (D^2u)_{i,j+\frac12} = u_{i,j+1} - u_{i,j} & i = 1,\dots,N,\ j = 1,\dots,N-1\ ; \end{cases} \tag{25}$$

As before, we denote Π^0_h the projection of a function $u \in L^1(\Omega)$ onto Q0 functions, consisting in replacing in each pixel the function with its average.

We introduce also dual variables defined on the edges of the mesh, and representing average fluxes through these edges: $p^1_{i+\frac12,j}$ is an horizontal average flux "in between" $u_{i+1,j}$ and $u_{i,j}$ while $p^2_{i,j+\frac12}$ is a vertical average flux between $u_{i,j+1}$ and $u_{i,j}$. We denote $\boldsymbol{p} = (p^1_\bullet, p^2_\bullet)$ this dual variable.

This variable is naturally extended into a Raviart-Thomas vector field $p(x) =$

$(p^1(x), p^2(x))$ by letting, in each pixel $S^h_{i,j}$, $p^1(x)$ be the affine interpolation between $p^1_{i-\frac{1}{2},j}$ on $\{ih-h\}\times[jh-h, jh]$ and $p^1_{i+\frac{1}{2},j}$ on $\{ih\}\times[jh-h, jh]$, and similarly $p^2(x)$ the affine interpolation between $p^2_{i,j-\frac{1}{2}}$ on $[ih-h, ih]\times\{jh-h\}$ and $p^1_{i,j+\frac{1}{2}}$ on $[ih-h, ih]\times\{jh\}$. In case of Neumann Boundary Conditions, we let by convention the boundary fluxes ($p^1_{\frac{1}{2},j}$, etc) be zero, while in case of Dirichlet B.C. such as for the inpainting problem these values are also important and need be defined.

With this convention, one has for a function $u \in BV(\Omega)$, letting $u^h = \Pi^h_0(u) = \sum_{i,j} u^h_{i,j}\chi_{S^h_{i-\frac{1}{2},j-\frac{1}{2}}}$ and assuming the boundary fluxes vanish:

$$\begin{aligned}\int_\Omega p\cdot Du &= -\int_\Omega u \operatorname{div} p\, dx = -\int_\Omega u^h \operatorname{div} p\, dx\\ &= h^2\sum_{i,j} u^h_{i,j}\frac{p^1_{i+\frac{1}{2},j}-p^1_{i-\frac{1}{2},j}+p^2_{i,j+\frac{1}{2}}-p^2_{i,j-\frac{1}{2}}}{h}\\ &= h\sum_{i,j} p^1_{i+\frac{1}{2},j}D^1u_{i+\frac{1}{2},j}+p^2_{i,j+\frac{1}{2}}D^2u_{i,j+\frac{1}{2}} = h\langle \boldsymbol{p}, \boldsymbol{D}u\rangle \end{aligned} \tag{26}$$

The Raviart-Thomas total variation on cubic meshes is defined as in (23) by

$$TV^h_{RT}(u) = \sup\left\{\int_\Omega u \operatorname{div} p\, dx : p \text{ as above}, |p(x)| \le 1\ \forall x\in\Omega\right\}, \tag{27}$$

observe that this is well defined, again, for any $u \in L^1(\Omega)$ even if it depends only on $\Pi^0_h(u)$.

This variant on cubic meshes enjoys exactly the same properties as the previous definition based on simplicial meshes, in particular the optimal error bound in Prop. 3.4 holds with the same proof. One also easily defines, as before, a Dirichlet version of this Raviart-Thomas total variation, suitable for inpainting problem: then, Prop. 4.1 is easily seen to hold also in this case. We show in Fig. 13c (see also the detail in Fig. 14) an example of denoising with this variant, where one sees that the reconstruction is very sharp and probably closest to the continuum limit with respect to other discretizations which we introduce later on (with results displayed in Fig. 13d and 13h). Unfortunately, this is not true for the inpainting problem (*cf* "RT" row in Fig. 11), despite the theoretical result Prop. 4.1. It seems here that there are many possible solution with optimal energy and that the optimization will not necessarily pick the sharpest one.

Using (26), we see that the discrete total variation introduced in this section can also be defined by a fully discrete expression: we have

$$TV^h_{RT}(u) = \sup\left\{h\langle \boldsymbol{p}, \boldsymbol{D}u\rangle : (p^1_{i\pm\frac12,j})^2 + (p^2_{i,j\pm\frac12})^2 \le 1\ \forall i,j\right\}.$$

In this expression, the constraints correspond to four norm constraints on the vectors $(p^1_{i+\frac12,j}, p^2_{i,j+\frac12})$, $(p^1_{i-\frac12,j}, p^2_{i,j+\frac12})$, $(p^1_{i+\frac12,j}, p^2_{i,j-\frac12})$, $(p^1_{i-\frac12,j}, p^2_{i,j-\frac12})$, which are equivalent to requiring that the Raviart-Thomas extension $p(x)$ has norm less than almost everywhere in the pixel $S^h_{i-\frac12,j-\frac12}$ (as they correspond to bounding $p(x)$ at the four extremal points of $S^h_{i-\frac12,j-\frac12}$).

In order to pave the way for the following sections, let us introduce four operators $L^{\pm,\pm}$, acting on the duals and defined at each pixel (i,j), which are given by

$$(L^{\pm,\pm}\boldsymbol{p})_{i,j} = \begin{pmatrix} p^1_{i\pm\frac12,j} \\ p^2_{i,j\pm\frac12} \end{pmatrix}, \quad 1 \le i,j \le N.$$

Then the above definition takes the form

$$TV^h_{RT}(u) = \sup\left\{h\langle \boldsymbol{p}, \boldsymbol{D}u\rangle : \|L^{\pm,\pm}\boldsymbol{p}\|_{2,\infty} \le 1\right\}. \tag{28}$$

where $\|L^{\pm,\pm}\boldsymbol{p}\|_{2,\infty}$ is the maximum (over i,j) of the 2-norm of the 2-vector $(p^1_{i\pm\frac12,j}, p^2_{i,j\pm\frac12})^T$. We now introduce a variant of this definition, first appeared in [51] and further analysed and implemented in [37]

4.2 Hintermüller et al / Condat's approach

In [51, 37] a variant of (28) is proposed in the following way: using the notation of [37], one introduces the operators

$$(L^{\updownarrow}\boldsymbol{p})_{i,j} = \begin{pmatrix} p^1_{i+\frac12,j} \\ \frac14\left(p^2_{i,j-\frac12} + p^2_{i,j+\frac12} + p^2_{i+1,j-\frac12} + p^2_{i+1,j+\frac12}\right) \end{pmatrix},$$

$$(L^{\leftrightarrow}\boldsymbol{p})_{i,j} = \begin{pmatrix} \frac14\left(p^1_{i-\frac12,j} + p^1_{i+\frac12,j} + p^1_{i-\frac12,j+1} + p^1_{i+\frac12,j+1}\right) \\ p^2_{i,j+\frac12} \end{pmatrix},$$

$$(L^{\bullet}\boldsymbol{p})_{i,j} = \frac12\begin{pmatrix} p^1_{i-\frac12,j} + p^1_{i+\frac12,j} \\ p^2_{i,j-\frac12} + p^2_{i,j+\frac12} \end{pmatrix}$$

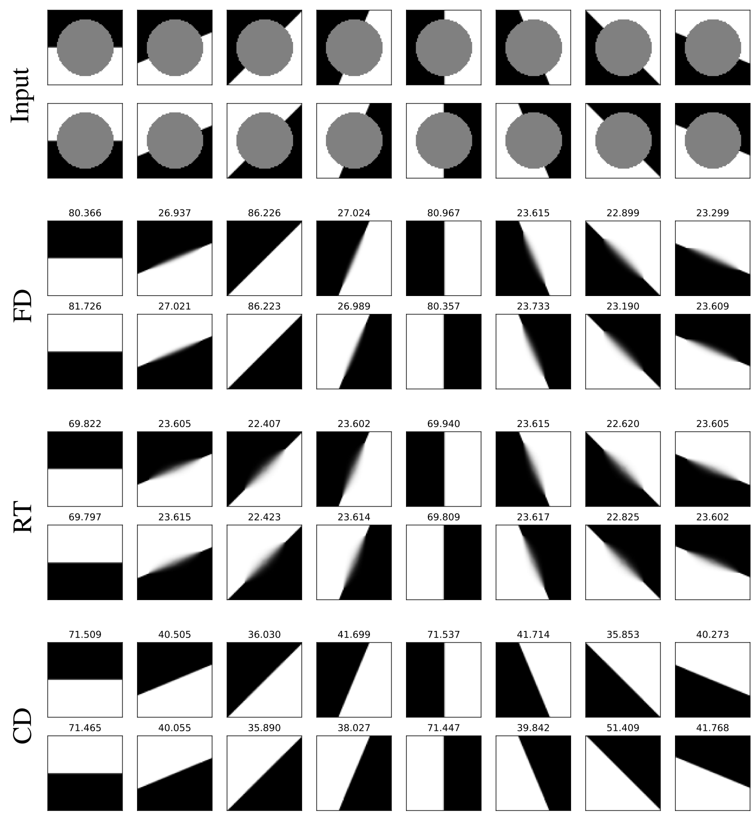


FIGURE 11 (From [35, Fig. 2]) Reconstruction quality of various discrezization schemes for the problem of inpainting a straight discontinuity (FD is (9), RT is (28), CD is (29)). The numbers above the images refer to the reconstruction error with respect to the ground truth, measured in PSNR.

and one lets

$$TV^h_{CD}(u) = \sup \left\{ h\langle \boldsymbol{p}, \boldsymbol{D}u\rangle : \|L^{\updownarrow,\leftrightarrow,\bullet}\boldsymbol{p}\|_{2,\infty} \leq 1 \right\}. \tag{29}$$

Enforcing the constraints in this way, it is not guaranteed any longer that the Raviart-Thomas interpolation of the dual field is feasible and in general, $TV^h_{CD}(u) \geq TV^h_{RT}(u)$. In particular, up to now, no error bound is known for this approximation, in addition, whether it is consistent was not even clear up to the recent work [35] (see Prop. 4.2 in Sec. 4.3). Yet, it turns out to be much more "isotropic" than any other approximation introduced so far, in particular for the inpainting problem. Numerous experiments are computed in [37] but the most striking results are the inpainting results shown in [35, Fig. 2], which demonstrate this approximation is able to recover sharp missing edges in almost any direction, see Fig. 11.

4.3 Learned dual discretizations

In [35], we have generalized the framework of the two previous sections, as we now describe. The idea is to replace the operators $L^{\leftrightarrow}$, $L^{\updownarrow}$, $L^{\bullet}$, by generic convolution operators $(L^1, \dots, L^K)$ of the form, for $k = 1, \dots, K$:

$$(L^k \boldsymbol{p})_{i,j} = \begin{pmatrix} (L^{k,1}\boldsymbol{p})_{i,j} \\ (L^{k,2}\boldsymbol{p})_{i,j} \end{pmatrix} = \begin{pmatrix} \sum_{m,n=-\nu}^{\nu} \xi^k_{m,n} p^1_{i+\frac{1}{2}-m,j-n} \\ \sum_{m,n=-\nu}^{\nu} \eta^k_{m,n} p^2_{i-m,j+\frac{1}{2}-n} \end{pmatrix}, \tag{30}$$

where $(\xi^k_{m,n}, \eta^k_{m,n})$ are weights with a small support of maximal size $(2\nu+1) \times (2\nu+1)$ for some integer $\nu \geq 0$ and which satisfy in addition:

$$\sum_{m,n} \xi^l_{m,n} = \sum_{m,n} \eta^l_{m,n} = 1. \tag{31}$$

Notice that the non-negativity of these coefficients is not required. A discrete total variation is defined by generalizing (28)–(29) as follows:

$$TV^h_L(u) = \sup\left\{ h\langle \boldsymbol{p}, \boldsymbol{D}u\rangle : \|L^k\boldsymbol{p}\|_{2,\infty} \leq 1, \quad k = 1, \dots, K \right\}.$$

We then have the following consistency result:

Proposition 4.2 ([35, Thm. 2.4])**.** *As $h \to 0$, the discrete total variations TV^h_L Γ-converge to*

$$TV(u;\Omega) = \begin{cases} |Du|(\Omega) & \text{if } u \in BV(\Omega), \\ +\infty & \text{else.} \end{cases}$$

We refer to [35] for a more precise statement, in particular concerning the boundary conditions. In addition to Proposition 4.2, a compactness result for sequences bounded in energy is also shown.

Now, what is interesting in this framework is that all coefficients which satisfy (31) allow to define a consistent total variation. This means that given a specific task, such as inpainting, one can try to "learn" the *best* coefficients, which solve this task best. This is a bi-level optimization program, which is hard to solve, but is stated in relatively low dimension (as the parameters which are learned are just the coefficients of the convolution operators L^k).

FD	RT	CD

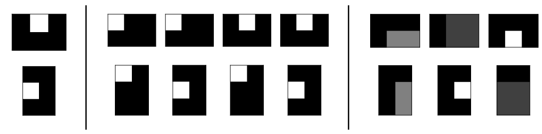


TABLE 1 Filters corresponding to respectively definitions (9), (28), and (29) of a discrete total variation, with which the inpainting results of Fig. 11 have been computed.

$K = 2$	$K = 2$ (sym)	$K = 8$ (sym)

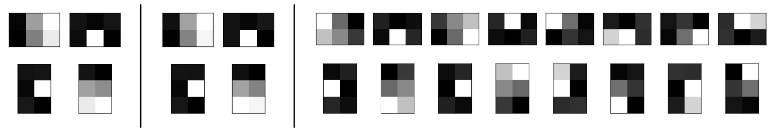


TABLE 2 Learned filters for the inpainting problem. The filters may contain negative values but are normalized to the interval $[0, 1]$. In the second column, a symmetry by transposition was enforced, while in the third a symmetry by 90° rotation was enforced.

As an illustration, we show in Table 2 pair of filters learned to best solve the inpainting task of Fig. 11. In these experiments, we consider small filters: the horizontal one (coefficients $\xi_\bullet$ in (30)) has 2×3 components and the vertical one (coefficients $\eta_\bullet$), 3×2. In this setting, the filters corresponding to Fig. 11 are shown in Table 1. The result produced by the learned filters are much sharper than with the ad hoc discretizations, see Fig. 12.

In Figures 13–14 we show the ROF denoising (Problem (4)) of the image in Fig. 13a, with various choices of the discrete total variation. Figs. 13b–13e show the results with the handcrafted discretizations discussed so far. In the three last figures (13f–13h), the discrete total variation is TV_L^h, with various strategies for learning the filters. In Fig. 13f, the filters are learned by trying to best recover orignal images from their noisy version. Observe that it produces visually the best reconstruction, with sharp edges but a result which does not look too artificial and almost no visible influence of the underlying pixels (the quality of Fig 13d is almost as good, but staircasing is reduced in 13f without altering the important details, see in particular the detail in Fig. 14). The filters in Fig. 13g are learned by trying to best recover the exact minimizer of (4) for g the characteristic of disks of various radii. The edges are sharper than in the previous, but the result is a bit less pleasing. In the last experiment, the filters are learned for inpainting (*cf.* Table 2, $K = 8$). The result for denoising is almost

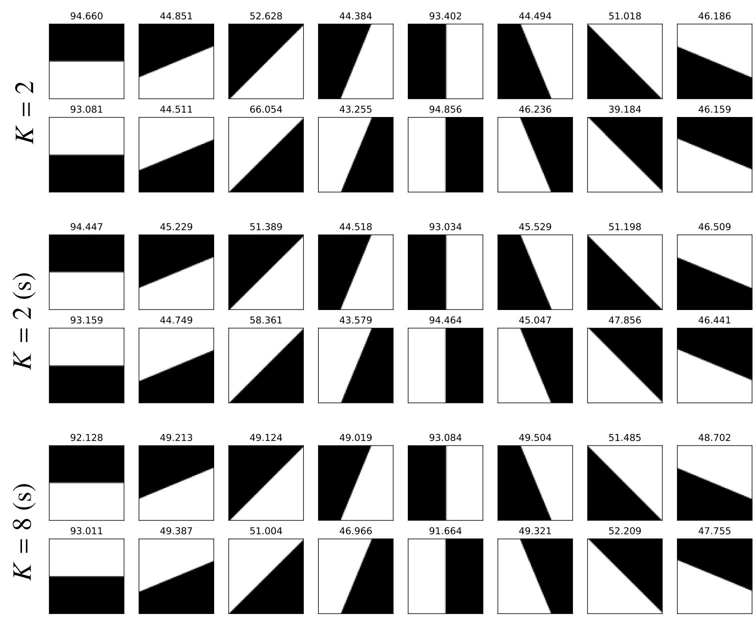


FIGURE 12 Reconstruction quality for the filters in Table 2. The results are much sharper, in all directions, than in Fig. 11, and the PSNR values much higher.

the same as in the previous figure, or with the Raviart-Thomas total variation (Fig. 13c), with very sharp edges and crisp details, maybe too aggressive for a natural image. See also Fig. 14 which shows an enlarged detail from Fig. 13. In [35], many more experiments are carried on and the process for learning filters for denoising synthetic or true images is explained with further details, still in the context of consistent discrete total variations. On the other hand, clearly, proving numerical errors for these general filters remains an open question.

5 CONCLUSION AND PERSPECTIVES

In this note, we have reviewed various approaches to consistently approximate, in an implementable way, the isotropic total variation. We have shown that different points of view can lead to qualitatively very different results, and sometimes provable optimal or non-optimal error estimates. In addition, the choice of a discretization should probably depend on the particular task the functional is used for. These ideas and approaches should be kept in mind when trying

(a) noisy **(b)** forward diff. **(c)** RT **(d)** Condat

(e) "Shannon" **(f)** learned (natural images) **(g)** learned (disk denoising) **(h)** learned (inpainting)

FIGURE 13 Original noisy image (Fig. 13a), and denoised variants.

to approximate similar but more complex variants of the total variation, such as non-homogeneous versions (or total variation on a Riemannian manifold), or non-local or singular variants (as for instance in [66], which could most probably be drastically improved).

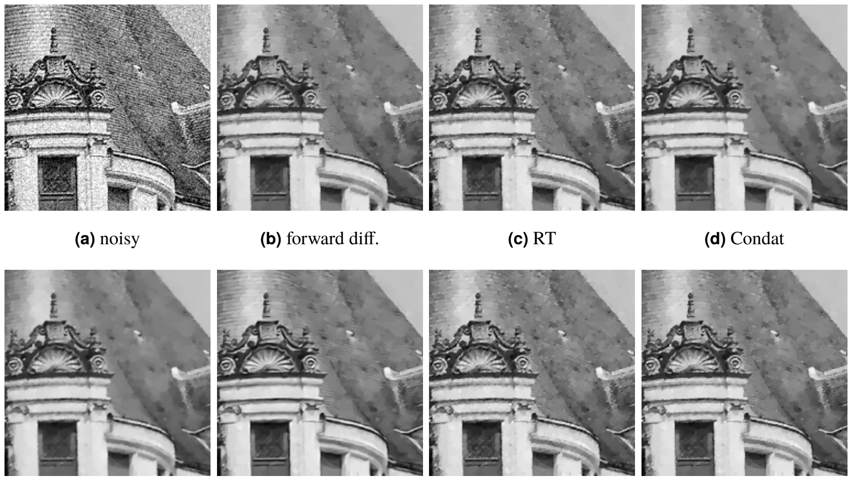
ACKNOWLEDGEMENTS

The authors would like to thank the Isaac Newton Institute for Mathematical Sciences, Cambridge, for its support and hospitality during the programme "Variational methods, new optimisation techniques and new fast numerical algorithms" (Sept.-Oct., 2017), when this work was undertaken. This work was supported by: EPSRC Grant N. EP/K032208/1. The work of A.C. was partially supported by a grant of the Simons Foundation. T.P. acknowledges support by the ERC starting grant HOMOVIS, No. 640156.

BIBLIOGRAPHY

[1] Rémy Abergel and Lionel Moisan. The Shannon Total Variation. *J. Math. Imaging Vision*, 59(2):341–370, 2017.

[2] Luigi Ambrosio, Nicola Fusco, and Diego Pallara. *Functions of bounded variation and free*



(a) noisy **(b)** forward diff. **(c)** RT **(d)** Condat

(e) “Shannon” **(f)** learned (natural images) **(g)** learned (disk denoising) **(h)** learned (inpainting)

FIGURE 14 Details of Fig. 13

discontinuity problems. Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press, New York, 2000.

[3] G. Anzellotti. Pairings between measures and bounded functions and compensated compactness. *Ann. Mat. Pura Appl. (4)*, 135:293–318, 1983.

[4] Sören Bartels. Total variation minimization with finite elements: convergence and iterative solution. *SIAM J. Numer. Anal.*, 50(3):1162–1180, 2012.

[5] Sören Bartels. Error control and adaptivity for a variational model problem defined on functions of bounded variation. *Math. Comp.*, 84(293):1217–1240, 2015.

[6] Sören Bartels. Error estimates for a class of discontinuous Galerkin methods for nonsmooth problems via convex duality relations, 2020. preprint arXiv:2004.09196.

[7] Sören Bartels. Nonconforming discretizations of convex minimization problems and precise relations to mixed methods, 2020. preprint arXiv:2002.02359.

[8] Sören Bartels and Marijo Milicevic. Stability and experimental comparison of prototypical iterative schemes for total variation regularized problems. *Comput. Methods Appl. Math.*, 16(3):361–388, 2016.

[9] Sören Bartels and Marijo Milicevic. Primal-dual gap estimators for $a posteriori$ error analysis of nonsmooth minimization problems. *ESAIM Math. Model. Numer. Anal.*, 54(5):1635–1660, 2020.

[10] Sören Bartels, Ricardo H. Nochetto, and Abner J. Salgado. Discrete total variation flows without regularization. *SIAM J. Numer. Anal.*, 52(1):363–385, 2014.

[11] Sören Bartels, Ricardo H. Nochetto, and Abner J. Salgado. A total variation diminishing

interpolation operator and applications. *Math. Comp.*, 84(296):2569–2587, 2015.

[12] Sören Bartels and Zhangxian Wang. Orthogonality relations of Crouzeix-Raviart and Raviart-Thomas finite element spaces, 2020. preprint arXiv:2005.02741.

[13] Amir Beck and Marc Teboulle. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. *SIAM J. Imaging Sci.*, 2(1):183–202, 2009.

[14] Yuri Boykov and Vladimir Kolmogorov. Computing geodesics and minimal surfaces via graph cuts. In *Proceedings Ninth IEEE International Conference on Computer Vision*, pages 26–33 vol.1, Oct 2003.

[15] Yuri Boykov, Olga Veksler, and Ramin Zabih. Fast approximate energy minimization via graph cuts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 23(11):1222–1239, 2001.

[16] Andrea Braides. *Γ-convergence for beginners*, volume 22 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford, 2002.

[17] Kristian Bredies, Thomas Pock, and Benedikt Wirth. A convex, lower semicontinuous approximation of Euler's elastica energy. *SIAM J. Math. Anal.*, 47(1):566–613, 2015.

[18] Susanne C. Brenner. Forty years of the Crouzeix-Raviart element. *Numer. Methods Partial Differential Equations*, 31(2):367–396, 2015.

[19] Jian-Feng Cai, Bin Dong, Stanley Osher, and Zuowei Shen. Image restoration: total variation, wavelet frames, and beyond. *J. Amer. Math. Soc.*, 25(4):1033–1089, 2012.

[20] Corentin Caillaud and Antonin Chambolle. Error estimates for finite differences approximations of the total variation. preprint hal-02539136, April 2020.

[21] Carsten Carstensen and DongJie Liu. Nonconforming FEMs for an optimal design problem. *SIAM J. Numer. Anal.*, 53(2):874–894, 2015.

[22] Vicent Caselles, Antonin Chambolle, and Matteo Novaga. The discontinuity set of solutions of the TV denoising problem and some extensions. *Multiscale Model. Simul.*, 6(3):879–894, 2007.

[23] Vicent Caselles, Antonin Chambolle, and Matteo Novaga. Total variation in imaging. In *Handbook of mathematical methods in imaging. Vol. 1, 2, 3*, pages 1455–1499. Springer, New York, 2015.

[24] Vicent Caselles, Khalid Jalalzai, and Matteo Novaga. On the jump set of solutions of the total variation flow. *Rend. Semin. Mat. Univ. Padova*, 130:155–168, 2013.

[25] Antonin Chambolle, Vicent Caselles, Daniel Cremers, Matteo Novaga, and Thomas Pock. An introduction to total variation for image analysis. In *Theoretical foundations and numerical methods for sparse recovery*, volume 9 of *Radon Ser. Comput. Appl. Math.*, pages 263–340. Walter de Gruyter, Berlin, 2010.

[26] Antonin Chambolle and Jérôme Darbon. *Image Processing and Analysis with Graphs: Theory and Practice*, chapter "A Parametric Maximum Flow Approach for Discrete Total Variation Regularization". CRC Press, 2012.

[27] Antonin Chambolle, Michael Goldman, and Matteo Novaga. Fine properties of the subdifferential for a class of one-homogeneous functionals. *Adv. Calc. Var.*, 8(1):31–42, 2015.

[28] Antonin Chambolle, Stacey E. Levine, and Bradley J. Lucier. An upwind finite-difference method for total variation-based image smoothing. *SIAM J. Imaging Sci.*, 4(1):277–299, 2011.

[29] Antonin Chambolle and Pierre-Louis Lions. Image recovery via total variation minimization and related problems. *Numer. Math.*, 76(2):167–188, 1997.

[30] Antonin Chambolle and Thomas Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *J. Math. Imaging Vision*, 40(1):120–145, 2011.

[31] Antonin Chambolle and Thomas Pock. A remark on accelerated block coordinate descent for computing the proximity operators of a sum of convex functions. *SMAI J. Comput. Math.*, 1:29–54, 2015.

[32] Antonin Chambolle and Thomas Pock. An introduction to continuous optimization for imaging. *Acta Numer.*, 25:161–319, 2016.

[33] Antonin Chambolle and Thomas Pock. Total roto-translational variation. *Numer. Math.*, 142(3):611–666, 2019.

[34] Antonin Chambolle and Thomas Pock. Crouzeix-Raviart approximation of the total variation on simplicial meshes. *J. Math. Imaging Vision*, 2020. (online).

[35] Antonin Chambolle and Thomas Pock. Learning Consistent Discretizations of the Total Variation. Preprint hal-02982082, October 2020.

[36] Antonin Chambolle, Pauline Tan, and Samuel Vaiter. Accelerated alternating descent methods for Dykstra-like problems. *J. Math. Imaging Vision*, 59(3):481–497, 2017.

[37] Laurent Condat. Discrete total variation: new definition and minimization. *SIAM J. Imaging Sci.*, 10(3):1258–1290, 2017.

[38] Gianni Dal Maso. *An introduction to* Γ*-convergence*, volume 8 of *Progress in Nonlinear Differential Equations and their Applications*. Birkhäuser Boston, Inc., Boston, MA, 1993.

[39] Philippe Destuynder, Mohamed Jaoua, and Hela Sellami. A dual algorithm for denoising and preserving edges in image processing. *J. Inverse Ill-Posed Probl.*, 15(2):149–165, 2007.

[40] Philippe Destuynder, Mohamed Jaoua, and Hela Sellami. An error estimate in image processing. *ARIMA Rev. Afr. Rech. Inform. Math. Appl.*, 15:61–81, 2012.

[41] Daniele A. Di Pietro and Simon Lemaire. An extension of the Crouzeix-Raviart space to general meshes with application to quasi-incompressible linear elasticity and Stokes flow. *Math. Comp.*, 84(291):1–31, 2015.

[42] Joan Duran, Michael Moeller, Catalina Sbert, and Daniel Cremers. Collaborative total variation: a general framework for vectorial TV models. *SIAM J. Imaging Sci.*, 9(1):116–151, 2016.

[43] Ernie Esser, Xiaoqun Zhang, and Tony F. Chan. A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. *SIAM J. Imaging Sci.*, 3(4):1015–1046, 2010.

[44] Lawrence C. Evans and Ronald F. Gariepy. *Measure theory and fine properties of functions*. Textbooks in Mathematics. CRC Press, Boca Raton, FL, revised from 1992 edition, 2015.

[45] Herbert Federer. *Geometric measure theory*. Die Grundlehren der mathematischen Wissenschaften, Band 153. Springer-Verlag New York Inc., New York, 1969.

[46] Enrico Giusti. *Minimal surfaces and functions of bounded variation*, volume 80 of *Monographs in Mathematics*. Birkhäuser Verlag, Basel, 1984.

[47] Bastian Goldluecke, Evgeny Strekalovskiy, and Daniel Cremers. The natural vectorial total variation which arises from geometric measure theory. *SIAM J. Imaging Sci.*, 5(2):537–564, 2012.

[48] Frédéric Hecht. New development in freefem++. *J. Numer. Math.*, 20(3-4):251–265, 2012.

[49] Duvan Henao, Carlos Mora-Corral, and Xianmin Xu. A numerical study of void coalescence and fracture in nonlinear elasticity. *Comput. Methods Appl. Mech. Engrg.*, 303:163–184, 2016.

[50] Marc Herrmann, Roland Herzog, Stephan Schmidt, José Vidal-Núñez, and Gerd Wachsmuth. Discrete total variation with finite elements and applications to imaging. *Journal of Mathematical Imaging and Vision*, 61(4):411–431, 2019.

[51] Michael Hintermüller, Carlos N. Rautenberg, and Jooyoung Hahn. Functional-analytic and numerical issues in splitting methods for total variation-based image reconstruction. *Inverse Problems*, 30(5):055014, 34, 2014.

[52] Dorit S. Hochbaum. An efficient algorithm for image segmentation, Markov random fields and related problems. *J. ACM*, 48(4):686–701 (electronic), 2001.

[53] Maria Klodt, Thomas Schoenemann, Kalin Kolev, Marek Schikora, and Daniel Cremers. An experimental comparison of discrete and continuous shape optimization methods. In David Forsyth, Philip Torr, and Andrew Zisserman, editors, *Computer Vision – ECCV 2008*, pages 332–345, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg.

[54] Ming-Jun Lai, Bradley Lucier, and Jingyue Wang. *The Convergence of a Central-Difference Discretization of Rudin-Osher-Fatemi Model for Image Denoising*, pages 514–526. Springer Berlin Heidelberg, Berlin, Heidelberg, 2009.

[55] Ming-Jun Lai and Leopold Matamba Messi. Piecewise linear approximation of the continuous Rudin-Osher-Fatemi model for image denoising. *SIAM J. Numer. Anal.*, 50(5):2446–2466, 2012.

[56] Jan Lellmann, Björn Lellmann, Florian Widmann, and Christoph Schnörr. Discrete and continuous models for partitioning problems. *Int. J. Comput. Vis.*, 104(3):241–269, 2013.

[57] Simon Masnou and Jean-Michel Morel. On a variational theory of image amodal completion. *Rend. Sem. Mat. Univ. Padova*, 116:211–252, 2006.

[58] Yves Meyer. *Oscillating patterns in image processing and nonlinear evolution equations*, volume 22 of *University Lecture Series*. American Mathematical Society, Providence, RI, 2001. The fifteenth Dean Jacqueline B. Lewis memorial lectures.

[59] Lionel Moisan. How to discretize the total variation of an image? *PAMM*, 7(1):1041907–1041908, 2007.

[60] Yurii Nesterov. *Introductory lectures on convex optimization*, volume 87 of *Applied Optimization*. Kluwer Academic Publishers, Boston, MA, 2004. A basic course.

[61] Yurii Nesterov. Smooth minimization of non-smooth functions. *Math. Program.*, 103(1):127–152, 2005.

[62] Yurii Nesterov. Gradient methods for minimizing composite functions. *Math. Program.*, 140(1,

Ser. B):125–161, 2013.

[63] Mark Nitzberg, David Mumford, and Takahiro Shiota. *Filtering, segmentation and depth*, volume 662 of *Lecture Notes in Computer Science*. Springer-Verlag, Berlin, 1993.

[64] Christoph Ortner. Nonconforming finite-element discretization of convex variational problems. *IMA J. Numer. Anal.*, 31(3):847–864, 2011.

[65] Christoph Ortner and Dirk Praetorius. On the convergence of adaptive nonconforming finite element methods for a class of convex variational problems. *SIAM J. Numer. Anal.*, 49(1):346–367, 2011.

[66] Thomas Pock, Daniel Cremers, Horst Bischof, and Antonin Chambolle. Global solutions of variational models with convex regularization. *SIAM J. Imaging Sci.*, 3(4):1122–1145, 2010.

[67] Sergey I. Repin. A variation-difference method for solving problems with functionals of linear growth. *Zh. Vychisl. Mat. i Mat. Fiz.*, 29(5):693–708, 798, 1989.

[68] Leonid Rudin, Stanley J. Osher, and Elad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992. [also in *Experimental Mathematics: Computational Issues in Nonlinear Science* (Proc. Los Alamos Conf. 1991)].

[69] Wenyi Tian and Xiaoming Yuan. Linearized primal-dual methods for linear inverse problems with total variation regularization and finite element discretization. *Inverse Problems*, 32(11):115011, 32, 2016.

[70] Fabio Viola, Andrew W. Fitzgibbon, and Roberto Cipolla. A unifying resolution-independent formulation for early vision. In *2012 IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, June 16-21, 2012*, pages 494–501, 2012.

[71] Jingyue Wang and Bradley J. Lucier. Error bounds for finite-difference methods for Rudin-Osher-Fatemi image smoothing. *SIAM J. Numer. Anal.*, 49(2):845–868, 2011.

[72] Xianmin Xu and Duvan Henao. An efficient numerical method for cavitation in nonlinear elasticity. *Math. Models Methods Appl. Sci.*, 21(8):1733–1760, 2011.

[73] Mingqiang Zhu and Tony. F. Chan. An efficient primal-dual hybrid gradient algorithm for total variation image restoration. CAM Reports 08-34, UCLA, Center for Applied Math., 2008.

[74] William P. Ziemer. *Weakly differentiable functions*. Springer-Verlag, New York, 1989. Sobolev spaces and functions of bounded variation.