



HAL
open science

Crouzeix-Raviart approximation of the total variation on simplicial meshes

Antonin Chambolle, Thomas Pock

► **To cite this version:**

Antonin Chambolle, Thomas Pock. Crouzeix-Raviart approximation of the total variation on simplicial meshes. *Journal of Mathematical Imaging and Vision*, 2020, 62 (6-7), pp.872–899. 10.1007/s10851-019-00939-3 . hal-01787012v3

HAL Id: hal-01787012

<https://hal.science/hal-01787012v3>

Submitted on 31 Mar 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Crouzeix-Raviart approximation of the total variation on simplicial meshes.*

Antonin Chambolle[†] Thomas Pock[‡]

March 31, 2020

We propose an adaptive implementation of a Crouzeix-Raviart based discretization of the total variation, which has the property of approximating from below the total variation, with metrication errors only depending on the local curvature, rather than on the orientation as is usual for other approaches.

keywords: Image processing, total variation, nonconforming P1 (Crouzeix-Raviart) finite elements, error estimates.

AMS MSC (2010): 49Q20 26A45 65K10 65N30

1. Introduction

Since [46], the total variation has been widely used in imaging as a basic denoising tool and a regularizer for inverse problems, obviously as it is one of the few convex regularizing energies which preserve discontinuities [18]. Classically, it is defined for a (here, to simplify, scalar) function $u : \Omega \rightarrow \mathbb{R}$ defined on a domain $\Omega \subset \mathbb{R}^d$ ($d = 2$ or 3 for most imaging applications) as the (total) variation of the distributional derivative, Du , which is assumed to be a measure. An equivalent definition, by duality, is simply

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

see for instance [3], and a function $u \in L^1(\Omega)$ is said to have bounded variation if and only if this quantity is finite. The space of such functions is usually denoted $BV(\Omega)$.

*Submitted to the editors DATE. This work was funded by EPSRC Grant N. EP/K032208/1 (Isaac Newton Institute, Cambridge), the Simons Foundation, FWF project EANOI, No. I1148, ERC grant HOMOVIS, No. 640156.

[†]CMAP, École Polytechnique, CNRS, 91128 Palaiseau, France
(antonin.chambolle@cmap.polytechnique.fr).

[‡]Institute of Computer Graphics and Vision, Graz University of Technology, 8010 Graz, Austria
(pock@icg.tugraz.at).

Clearly, such a definition allows for discontinuous functions, as it is obvious that the characteristic function of a sufficiently regular set is a BV function.

For this reason, it is not trivial to correctly discretize the functional $|Du|(\Omega)$. An obvious discretization, assuming to simplify $\Omega = [0, 1]^2$, N is an integer, $h = 1/N > 0$ would be, given the discrete function $u^h = (u_{i,j}^h)_{0 \leq i,j \leq N}$,

$$J_h(u^h) = h \sum_{i < N, j < N} \sqrt{(u_{i+1,j}^h - u_{i,j}^h)^2 + (u_{i,j+1}^h - u_{i,j}^h)^2} + h \sum_{i=1}^{N-1} |u_{i+1,N}^h - u_{i,N}^h| + h \sum_{j=1}^{N-1} |u_{N,j+1}^h - u_{N,j}^h|. \quad (2)$$

The scaling h is the volume of the elementary “pixel” h^d (here $d = 2$) divided by the scale h which appears in the discrete gradients. Naively, this discretization appears to be wrong, as for instance the measure of a characteristic function ($u_{i,j}^h \in \{0, 1\}$) will always be exaggerated, more or less strongly depending on the orientation of the boundary, see Fig. 1.

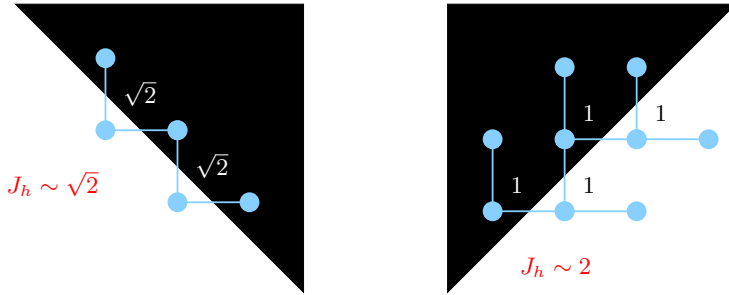


Figure 1: How J_h would measure a sharp slanted edge in a 1×1 square: on the left, we get the correct measure, on the right the result is biased by a factor $\sqrt{2}$ (the numbers show the contribution of a triple of pixels (i, j) , $(i + 1, j)$, $(i, j + 1)$ to the energy), while a smoothed edge in the same direction can have a much lower energy. This explains why minimizers of this discrete approximation will have sharp edges in some orientation, and smoother edges in other.

However, this analysis is not really correct: as we are in practice interested in minimization problems involving J_h , the right question should be whether minimizers of such problems will approximate correctly the minimizer of some related continuous problem. This is expressed by the fact that J_h , as $h \rightarrow 0$, actually “ Γ -converges” [12, 29] to $|Du|(\Omega)$ (the proof is more or less trivial, for instance see [15, 26] for examples of similar proofs for variants of J_h). But this result suffers from a slight drawback: since the measure of characteristic functions with J_h is wrong, it means that in this process, such

a function in the limit will *not*, in general, be approximated by discrete characteristic functions. In practice, it is easy to see that it suffices to slightly smooth the limiting function u to obtain after discretization an image u^h with $J_h(u^h) \approx |Du|(\Omega)$. (Rigorous error estimates can be found for instance in [47, 40, 39], see also some references below for P1 finite elements error estimates.) This means that sharp discontinuities will exceptionally appear in minimizers of energies involving J_h (see [27] for many interesting examples). In a work in preparation [16], we will show that the order of approximation, as $h \rightarrow 0$, for a denoising problem based on the discretization J_h is actually worse when denoising a slanted edge oriented as in Fig. 1, right, than for an edge oriented as on the left ($O(h^{2/3})$ vs. $O(h)$ for the energy).

This issue has been addressed many times in the imaging literature, in many contexts: finite differences [21, 24, 27], graph-based total variations [45, 11] (these references can by no means be exhaustive), more complex grids [38]... Others have advocated for finite elements discretizations [33, 34, 32], obtaining in particular very good results with adaptive meshes [5, 7, 6, 8, 9]. An obvious issue with P1 elements is the following: again, if one wants to approximate a discontinuous function, then the gradient in some elements should be very large, and its orientation will be mostly determined by the directions of the edges of the element. This leads again to an exaggeration of the total variation of characteristic functions, and as a result, to a smoothing of the discrete variational solutions, unless adaption is (well) implemented. It is difficult to expect this to improve using higher order approximations [37], as difficulties precisely arise when one needs to approximate discontinuous functions. One direction to improve this is suggested in [44], which proposes to use discontinuous P1 finite elements. There, again, the discretization is conforming in the sense that the jump energy is actually taken into account in the discretized functional. A notable interest here should be a much more precise discretization of the discontinuities, however, once again, in case the edges of the mesh are not parallel to a jump, its measure will be exaggerated and a smoothing will be necessary in order to better approximate the energy.

In this paper, we discuss the merits of a finite elements discretization of the total variation based on nonconforming P1 elements, also known as “Crouzeix-Raviart” finite elements since they were introduced in [28] (see also [13]). Contrarily to [44], the discretizations we consider will be truly nonconforming, in the sense that the jump part across edges of the mesh are not part of the energy.

It has been observed many times that such elements can be useful to discretize some nonlinear variational problems such as Cavitation and Fracture [35, 48], Nonlinear Elasticity and Stokes’s equation [31] (where it is extended to non-triangular meshes, which should also be interesting for total variation-based energies), Nonlinear Elasticity with mesh adaption [42, 41], Topology Optimization [17] — see [13] for a general overview of the use of this discretization over the years.

It turns out that indeed these elements enjoy some very nice properties: in particular, the direction of the gradient in an element is entirely free and not determined by the shape of the element itself. This makes them particularly desirable to approximate functions with discontinuities, without altering the total measure of the singularity (see the quite elementary Prop 3.3 below). On the other hand, minimizers of a Crouzeix-

Raviart-based total variation can be quite diffusive, for reasons different than the ones pointed out before, so that the need for adaptivity is not totally eliminated by this choice. We can propose, on square grids, a particular adaptive strategy which automatically derives the best way to cut elementary squares into two triangles in order to find the best approximation of a given image, from the point of view of diffusivity.

Throughout the paper we will mostly focus on the following “denoising” problem [46]:

$$\min_{u \in BV(\Omega)} |Du|(\Omega) + \frac{1}{2\tau} \int_{\Omega} (u - g)^2 dx. \quad (3)$$

Here, $g \in L^2(\Omega)$ is the original signal (in [46], a noisy image), u its regularized version, and $\tau > 0$ a parameter (which obviously controls the degree of smoothing).

The main reason for focusing on this problem 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 minimization algorithms involving the same functional (see [25] for an overview). 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 (and simple variants of the first).

The paper is organized as follows: in the next Section 2 we discuss the issues of approximating the gradient of BV functions and introduce the Crouzeix-Raviart finite elements, discussing their most useful properties in our context. Then, in Section 3, we define the Crouzeix-Raviart total variation and analyse some properties. We show in particular that straight lines (more precisely, step functions with straight jump) are measured perfectly by this energy, independently of the mesh. On the other hand, we can also build diffusive approximations to step functions. In Section 4 we introduce a simple adaptive strategy, in 2D, to overcome this issue. We discuss error bounds for the minimization of (3) in Section 5; then we show numerical experiments, and comparison with other discretizations, in Section 6. Eventually, we propose in Section B a variant of our adaptive Crouzeix-Raviart discretization for 2D images, which has only one degree of freedom per pixel.

Acknowledgements

The authors would like to thank the Isaac Newton Institute for Mathematical Sciences, Cambridge, for support and hospitality during the programme “Variational methods, new optimisation techniques and new fast numerical algorithms” (Sept.-Oct., 2017), when this work was undertaken. It has benefitted the support of the 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 Austrian science fund (FWF) under the project EANOI, No. I1148 and the ERC starting grant HOMOVIS, No. 640156. The authors were sharing Mila Nikolova’s office at the Isaac Newton Institute and were very happy to enjoy many interesting discussions with her on this project. We dedicate this paper to her memory.

2. Crouzeix-Raviart approximation of BV functions

2.1. Discretization of the gradient of a BV function u

Consider $u \in BV(\Omega)$, $\Omega \subset \mathbb{R}^2$, a function with bounded variation and $\mathcal{T}^h = \{T_i : i = 1, \dots, N\}$ a triangular mesh¹ of a polygonal approximation $\Omega^h \subset \Omega$ of Ω . Here $h > 0$, is the maximal size of an edge of a triangle T_i , $i = 1, \dots, N$, moreover one assumes $\text{dist}(\Omega^h, \mathbb{R}^2 \setminus \Omega) \leq h$. Let for each triangle T_i ,

$$\mathbf{p}_i := \frac{1}{|T_i|} \int_{T_i} Du. \quad (4)$$

Remark 2.1. *Since Du is a measure, unless $|Du|(\bigcup_{i=1}^N (\partial T_i)) = 0$ one needs to be more precise when defining \mathbf{p}_i as above. Our analysis will be rigorous for general BV functions provided we first arbitrarily assign each facet of the T_i to one of the adjacent elements it belongs to: that is, for instance, replace T_1 with \bar{T}_1 , T_2 with $\bar{T}_2 \setminus \bar{T}_1$, T_i with $\bar{T}_i \setminus \bigcup_{j < i} \bar{T}_j$, etc. In this case, given a facet $F = \partial T \cap \partial T'$ and assuming $F \subset T$, the restriction $u|_F$ should be understood as the trace of $u|_T$, which might differ from the trace of $u|_{T'}$ in case $|Du|(F) \neq 0$. To simplify, we will not stress this point in each of our statements, however this is how they should be understood in general.*

Based on the vectors \mathbf{p}_i , we define, for $x \in \Omega^h$ the piecewise constant function

$$\mathbf{p}^h(x) = \sum_i \mathbf{p}_i \chi_{T_i}(x), \quad (5)$$

where χ_{T_i} denotes the characteristic function of the triangle T_i . By construction, obviously (thanks to Jensen's inequality),

$$\int_{\Omega^h} |\mathbf{p}^h(x)| dx \leq |Du|(\Omega^h) \leq |Du|(\Omega). \quad (6)$$

More precisely, if we introduce $\nu^h(x) = \mathbf{p}^h(x)/|\mathbf{p}^h(x)|$ (if $\mathbf{p}_i = 0$ for some triangle T_i , we can either choose arbitrarily $\nu^h(x)$ in T_i , or let $\nu^h(x) = 0$), we can derive the following estimate:

Lemma 2.2. *Let $u \in BV(\Omega^h)$ be a function with bounded variation and let $\mathbf{p}^h(x)$ be defined as in (5). Then,*

$$|Du|(\Omega^h) = \int_{\Omega^h} |\mathbf{p}^h(x)| dx + \frac{1}{2} \int_{\Omega^h} |\nu_u(x) - \nu^h(x)|^2 |Du|. \quad (7)$$

¹For simplicity, most of our results are stated in two dimensions, however unless otherwise specified, they trivially extend to higher dimension up to Section 3.3 (included) and in Section 5, replacing triangles by simplices, etc., and possibly changing some constants as in Remark 2.3.

Proof. By direct calculation,

$$\begin{aligned}
|Du|(\Omega^h) &= \int_{\Omega^h} |Du| = \int_{\Omega^h} \nu_u(x) \cdot Du \\
&= \sum_{i=1}^N \int_{T_i} \nu^h(x) \cdot Du + \int_{\Omega^h} (\nu_u(x) - \nu^h(x)) \cdot Du \\
&= \sum_{i=1}^N |\mathbf{p}_i| |T_i| + \int_{\Omega^h} |Du| - \nu^h(x) \cdot \nu_u(x) |Du| \\
&= \int_{\Omega^h} |\mathbf{p}^h(x)| dx + \int_{\Omega^h} (1 - \nu^h(x) \cdot \nu_u(x)) |Du| \\
&= \int_{\Omega^h} |\mathbf{p}^h(x)| dx + \frac{1}{2} \int_{\Omega^h} (|\nu_u(x)|^2 + |\nu^h(x)|^2 - 2\nu^h(x) \cdot \nu_u(x)) |Du|,
\end{aligned}$$

and the final estimate follows. \square

The error estimate shows that the approximation will be better if the gradient direction $\nu_u = Du/|Du|$ does not oscillate too much in each triangle. In particular if u is the characteristic of a half-space, then the value of the discrete variation is exact.

Remark 2.3. *Interestingly, it is easy to deduce that if $u = \chi_E$ is the characteristic function of a $C^{1,1}$ set E which satisfies both a R -inner and outer ball condition² everywhere on ∂E , then if $h \leq R$,*

$$\left(1 - \frac{\pi^2}{18} \left(\frac{h}{R}\right)^2\right) |D\chi_E(\Omega^h)| \leq \int_{\Omega^h} |\mathbf{p}^h(x)| dx \leq |D\chi_E(\Omega^h)|. \quad (8)$$

Indeed, if $h \leq R$ then, thanks to the ball condition, for any triangle $T \in \mathcal{T}^h$, $\partial E \cap T$ is the intersection of T with a small piece of $C^{1,1}$ curve of length at most $\pi h/3$. Denoting by θ the angle of the normal vector of this curve with respect to e_1 and $\bar{\theta}$ the angle of ν^h , obviously $|\nu(x) - \nu^h| \leq |\theta(x) - \bar{\theta}|$. Observing that there must be a point on the piece of curve where $\nu^h = \nu(\bar{x})$ (as ν^h is in the cone generated by $\nu(x)$ for x on the curve), one has (using that θ is $(1/R)$ -Lipschitz, and denoting dist the distance along the curve)

$$|\nu(x) - \nu^h| \leq |\theta(x) - \bar{\theta}| \leq \frac{1}{R} \text{dist}(x, \bar{x}) \leq \frac{\pi h}{3R}.$$

Hence,

$$\int_T |\nu(x) - \nu^h|^2 |Du| \leq \frac{\pi^2 h^2}{9R^2} \mathcal{H}^1(\partial E \cap T)$$

and (8) follows from (7). (Under the very strong inner and outer ball conditions, this can be extended to any dimension, with a different constant.) A natural question is whether a similar estimate would hold, in $2D$, for $u \in BV(\Omega)$ such that there exists $z \in L^\infty(\Omega; B(0, 1))$ (or maybe continuous) with $\text{div } z \in L^\infty(\Omega)$ and $z \cdot Du = |Du|$.

²that is, for any $x \in \partial E$, there are balls $B(y, R) \subset E$, $B(z, R) \subset E^c$, with $\{x\} = \partial B(y, R) \cap \partial B(z, R)$; in particular $|\kappa_{\partial E}| \leq 1/R$.

2.2. Approximation of u

Now an important question is whether \mathbf{p}^h can be considered as the discrete gradient of some discretized function u^h . It is in fact well known that it is the case, with the function u^h defined in a space of nonconforming finite elements interpolation called ‘‘Crouzeix-Raviart’’ (CR) finite elements [28, 13].

For the readers’ convenience we give here shortly simple arguments which explain this fact (which are known, see for instance [42, Lemma 2]). Although CR elements are defined in any dimension, let us first expose basic arguments in 2D.

We first claim that in a discrete sense, $\text{curl } \mathbf{p}^h = 0$. The precise sense is (obviously) as follows: For any continuous and piecewise linear (P1) function $\varphi : \mathcal{T}^h \mapsto \mathbb{R}$ one has

$$\int_{\Omega^h} \mathbf{p}^h \cdot (\nabla \varphi)^\perp dx = 0, \quad (9)$$

where we denoted $(a, b)^\perp := (-b, a)$, for $(a, b) \in \mathbb{R}^2$ a counter-clock rotation by 90 degrees. Indeed, by definition, $\nabla \varphi$ is constant in each triangle T_i and will denote its value by $(\nabla \varphi)_i$. Clearly, \mathbf{p}^h is also constant on each triangle (with value \mathbf{p}_i), and therefore (9) becomes

$$\sum_{i=1}^N |T_i| \mathbf{p}_i \cdot (\nabla \varphi)_i^\perp = \sum_{i=1}^N \left(\int_{T_i} Du \right) \cdot (\nabla \varphi)_i^\perp = \int_{\Omega} (\nabla \varphi)^\perp \cdot Du = 0.$$

as $\text{curl } Du = 0$. To prove this rigorously, if $|Du|(\bigcup_{i=1}^N (\partial T_i)) = 0$ one can first approximate u with smooth functions, for which the integral is trivially zero, and pass to the limit, if not, one can first approximate the mesh and φ with a slightly perturbed mesh $\mathcal{T}' = \{T'_i : i = 1, \dots, N\}$ and φ' such that $|Du|(\bigcup_{i=1}^N (\partial T'_i)) = 0$ and $\int_{\Omega} (\nabla \varphi')^\perp \cdot Du \approx \int_{\Omega} (\nabla \varphi)^\perp \cdot Du = 0$ and pass to the limit.

Let us now show that one can integrate back any discrete field \mathbf{p}^h satisfying (9) into a function u^h , which is, if \mathbf{p}^h was obtained from (4), an approximation of u . In Figure 2 we consider a vertex \bar{v} which is common to a set of ordered triangles T_i , $i = 1, \dots, N$ of a triangulation \mathcal{T}^h , defined via the triplets (v_i, v_{i+1}, \bar{v}) and $v_{N+1} = v_1$. Moreover, we consider for φ in (9) a function equal to 1 in the vertex \bar{v} and zero in all other vertices v_1, \dots, v_{N+1} . The rotated gradients $(\nabla \varphi)_i^\perp$ in the triangles are given by

$$(\nabla \varphi)_i^\perp = \frac{v_{i+1} - v_i}{h_i |v_{i+1} - v_i|} = \frac{v_{i+1} - v_i}{|T_i|},$$

where h_i denotes the height of the triangle relative to the edge (v_i, v_{i+1}) . Then, (9) becomes

$$\int_{\Omega^h} \mathbf{p}^h \cdot (\nabla \varphi)^\perp dx = \sum_{i=1}^N |T_i| \mathbf{p}_i \cdot (\nabla \varphi)_i^\perp = \sum_{i=1}^N \mathbf{p}_i \cdot (v_{i+1} - v_i) = 0$$

that is, the circulation of \mathbf{p}^h around the loop v_1, \dots, v_{N+1} vanishes. In particular, the circulation of \mathbf{p}^h also vanishes for all loops passing the points $\alpha v_i + (1 - \alpha) \bar{v}$ for $\alpha \in [0, 1]$.

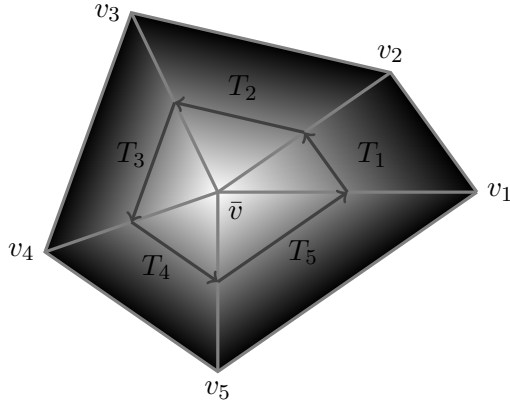


Figure 2: A set of ordered triangles T_i , $i = 1, \dots, 5$ defined through the triplets (v_i, v_{i+1}, \bar{v}) with a common vertex \bar{v} in the middle. In the background, we show the P1 function φ which is one at \bar{v} and zero at the other vertices. In red we show the loop of vectors originating by the rotated and scaled gradients of the function ϕ passing through the middle of each edge.

The most interesting case is $\alpha = \frac{1}{2}$ since it yields points in the middle of the edges which are also part of neighboring loops. It follows that one can assign a (unique up to a global constant) values u_e in the middle of each edge e of the whole triangulation (at least if the domain is simply connected). We will soon see a more global characterization of \mathbf{p}^h which allows to consider it as a discrete gradient in arbitrary domains (and dimension), see Lemma 2.4.

Actually, if \mathbf{p}^h is obtained through (4), one checks easily that in addition, the correct value to assign in the middle of an edge is the average of u on the edge. Indeed, Figure 3 shows a triangle $T = (v_1, v_2, v_3) \in \mathcal{T}^h$, where we have chose the normalized edge direction $e = (v_3 - v_2)/|v_3 - v_2|$. Then, one has (assuming $|Du|(\partial T) = 0$, else taking into account Remark 2.1)

$$\int_T e \cdot Du = \int_{\partial T} u(x) e \cdot \nu_T d\mathcal{H}^1 = \sin(\theta_3) \int_{[v_1, v_3]} u dx - \sin(\theta_2) \int_{[v_1, v_2]} u dx.$$

Denoting h_1 the height from v_1 , one has $h_1 = |v_2 - v_1| \sin(\theta_2) = |v_3 - v_1| \sin(\theta_3)$ and $|T| = h_1 |v_3 - v_2|/2$, so that we have

$$\mathbf{p}^h \cdot e = \frac{1}{|T|} \int_T e \cdot Du = \frac{2}{|v_3 - v_2|} \left(\int_{[v_1, v_3]} u dx - \int_{[v_1, v_2]} u dx \right).$$

The right-hand side is exactly the gradient, in the direction e , of the affine function which is equal in the middle of each edge $[v_i, v_j]$ to $\int_{[v_i, v_j]} u dx$.

In the next section, we consider a better characterization of Crouzeix-Raviart gradients (which is also easier to handle in arbitrary dimension).

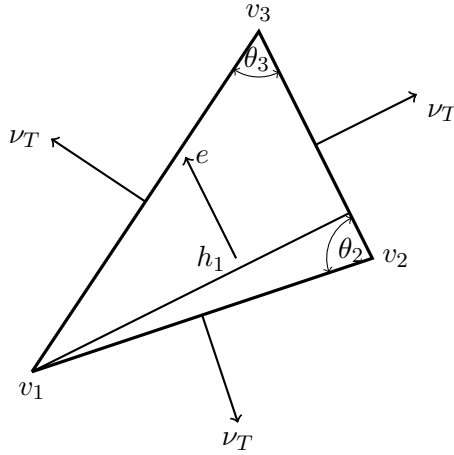


Figure 3: A triangle $T = (v_1, v_2, v_3) \in \mathcal{T}^h$ together with its outer normals ν_T , the normalized edge direction $e = (v_3 - v_2)/|v_3 - v_2|$ shown and the two angles $\theta_3 = \widehat{v_1 v_3 v_2}$ and $\theta_2 = \widehat{v_3 v_2 v_1}$.

2.3. Characterization of Crouzeix-Raviart gradients

We introduce, given the mesh \mathcal{T}^h , the set $\mathcal{N}(\mathcal{T}^h)$ of nodes which are the middle points of the edges of the triangles of \mathcal{T}^h . Moreover, we define the nonconforming Crouzeix-Raviart (CR) finite-element space

$$V(\mathcal{T}^h) = \left\{ u : \Omega^h \rightarrow \mathbb{R} : u|_T \text{ affine } \forall T \in \mathcal{T}^h, u \text{ continuous at } x, \forall x \in \mathcal{N}(\mathcal{T}^h) \right\}. \quad (10)$$

In this space, the gradient $\mathbf{p}^h = D_h u$ of a function u is defined as the vectorial function which coincides with $\nabla u|_T$ on each triangle (and we will denote $D_h u(T)$ its value in the triangle T). We call it a ‘‘Crouzeix-Raviart (CR) gradient’’, and observe that this field is a P0 finite elements vector field, that is, a vector field which is constant on each triangle. The analysis in the previous section shows that for any $u \in BV(\Omega)$ with $|Du|(\bigcup_{i=1}^N \partial T_i) = 0$, the function u^h defined by assigning to $x \in \mathcal{N}(\mathcal{T}^h)$ the average value of u on the edge through x , and affine with slope \mathbf{p}_i , given by (4), in each T_i , belongs to $V(\mathcal{T}^h)$. (We call it the projection of u to $V(\mathcal{T}^h)$, and it is naturally extended to any $u \in BV(\Omega)$ by assigning the boundaries $\bigcup_{i=1}^N \partial T_i$ arbitrarily to one of the neighbouring triangle, such as the one with lower index as already suggested). In particular, \mathbf{p}^h , given by (5), is a CR gradient.

In the following, we give a characterization of CR gradients which is more precise than (9). It is based on zeroth order ‘‘Raviart-Thomas’’ (RT0) vector fields [43], which are defined by their fluxes across the edges of the triangulation. Inside the triangles, RT0 fields are affine functions and they are continuous across the edges of the mesh. Given a triangle $T = (v_1, v_2, v_3)$, the RT0 field $\phi_T(x)$ inside the triangle T is given by

$$\phi_T(x) = \frac{f_1}{h_1|v_3 - v_2|}(x - v_1) + \frac{f_2}{h_2|v_3 - v_1|}(x - v_2) + \frac{f_3}{h_3|v_2 - v_1|}(x - v_3),$$

where f_i are the fluxes through the edges e_i which are opposite to the vertex v_i and h_i are the heights relative to the vertex v_i . With this choice the flux of the field $\phi_T(x)$ through each edge e_i with outward normal ν_i is constant. Indeed, for $x \in e_1 = [v_2, v_3]$,

$$\phi_T(x) \cdot \nu_1 = \frac{f_1(x - v_1) \cdot \nu_1}{h_1|v_3 - v_2|} + \frac{f_2(x - v_2) \cdot \nu_1}{h_2|v_3 - v_1|} + \frac{f_3(x - v_3) \cdot \nu_1}{h_3|v_2 - v_1|} = \frac{f_1}{|v_3 - v_2|}$$

since $(x - v_1) \cdot \nu_1 = h_1$ and $(x - v_2) \cdot \nu_1 = (x - v_3) \cdot \nu_1 = 0$. Hence

$$\int_{e_1} \phi_T(x) \cdot \nu_1 d\mathcal{H}^1 = f_1,$$

and the total flux through the edges of T is given by (denoting $\nu_T \in \{\nu_1, \nu_2, \nu_3\}$ the outwards normal to T):

$$\int_{\partial T} \phi_T \cdot \nu_T d\mathcal{H}^1 = f_1 + f_2 + f_3.$$

We denote $RT0(\mathcal{T}^h)$ the space of RT0 vector fields relative to the mesh \mathcal{T}^h , while $RT0_0(\mathcal{T}^h) \subset RT0(\mathcal{T}^h)$ is the subspace of Raviart-Thomas vector fields with zero flux through $\partial\Omega_h$.

The next result establishes a relationship between CR gradients and zero-divergence RT0 fields.

Lemma 2.4. *The P0 field \mathbf{p}^h is a CR gradient if and only if it is orthogonal to all zero-divergence RT0 fields ϕ with vanishing fluxes on $\partial\Omega^h$, that is*

$$\int_{\Omega^h} \phi \cdot \mathbf{p}^h dx = 0 \quad \forall \phi \in RT0_0(\mathcal{T}^h).$$

Remark 2.5. *Strangely, we have not found this statement in this form in the literature. It is however related to Helmholtz Decomposition type results and is for instance used in [17], where a (more complicated and 2D) proof of the duality result (Sec. 3.3 below) is found also in a nonlinear setting.*

Proof. First, we show that CR gradients are orthogonal to zero divergence RT0 fields. Given a RT0 field ϕ defined in Ω^h and $u^h \in V(\mathcal{T}^h)$ with $\mathbf{p}^h = D_h u^h$ (which can be the projection, as mentioned, of a BV function u), then for $T_i \in \mathcal{T}^h$,

$$\int_{T_i} \phi \cdot \mathbf{p}^h dx = \int_{\partial T_i} u^h \phi \cdot \nu_{T_i} d\mathcal{H}^1 - \int_{T_i} u^h \operatorname{div} \phi dx,$$

where in the second integral, u^h is the inner trace of the affine function u^h in the triangle (recall that this function may be discontinuous through the edges of the triangle). Since $\phi \cdot \nu_{T_i}$ is constant on each edge $(e_j^i)_{j=1,2,3}$ of T_i , and since u^h has, on e_j^i , average value $u^h(x_j^i)$ where $x_j^i \in \mathcal{N}(\mathcal{T}^h)$ is the middle point of e_j^i , it yields

$$\int_{T_i} \phi \cdot \mathbf{p}^h dx = \sum_{j=1}^3 |e_j^i| u^h(x_j^i) \phi \cdot \nu_j^i - \int_{T_i} u^h \operatorname{div} \phi dx.$$

If $T_{j'}$ is a neighbouring triangle and $e_{j'}^{i'} = e_j^i$ the common edge, as u^h is continuous at $x_j^i = x_{j'}^{i'}$ and $\nu_j^i = -\nu_{j'}^{i'}$,

$$|e_j^i|u^h(x_j^i)\phi \cdot \nu_j^i + |e_{j'}^{i'}|u^h(x_{j'}^{i'})\phi \cdot \nu_{j'}^{i'} = 0.$$

Hence summing on all triangles, we obtain the general Green formula:

$$\int_{\Omega^h} \phi \cdot D_h u^h dx = \int_{\partial\Omega^h} u^h \phi \cdot \nu_{\Omega^h} d\mathcal{H}^1 - \int_{\Omega^h} u^h \operatorname{div} \phi dx. \quad (11)$$

(Remark that if u^h was obtained as the projection of a BV function u , then the second integral is also $\int_{\partial\Omega^h} u \phi \cdot \nu_{\Omega^h} d\mathcal{H}^1$.) In particular, if $\operatorname{div} \phi = 0$ and ϕ vanishes on $\partial\Omega^h$, we find

$$\int_{\Omega^h} \phi \cdot \mathbf{p}^h dx = 0.$$

Conversely, we show that if \mathbf{p}^h is orthogonal to all zero-divergence RT0 field with vanishing flux through $\partial\Omega_h$, then it is a CR gradient. Assume we are given a 2D P0 vector field \mathbf{p}^h such that $\int_{\Omega^h} \phi \cdot \mathbf{p}^h dx = 0$ for all $\mathbf{p}^h \in RT0_0(\mathcal{T}^h)$ with zero divergence. Let $x^0, x^1 \in \mathcal{N}(\mathcal{T}^h)$ be two midpoints of edges of the triangulation and assume we are given two different piecewise linear simple paths from x^0 to x^1 , intersecting at most once each triangle by joining two nodes of $\mathcal{N}(\mathcal{T}^h)$ by a straight segment. Letting $u(x^0) = 0$, we can integrate \mathbf{p}^h along each path Γ^i , $i = 1, 2$, to obtain two different values $u^i(x^1)$, $i = 1, 2$. If these values are the same (i.e. independent on the path), the circulation of \mathbf{p}^h along the path Γ^1 followed by Γ^2 will be zero, which in turn implies that \mathbf{p}^h is a CR gradient.

In fact, we can identify each path with a Raviart-Thomas vector field ϕ^i , such that the flux through an edge is 1 when the path crosses the edge, moving forward. Then, $\phi = \phi^2 - \phi^1$ is divergence free so that

$$\int \mathbf{p}^h \cdot \phi^1 dx = \int \mathbf{p}^h \cdot \phi^2 dx.$$

The same computation as before shows that this is equivalent to $u^1(x^1) = u^2(x^1)$. \square

2.4. Further obvious remarks

Given $u \in V(\mathcal{T}^h)$, we may define a P0 function u_0 by averaging u in each triangle $T \in \mathcal{T}^h$. Obviously (as u is affine in T), the value thus obtained is the same as the value in the center (of mass) c_T of the triangle, and the average of the three mid-point values of the edges. One has

$$\int_T |u(x) - u(c_T)| dx \leq |D_h u(T)| \int_T |x - c_T| dx \leq h|T| |D_h u(T)|.$$

In particular, summing on all triangles we find that:

$$\|u - u_0\|_{L^1(\Omega^h)} \leq h \int_{\Omega^h} |D_h u| dx. \quad (12)$$

The same would also clearly hold with a L^p norm on both sides, replacing the L^1 norm, for $p \in [1, \infty]$. See for instance [14] for more general estimates.

3. Definition of a Crouzeix-Raviart discrete total variation

3.1. Definition, and approximation properties

Given then $u \in V(\mathcal{T}^h)$, we define the approximate total variation

$$J_h(u) := \int_{\Omega^h} |D_h u| dx = \sum_{T \in \mathcal{T}^h} |T| |D_h u(T)|. \quad (13)$$

Here we recall that $D_h u(T)$ is the gradient of u in each triangle T (and not across the edges of the triangles). In this section, we show that in a variational sense, J_h is an approximation of the total variation.

The following result is obvious, considering the analysis in the previous sections. We assume that we are given an arbitrary family of triangulations \mathcal{T}^h , with $h \rightarrow 0$.

Proposition 3.1. *J_h Γ -converges to the total variation (1) as $h \rightarrow 0$, in $L^1(\Omega)$, as well as for the distributional convergence.*

Although the latter convergence is quite weak, we will see later on that without further assumption on the triangulations we can hardly hope for compactness in a better sense. For functional with growth $p > 1$, [42] propose a more precise approach to convergence (for regular meshes).

Proof. As we have seen, for any $u \in L^1(\Omega)$ with finite total variation, one can build u^h which will go to u as $h \rightarrow 0$, and such that $J_h(u^h) \rightarrow |Du|(\Omega)$. Conversely, if u^h is a sequence which converges to some u and $\sup_h J_h(u^h) < +\infty$, then clearly $\mathbf{p}^h(x)$ is bounded as a measure and converges weakly-* (up to subsequences) to some measure $\mathbf{p} \in \mathcal{M}(\Omega; \mathbb{R}^2)$. The fact that \mathbf{p} is orthogonal to zero divergence fields is easy to show by approximating compactly supported smooth vanishing divergence fields with RT0 fields and using Lemma 2.4. Moreover one obviously have

$$\int_{\Omega} |\mathbf{p}| \leq \liminf_{h \rightarrow 0} J_h(u^h).$$

It remains therefore to show that $\mathbf{p} = Du$. This will follow from the compactness result which we now state in Proposition 3.2. \square

Proposition 3.2. *Let $u^h \in V(\mathcal{T}^h)$ such that $\sup_h J(u^h) < +\infty$. Then, there exists a subsequence (u^{h_k}) and a constant $c_{h_k} \in \mathbb{R}$, and $u \in BV(\Omega)$, such that $u^{h_k} - c_{h_k} \rightarrow u$ in the sense of distributions in Ω as $k \rightarrow \infty$ (and $|Du|(\Omega) \leq \liminf_k J_{h_k}(u^{h_k})$).*

Moreover, if we assume that the triangulations are uniformly regular (in the classical sense: there exists $\bar{\theta} > 0$ such that the angles of the triangles are all larger than $\bar{\theta}$, or equivalently, there exists $\delta > 0$ such that for any triangle $T = (v_1, v_2, v_3) \in \mathcal{T}^h$, $h_3 \geq \delta |v_3 - v_1|$ where $h_3 = \text{dist}(v_3, (v_1, v_2))$ is the height of T from v_3), then the convergence $u^{h_k} - c_{h_k} \rightarrow u$ is strong in $L^p(\Omega)$ for any $p < 2$, and weak in $L^2(\Omega)$.

Proof. As we have seen, up to a subsequence, $\mathbf{p}^{h_k} := D_{h_k} u^{h_k} \xrightarrow{*} \mathbf{p}$ as measures and $\mathbf{p} = Du$ for some function $u \in BV(\Omega)$. We need to show that u is the limit of the u^{h_k} (up to constants).

First, without further assumption on the triangulations, the following is true: given ϕ a C^1 , compactly supported field and ϕ^h the RT0 fields obtained by evaluating the fluxes of ϕ across the edges of \mathcal{T}^h , one still has thanks to (11):

$$\int_{\Omega^{h_k}} u^{h_k} \operatorname{div} \phi^{h_k} dx = - \int_{\Omega^{h_k}} \phi^{h_k} \cdot D_{h_k} u^{h_k} dx \rightarrow - \int_{\Omega} \phi \cdot Du = \int_{\Omega} u \operatorname{div} \phi dx$$

as $k \rightarrow \infty$. The first integral, on the other hand (using the notation of Section 2.4) is also

$$\sum_{T \subset \Omega} \int_T u^{h_k}(x) dx \frac{1}{|T|} \int_T \operatorname{div} \phi(y) dy = \sum_{T \subset \Omega} u_0^{h_k}(T) \int_T \operatorname{div} \phi(y) dy = \int_{\Omega^h} u_0^{h_k} \operatorname{div} \phi(y) dy$$

so that we conclude that $u_0^{h_k} \rightarrow u$ (up to constants, one should for instance remove the averages of all the functions) in the distributional sense (and u^{h_k} as well, thanks to (12)).

We now assume in addition that the triangulations are uniformly regular, and consider again the P0 functions u_0^h . Observe that the jump of u_0^h across an edge $[v_2, v_3]$ common to two triangles $T = (v_1, v_2, v_3)$ and $T' = (v_2, v_3, v_4)$ is given by $D_h u^h(T) \cdot (2v_1 - v_2 - v_3)/6 - D_h u^h(T') \cdot (2v_4 - v_2 - v_3)/6$ (we recall that $u_0^h(T)$ is both given by the average of u^h in T triangle and by the middle value $u^h((v_1 + v_2 + v_3)/3)$). Hence,

$$\begin{aligned} & |u_0^h(T) - u_0^h(T')| |v_3 - v_2| \\ & \leq |D_h u^h(T)| \frac{|v_3 - v_2| |2v_1 - v_2 - v_3|}{6} + |D_h u^h(T')| \frac{|v_3 - v_2| |2v_4 - v_2 - v_3|}{6}. \end{aligned}$$

By the regularity assumption, $|v_3 - v_2| |2v_1 - v_2 - v_3| \leq |v_3 - v_2| |v_1 - v_2| + |v_3 - v_2| |v_1 - v_3| \leq 2|T|/\delta$, hence

$$|u_0^h(T) - u_0^h(T')| |v_3 - v_2| \leq \frac{1}{3\delta} \left(|T| |D_h u^h(T)| + |T'| |D_h u^h(T')| \right).$$

Summing on all the edges, it follows that (seing the P0 function u_0^h as a piecewise constant function with bounded variation)

$$\int_{\Omega^h} |Du_0^h| \leq \frac{1}{\delta} J_h(u^h).$$

Hence, up to a constant, u_0^h is bounded in $L^2(\Omega)$ and compact in $L^p(\Omega)$, $p < 2$. We conclude thanks to (12). We could have used general estimates for nonconforming finite elements, such as found in [14, Chap. 10]. \square

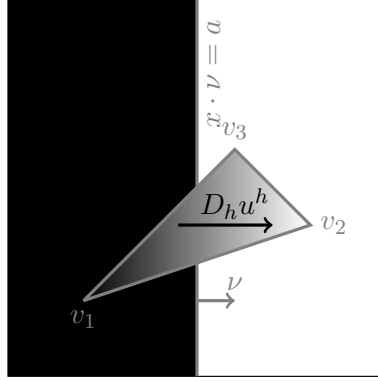


Figure 4: The figure shows the function u^h and its gradient for one triangle $T = (v_1, v_2, v_3)$ approximating the characteristic function of the half space $x \cdot \nu > a$ with $\nu = (1, 0)$ and $a \in \mathbb{R}$. The function u^h inside the triangle is affine with a gradient $D_h u^h$ which is simply a multiple of the normal ν .

3.2. The measure of straight lines

The Green formula (11) shows that J_h satisfies the following lower estimate, for $u \in V(\mathcal{T}^h)$ and ϕ^h a RT0 field:

$$J_h(u) \geq \sup \left\{ \int_{\partial\Omega^h} u \phi^h \cdot \nu_{\Omega^h} d\mathcal{H}^1 - \int_{\Omega^h} u \operatorname{div} \phi^h dx : |\phi^h(c_T)| \leq 1 \forall T \in \mathcal{T}^h \right\}, \quad (14)$$

where we recall that c_T refers to the center of the respective triangle T .

This formula would be an interesting way to build lower estimates for variational problems involving J_h if it were easy to build test RT0 fields ϕ^h satisfying the constraint, possibly from fields in the continuum. However, it is easy to check that given $\phi \in C_c^\infty(\Omega; \mathbb{R}^2)$ with $|\phi| \leq 1$ everywhere, the RT0 projection ϕ^h defined by assigning on each edge of the triangulation the flux of ϕ through the edge needs not satisfy $|\phi^h(c_T)| \leq 1$ (and can be substantially larger, independently on the mesh size h).

There is however one trivial situation where this can be used, and we obtain the following result, valid for any triangulation \mathcal{T}^h of Ω :

Proposition 3.3. *Let $\nu \in \mathbb{S}^1$ be a unit vector, $a \in \mathbb{R}$ and $u = \chi_{\{x \cdot \nu > a\}}$. Let u^h be the projection of u on CR functions, obtained by letting $D_h u^h(T) = \nu \mathcal{H}^1(\{x \cdot \nu = a\} \cap T) / |T|$ in each triangle $T \in \mathcal{T}^h$. Then for any CR function v with $v = u^h$ on $\mathcal{N}(\mathcal{T}^h) \cap \partial\Omega^h$,*

$$J_h(v) \geq J_h(u^h) = \mathcal{H}^1(\{x \cdot \nu = a\} \cap \Omega^h).$$

The proposition shows (see also Figure 4) that in some sense, the discrete functional J_h is perfectly isotropic, as its minimal value for a straight edge coincides with the length of the edge whatever the direction ν . We will see soon that unfortunately, this is not enough to make J_h a “perfect” approximation of the total variation.

Proof. The fact that $J_h(u^h) = \mathcal{H}^1(\{x \cdot \nu = a\} \cap \Omega^h)$ follows from (7). The other inequality follows from (14). Indeed, consider the constant field ν , which is of course a particular RT0 field defined on \mathcal{T}^h with norm less or equal to 1. We have therefore:

$$J_h(v) \geq \int_{\Omega^h} \nu \cdot D_h v = \int_{\Omega^h} \nu \cdot D_h u^h + \int_{\Omega^h} \nu \cdot D_h(v - u^h) = J_h(u^h) + \int_{\Omega^h} \nu \cdot D_h(v - u^h).$$

Now, thanks to (11),

$$\int_{\Omega^h} \nu \cdot D_h(v - u^h) = \int_{\partial\Omega^h} (v - u^h) \nu \cdot \nu_{\Omega^h}$$

and this vanishes since we have assumed that $v = u^h$ on all the midpoints of the edges of the triangles which form $\partial\Omega^h$. \square

3.3. A general duality formula

A natural question is whether equality holds in (14). The answer is clearly no, since one can check that the right-hand side of (14) depends only on the average values (in the middle of each simplex) of u , while J_h depends on the full function. The precise answer is as follows: give \bar{u} a P0 function, constant with value u_T on each element T , we let

$$J_h^0(\bar{u}) = \min \left\{ J_h(u) : u \in V(\mathcal{T}^h), u(c_T) = \bar{u}_T \forall T \in \mathcal{T} \right\}. \quad (15)$$

Then, it holds the following

Theorem 1. *For any $u \in V(\mathcal{T}^h)$ and $\bar{u} := (u(c_T))_{T \in \mathcal{T}^h}$,*

$$J_h^0(\bar{u}) = \sup \left\{ - \int_{\Omega^h} u \operatorname{div} \phi^h dx : \phi^h \in RT0_0(\mathcal{T}^h), |\phi^h(c_T)| \leq 1 \forall T \in \mathcal{T}^h \right\}. \quad (16)$$

We give a proof of this in Appendix A, for more general discretizations of Sobolev semi-norms. Note also that this can be derived from a similar result in [17] (the proof we propose is however simpler and easily seen to hold in any dimension).

3.4. The bad news: diffuse solutions

Proposition 3.3 shows that in theory, J_h is an excellent approximation of the total variation, which does not suffer from metrication errors or anisotropy as most other, in both the finite differences (such as graph based TVs [10], “isotropic” ℓ_2 -TV [20]) or finite elements (P1 based [5]) settings. However, we will show that it does not mean that the corresponding solution is always sharp: we now show a particular example where in addition to the projection u^h (which is sharp), we can build infinitely many other solutions with the same energy (and a transition width which of course will go to zero as $h \rightarrow 0$), for which the transition occurs across a large number of elements.

The setting is as follows: $\Omega = \Omega^h = [0, 1]^2$, $h = 1/n$, $n \geq 1$ divided in squares $C_{i,j} = [(ih, jh), ((i+1)h, jh)] \times [(ih, jh), (ih, (j+1)h)]$ for $i = 0, \dots, n-1$, $j = 0, \dots, n-1$. Then,

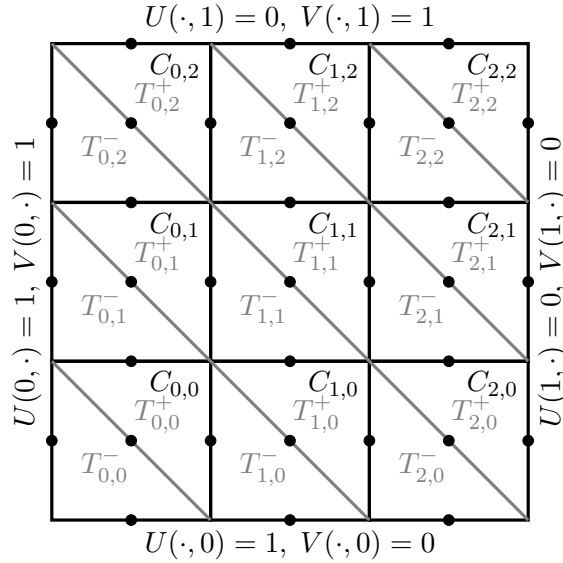


Figure 5: Discretization of the unit square $\Omega^h = [0, 1]^2$, $h = 1/3$ using squares $C_{0,0}, \dots, C_{2,2}$ each of size h^2 . Each square $C_{i,j}$ is then divided into an upper triangle $T_{i,j}^+$ and a lower triangle $T_{i,j}^-$. The nodes of the mesh, represented as black dots, are in the middle of the edges of the triangles. Note the boundary conditions imposed by the functions $U(x, y)$ and $V(x, y)$.

each square $C_{i,j}$ is divided into two triangles $T_{i,j}^\pm$ with $T_{i,j}^- = C_{i,j} \cap \{x + y \leq (i + j + 1)h\}$ and $T_{i,j}^+ = C_{i,j} \cap \{x + y > (i + j + 1)h\}$, see Figure 5. We recall that the nodes $\mathcal{N}(\{T_{i,j}^\pm\})$ of the mesh are the middle of the edges of the triangles, hence, here, the points $((i + 1/2)h, jh)$, $(ih, (j + 1/2)h)$, and $((i + 1/2)h, (j + 1/2)h)$.

Consider now the boundary conditions $U(x, y) = 1$ if $x = 0$ or $y = 0$, $U(x, y) = 0$ if $x = 1$ or $y = 1$, and $V(x, y) = 1$ if $x = 0$ or $y = 1$, and $V(x, y) = 0$ if $x = 1$ or $y = 0$. Clearly, Proposition 3.3 shows that the problems

$$\min_{u=U \text{ on } \partial\Omega \cap \mathcal{N}(\{T_{i,j}^\pm\})} J_h(u) \quad \text{and} \quad \min_{v=V \text{ on } \partial\Omega \cap \mathcal{N}(\{T_{i,j}^\pm\})} J_h(v) \quad (17)$$

where the minimizers are taken on CR functions on the mesh $\{T_{i,j}^\pm\}$, have both value $\sqrt{2}$. Moreover, a solution is given, for the first, by $u = \chi_{\{x+y \leq 1\}}$ (that is, the exact solution) on all nodes with $i + j \neq n$, and $u \in [0, 1]$ in any node with $i + j = n$. For the second we can find a solution considering the projection of $\chi_{\{x \leq y\}}$ on the CR functions: that is, the functions v with $v(x, y) = 1$ on all nodes with $x < y$, $v(x, y) = 0$ on all nodes with $x > y$, and $v(x, y) = 1/2$ on the nodes with $x = y$. We claim that if there are infinitely many solutions to the first problem in (17), they are all concentrated (their gradient is not zero only near the axis $x + y = 1$) and may differ from u only near the line $\{x + y = 1\}$. On the other hand, there are also infinitely many solutions to the second, but now they can be very diffuse, and in particular solutions v' with $0 < v' < 1$ on all the interior nodes.

To prove the first claim, we observe that for any solution u and for all i , u has to go from 1 to 0 along the “vertical” chain of vertices $((i + 1/2)h, j)$, $j = 0, \dots, n$. It follows that $\int_{\Omega} (D_h u)_1 dx = -1$. Similarly, $\int_{\Omega} (D_h u)_2 dx = -1$. Hence,

$$2 = - \int_{\Omega} (D_h u)_1 + (D_h u)_2 dx \leq \sqrt{2} \int_{\Omega} |D_h u| dx$$

using the 2D Cauchy-Schwartz inequality, pointwise. Since the right-hand side of this expression is also 2, it means that these inequalities are in fact equalities, and in particular that for a.e. x , $(D_h u)_1(x) = (D_h u)_2(x)$. As a consequence, in each triangle $T_{i,j}^{\pm}$, the values of u at the middle of the two shorter edges are equal. Using the boundary condition, it follows that u is unique except possibly on the middle vertices with $x + y = 1$ and thus in the triangles which contain them. It is easy to see that one can assign any arbitrary value $u(ih + h/2, (n - i)h - h/2) \in [0, 1]$ at each of these nodes without changing the value of the energy.

We now focus on the second claim. Observe that if v is a minimizer of the second problem in (17), one must have now, for the same reasons as for u ,

$$2 = \int_{\Omega} (D_h v)_1 - (D_h v)_2 dx \leq \sqrt{2} \int_{\Omega} |D_h v| dx = 2$$

so that $(D_h v)_1 = -(D_h v)_2 \geq 0$ a.e.

But now, this just imposes that the value of v in the middle of the longer edge of each triangle is the average of the two values in the middle of the shorter edges. Moreover, any CR function v which satisfies this condition and is globally nondecreasing from 0 to 1 along the nodes in the direction $(-1, 1)$ has minimal energy (as it will satisfy equality in the above equation).

Let us build a solution v “as diffusive as possible” (we would like to make this claim rigorous, however it is not clear how).

To simplify: we will look for solutions which satisfy the symmetry, for all $(x, y) \in (0, 1)^2$,

$$v(x, y) + v(y, x) = 1.$$

In fact, if v is an arbitrary solution, then so is $(x, y) \mapsto 1 - v(y, x)$, hence so is also $1/2 + (v(x, y) - v(y, x))/2$ which has the above symmetry and essentially a larger transition than v . Hence assuming this symmetry is not too restrictive when looking for diffuse solutions. Similarly, we will assume that v is symmetric with respect to the axis $\{x + y = 1\}$: $v(x, y) = v(1 - y, 1 - x)$.

In particular, we have that $v'(x, x) = 1/2$, and $v'(x, y) \leq 1/2$ if $x \geq y$. We build a v in $\{y \leq x \leq 1 - y\}$. We claim (and this is all we know for sure) that this is the solution with the largest values $v_{i-1,1/2}, v_{i,1/2}$, $1 \leq i \leq n$.

The first observation is that if $v_{1/2,1/2} = 1/2$, then the largest possible value for $v_{1,1/2}$ is also $1/2$. But then, since $v_{3/2,0} = 0$, one has $v_{3/2,1/2} = 1/4$. Then the largest possible value $v_{2,1/2}$ is also $1/4$, etc. One finds that the largest possible value of v in the first line is by taking $v_{(i-1/2),1/2} = v_{i,1/2} = 1/2^{i+1}$.

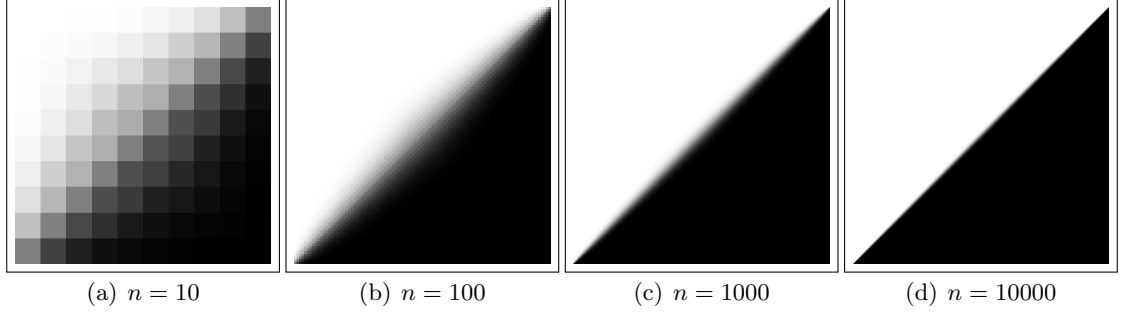


Figure 6: The smooth transition for different numbers of n , where we only plot the values $v_{i+1/2,j+1/2}$. Observe that only for a very fine mesh ($n = 10000$) we obtain a reasonably sharp transition.

We build then v by assuming that it is constant on the triangles $T_{i,j}^+$, at least as long as $i + j < n$. It will be clear later on why a “bad” v has to have such an oscillating gradient. One has

- $v_{i+1/2,i+1/2} = 1/2$ for $i \geq 0$;
- $v_{i+1/2,0} = 0$ for $i \geq 0$;
- $v_{i+1/2,j+1/2} = v_{i+1,j+1/2} = v_{i+1/2,j+1}$ for $i \geq j \geq 0, i + j \leq n - 1$;
- $v_{i+1/2,j+1/2} = (v_{i,j+1/2} + v_{i+1/2,j})/2$.

In particular, we deduce that

$$v_{i+1/2,j+1/2} = \frac{v_{i-1/2,j+1/2} + v_{i+1/2,j-1/2}}{2}.$$

Now, we introduce the variables $\xi_{i-j}^{i+j} = v_{i+1/2,j+1/2}$. One has $\xi_0^{2i} = 1/2$ for $i \geq 0$, and

$$\xi_l^m = \frac{\xi_{l-1}^{m-1} + \xi_{l+1}^{m-1}}{2} = \frac{1}{4}\xi_{l-2}^{m-2} + \frac{1}{2}\xi_l^{m-2} + \frac{1}{4}\xi_{l+2}^{m-2}.$$

For m even and $l = 2$, this reduces to

$$\xi_2^m = \frac{1}{8} + \frac{1}{2}\xi_2^{m-2} + \frac{1}{4}\xi_3^{m-2}.$$

Denoting now, $\zeta_l^m = \xi_{2l}^{2m}$, $l \geq 1$, we find that ($\zeta_l^0 = 0$ and) for $m \geq 1$,

$$\zeta^m = \frac{1}{4} \begin{pmatrix} 2 & 1 & 0 & 0 & \dots \\ 1 & 2 & 1 & 0 & \ddots \\ 0 & 1 & 2 & 1 & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \end{pmatrix} \zeta^{m-1} + \begin{pmatrix} \frac{1}{8} \\ 0 \\ 0 \\ \vdots \end{pmatrix}$$

Denoting B the matrix appearing in the recursion, we find that we must choose

$$\zeta^m = \left(\sum_{l=0}^{m-1} B^l \right) \begin{pmatrix} \frac{1}{8} \\ 0 \\ \vdots \end{pmatrix} = (I - B)^{-1}(I - B^m) \begin{pmatrix} \frac{1}{8} \\ 0 \\ \vdots \end{pmatrix}. \quad (18)$$

From there, we can go back to build v in $\{y < x < 1 - y\}$, and then by symmetry in the whole square $[0, 1]^2$. This shows that one can find a solution v , in that case, with a quite large transition layer. Of course any convex combination between this function and the function which is $1/2$ only at nodes where $x = y$, and equal to $\chi_{\{y > x\}}$ elsewhere, provides another solution.

In (18), we did not specify the size of the matrices: in fact, it is enough to take these of size larger than $n/2$ to compute v in a mesh of size $n \times n$. Choosing therefore $M > n/2$ we consider the $M \times M$ matrix

$$\frac{1}{4} \begin{pmatrix} 2 & 1 & 0 & 0 & \dots \\ 1 & 2 & 1 & 0 & \ddots \\ 0 & 1 & 2 & 1 & \ddots \\ \vdots & \ddots & \ddots & \ddots & \ddots \\ 0 & \dots & 0 & 1 & 2 \end{pmatrix}$$

This matrix has the M eigenvectors $e_k = (\sin(lk\pi/(M+1)))_{l=1}^M$, for $k = 1, \dots, M$, corresponding to the eigenvalues, respectively, $(1 + \cos(k\pi/(M+1)))/2 = \cos^2(k\pi/(2(M+1)))$. One can therefore compute (at least numerically) solutions ζ^m for large values of m , using that

$$\langle \zeta^m, e_k \rangle = \frac{1 - \cos^{2m}(k\pi/(2(M+1)))}{4 \sin(k\pi/(2(M+1)))} \cos(k\pi/(2(M+1))) \quad (19)$$

and $\zeta^m = \sum_k \langle \zeta^m, e_k \rangle e_k / \|e_k\|^2$. We show in Fig. 6 a plot of the corresponding function v for various n .

4. Implementation with adaptive meshes in 2D

In this section we propose, for a quadrangular mesh in 2D, an adaptive CR finite elements discretization which is based on the important observation that the triangulation $(T_{i,j}^\pm)$ in Section 3.4 shares the same nodes as the triangulation obtained by dividing each $C_{i,j}$ in the other possible way, that is, along the axis $\{x - ih = y - jh\}$.

Let us define v' as the function which is equal to v (defined in the previous section) on the nodes, but is now a CR function on the flipped mesh. Then, in all the squares $C_{i,j}$ where v was constant in one of the triangles $T_{i,j}^\pm$ ($T_{i,j}^+$ for $i+j < n$, $T_{i,j}^-$ for $i+j > n$), one observes that $D_h v'$ is either horizontal or vertical so that $|D_h v'| = |(D_h)_1 v'| + |(D_h)_2 v'| =: |D_h v'|_1$ (with an obvious notation).

Let $S_h = \bigcup_{i+j=n} C_{i,j}$ be the remaining squares (where, in fact, it is easy to see that $v' = v$), and $\tilde{S}_h = \{1 - h \leq x + y \leq 1 + h\} \supset S_h$. By minimality, we must have that (we use $D_h v = 0$ in $\tilde{S}_h \setminus S_h$, $v = v'$ in S_h)

$$\sqrt{2}h = \int_{\tilde{S}_h} |D_h v| dx = \int_{S_h} |D_h v| dx = \int_{S_h} |D_h v'| dx.$$

It yields that

$$\begin{aligned} 2 &= \int_{\Omega} |D_h v'|_1 dx = \int_{\Omega \setminus S_h} |D_h v'| dx + \int_{S_h} |D_h v'|_1 dx \\ &= \int_{\Omega} |D_h v'| dx + \int_{S_h} |D_h v'|_1 - |D_h v'| dx \leq \int_{\Omega} |D_h v'| dx + (\sqrt{2} - 1) \int_{S_h} |D_h v'| dx. \end{aligned}$$

Hence, we see that in the new triangulation,

$$\int_{\Omega} |D_h v'| dx \geq 2 - (2 - \sqrt{2})h. \quad (20)$$

This is much larger than the optimal value of $\sqrt{2}$. It means that if we had minimized here the *maximum* of the discrete total variation over the two possible triangulations, such a bad solution would have been ruled out and we would have recovered in the best solution.

We clearly see that we could even do better: we could choose in *each* square $C_{i,j}$ the best triangulation. This is what we describe in the next section.

4.1. Total variation on a square mesh

As before, we use an image domain $\Omega = [0, 1]^2$, set the scale parameter $h = 1/N$, with $N \geq 1$, and discretize Ω into $N \times N$ squares of size $h \times h$. As mentioned above, given such a square mesh in 2D, we can build (2^{N^2}) different triangulations by cutting each square into two triangles, in two possible ways. It is particularly remarkable that in this case, the nodes of the Crouzeix-Raviart spaces associated to these triangulations, which are the middle of the edges, remain the same. As a consequence, given the values on the nodes, we can introduce an approximation of the total variation given by the maximum, over all possible triangulations, of the discrete functional (13).

As shown in Figure 5 we need to store values of the image u in the middle of each horizontal and vertical edge as well as in the centers of each square. Therefore we introduce the three index sets \mathcal{I}_h , \mathcal{I}_v , and \mathcal{I}_c which are given by

$$\begin{aligned} \mathcal{I}_h &= \left\{ \mathbf{i} = \left(\left(i + \frac{1}{2} \right), j \right) : 0 \leq i \leq N - 1, 0 \leq j \leq N \right\}, \\ \mathcal{I}_v &= \left\{ \mathbf{i} = \left(i, \left(j + \frac{1}{2} \right) \right) : 0 \leq i \leq N, 0 \leq j \leq N - 1 \right\}, \\ \mathcal{I}_c &= \left\{ \mathbf{i} = \left(\left(i + \frac{1}{2} \right), \left(j + \frac{1}{2} \right) \right) : 0 \leq i \leq N - 1, 0 \leq j \leq N - 1 \right\}. \end{aligned}$$

We also set $\mathcal{I} = \mathcal{I}_h \cup \mathcal{I}_v \cup \mathcal{I}_c$, which refers to the complete set of nodal values. Note that $|\mathcal{I}_h| = |\mathcal{I}_v| = N(N + 1)$ and $|\mathcal{I}_c| = N^2$, hence, $|\mathcal{I}| = 3N^2 + 2N$. This shows

that the proposed discretization of the total variation increases the number of variables roughly by a factor of three. A more memory efficient (but also more diffusive) version is proposed in section B.

The discrete image is now given by a vector $u \in \mathbb{R}^{\mathcal{I}}$, where we can (by a slight abuse of notation) identify the values of the discrete image with the values of the underlying piecewise affine CR function by means of $u_{\mathbf{i}} = u(\mathbf{i}h)$ for all $\mathbf{i} \in \mathcal{I}$.

Next, we introduce the linear operators $D_1, D_2 : \mathbb{R}^{\mathcal{I}} \mapsto \mathbb{R}^{\mathcal{I}_c \times 4}$, each of them computing per square $\mathbf{i} \in \mathcal{I}_c$ the 4 possible differences out of the 5 nodal values.

$$(D_1 u)_{\mathbf{i}} = \begin{pmatrix} (D_1 u)_{\mathbf{i},1} \\ (D_1 u)_{\mathbf{i},2} \\ (D_1 u)_{\mathbf{i},3} \\ (D_1 u)_{\mathbf{i},4} \end{pmatrix} = \begin{pmatrix} u_{\mathbf{i}} - u_{\mathbf{i}-(\frac{1}{2},0)} \\ u_{\mathbf{i}} - u_{\mathbf{i}-(0,\frac{1}{2})} \\ u_{\mathbf{i}+(\frac{1}{2},0)} - u_{\mathbf{i}} \\ u_{\mathbf{i}+(0,\frac{1}{2})} - u_{\mathbf{i}} \end{pmatrix},$$

and

$$(D_2 u)_{\mathbf{i}} = \begin{pmatrix} (D_2 u)_{\mathbf{i},1} \\ (D_2 u)_{\mathbf{i},2} \\ (D_2 u)_{\mathbf{i},3} \\ (D_2 u)_{\mathbf{i},4} \end{pmatrix} = \begin{pmatrix} u_{\mathbf{i}+(\frac{1}{2},0)} - u_{\mathbf{i}} \\ u_{\mathbf{i}} - u_{\mathbf{i}-(\frac{1}{2},0)} \\ u_{\mathbf{i}} - u_{\mathbf{i}-(0,\frac{1}{2})} \\ u_{\mathbf{i}+(0,\frac{1}{2})} - u_{\mathbf{i}} \end{pmatrix},$$

for all $\mathbf{i} \in \mathcal{I}_c$, where D_1 computes the differences for the first triangulation (the one shown in Figure 5) and D_2 computes the differences of the second triangulation. Moreover, we define the operator $D : \mathbb{R}^{\mathcal{I}} \mapsto \mathbb{R}^{\mathcal{I}_c \times 4 \times 2}$, which simply combines the two preceding operators such that

$$(Du)_{\mathbf{i}} = ((D_1 u)_{\mathbf{i}}, (D_2 u)_{\mathbf{i}}), \quad \forall \mathbf{i} \in \mathcal{I}_c.$$

We are now ready to state the proposed discretization of the total variation, denoted as adaptive Crouzeix-raviart (ACR). It is defined as the energy

$$J_h(u) = h \sum_{\mathbf{i} \in \mathcal{I}_c} f((Du)_{\mathbf{i}}). \quad (21)$$

Let $\xi \in \mathbb{R}^{4 \times 2}$ with

$$\xi = \begin{pmatrix} \xi_{1,1} & \xi_{1,2} \\ \xi_{2,1} & \xi_{2,2} \\ \xi_{3,1} & \xi_{3,2} \\ \xi_{4,1} & \xi_{4,2} \end{pmatrix} = \begin{pmatrix} (D_1 u)_{\mathbf{i},1}, (D_2 u)_{\mathbf{i},1} \\ (D_1 u)_{\mathbf{i},2}, (D_2 u)_{\mathbf{i},2} \\ (D_1 u)_{\mathbf{i},3}, (D_2 u)_{\mathbf{i},3} \\ (D_1 u)_{\mathbf{i},4}, (D_2 u)_{\mathbf{i},4} \end{pmatrix},$$

the function $f : \mathbb{R}^{4 \times 2} \mapsto \mathbb{R}$ is given by

$$f(\xi) = \max \left\{ \left(\sqrt{\xi_{1,1}^2 + \xi_{2,1}^2} + \sqrt{\xi_{3,1}^2 + \xi_{4,1}^2} \right), \left(\sqrt{\xi_{1,2}^2 + \xi_{2,2}^2} + \sqrt{\xi_{3,2}^2 + \xi_{4,2}^2} \right) \right\}. \quad (22)$$

The function f computes on each square the maximum of the total variation for each of the two possible triangulations. Then, the local contributions are summed over all squares, which gives our proposed adaptive total variation.

In some applications, the given image datum might not be given as averages along the edges of the triangles but rather as averages across pixels. In order to relate the edge values of the triangles to such pixel averages, we introduce a linear operator $A : \mathbb{R}^{\mathcal{I}} \mapsto \mathbb{R}^{\mathcal{I}_c}$ which computes the averages of the two triangles corresponding to the images pixels:

$$(Au)_{\mathbf{i}} = \frac{2u_{\mathbf{i}} + u_{\mathbf{i}-(\frac{1}{2},0)} + u_{\mathbf{i}+(\frac{1}{2},0)} + u_{\mathbf{i}-(0,\frac{1}{2})} + u_{\mathbf{i}+(0,\frac{1}{2})}}{6}, \quad \forall \mathbf{i} \in \mathcal{I}_c.$$

Remark 4.1. *Before proceeding with imaging applications, we would like to comment on the proposed discrete ACR energy.*

- *In (22), we are minimizing the maximum of two convex functions, which in general is larger than the minimal value of the single functions. Hence, it could happen that we slightly overestimate the value of the total variation. In numerical results it turns out that such a case is rare, but may occurs, for example, if the orientation of a line discontinuity is $\{\pi/8, 3\pi/8, \dots\}$, see the second row in Table 1. However, this difference is only marginal and has almost no influence on the quality of the solution.*
- *Instead of minimizing the maximum of the two functions, we could have minimized the average as well. While this seems slightly simpler from an optimization point of view, it leads to blurrier results and does not allow to identify the mesh afterwards.*

4.2. Application to image processing problems

In order to apply the proposed ACR discretization of the total variation to image processing problems, we consider generic optimization problems of the form

$$\min_u F(Du) + G(Au) + H(u), \quad (23)$$

where $F(Du) = \lambda J_h(u)$ with $\lambda > 0$ being a regularization parameter. For simplicity we set $h = 1$, because if $h \neq 1$, its value can always be lagged into the regularization parameter λ .

The functions G and H are convex functions representing boundary conditions or data fidelity terms. Observe that while the function G is defined on pixel averages the function H is defined on nodes of the triangles which corresponds to the edge averages. This allows to adapt the proposed discrete version of the total variation to the specific properties of different applications. For example, a data fidelity term with respect to a noisy image might be better specified based on the pixel averages while a boundary condition might be better specified based on the edge averages. We shall assume here that the functions G and H are of the form $G(v) = \sum_{\mathbf{i} \in \mathcal{I}_c} g_{\mathbf{i}}(v_{\mathbf{i}})$ and $H(u) = \sum_{\mathbf{i} \in \mathcal{I}} h_{\mathbf{i}}(u_{\mathbf{i}})$ which will allow to efficiently compute their proximal maps.

As the problem (23) represents a non-smooth convex optimization problem, we follow [22] and consider instead its saddle-point formulation

$$\min_u \max_{\mathbf{p}, q} \langle Du, \mathbf{p} \rangle_P + \langle Au, q \rangle_Q + H(u) - G^*(q) - F^*(\mathbf{p}), \quad (24)$$

where $\mathbf{p} = (\mathbf{p}_i)_{i \in \mathcal{I}_c} \in \mathbb{R}^{\mathcal{I}_c \times 4 \times 2}$, with $\mathbf{p}_i \in \mathbb{R}^{4 \times 2}$, and $q = (q_i)_{i \in \mathcal{I}_c} \in \mathbb{R}^{\mathcal{I}_c}$ with $q_i \in \mathbb{R}$ are the dual variables. The function F^* denotes the convex conjugate of the function F . It is given by

$$F^*(\mathbf{p}) = \sum_{i \in \mathcal{I}_c} f^*(\mathbf{p}_i / \lambda),$$

where f^* refers to the convex conjugate of the function f defined in (22). Likewise, the function G^* denotes the convex conjugate of the function G . The inner products $\langle \cdot, \cdot \rangle_P$ and $\langle \cdot, \cdot \rangle_Q$ are given by

$$\langle Du, \mathbf{p} \rangle_P = \sum_{i \in \mathcal{I}_c} \sum_{m=1}^4 \sum_{n=1}^2 (Du)_{i,m,n} \mathbf{p}_{i,m,n}, \quad \langle Au, q \rangle_Q = \sum_{i \in \mathcal{I}_c} (Au)_i q_i.$$

We will also make use of the adjoint operators $D^* : \mathbb{R}^{\mathcal{I}_c \times 4 \times 2} \mapsto \mathbb{R}^{\mathcal{I}}$ and $A^* : \mathbb{R}^{\mathcal{I}_c} \mapsto \mathbb{R}^{\mathcal{I}}$ which are defined through the identities $\langle Du, \mathbf{p} \rangle_P = \langle u, D^* \mathbf{p} \rangle_U$ and $\langle Au, q \rangle_Q = \langle u, A^* q \rangle_U$ with inner products $\langle \cdot, \cdot \rangle_U$ given by

$$\langle u, D^* \mathbf{p} \rangle_U = \sum_{i \in \mathcal{I}_c} u_i (D^* \mathbf{p})_i, \quad \langle u, A^* q \rangle_U = \sum_{i \in \mathcal{I}_c} u_i (A^* q)_i.$$

Moreover, we need to evaluate the proximal operators with respect to the functions F^* , G^* and H . Thanks to their structure, they decompose into local operations:

$$\begin{aligned} \hat{u} &= \text{prox}_{\tau H}(\bar{u}) &\iff &\hat{u}_i = \text{prox}_{\tau h_i}(\bar{u}_i), \quad \forall i \in \mathcal{I}, \\ \hat{q} &= \text{prox}_{\sigma G^*}(\bar{q}) &\iff &\hat{q}_i = \text{prox}_{\sigma g_i^*}(\bar{q}_i), \quad \forall i \in \mathcal{I}_c, \\ \hat{\mathbf{p}} &= \text{prox}_{\sigma F^*}(\bar{\mathbf{p}}) &\iff &\hat{\mathbf{p}}_i = \text{prox}_{\sigma f^*(\cdot/\lambda)}(\bar{\mathbf{p}}_i) \quad \forall i \in \mathcal{I}_c. \end{aligned}$$

The proximity operators with respect to the functions σg_i^* and τh_i are usually easy to compute (also thanks to Moreau's identity) as the functions g_i and h_i usually represent just boundary conditions or simple 1D functions, see [25] for standard examples. The proximity operator with respect to the function $\sigma f^*(\cdot/\lambda)$ is more complicated but fortunately it also admits a closed form solution, see Appendix C for details.

We solve the saddle-point problem (24) using the first-order primal-dual algorithm studied in [22, 23]. The algorithm is as follows: Choose $u^0 \in \mathbb{R}^{\mathcal{I}}$, $\mathbf{p}^0 \in \mathbb{R}^{\mathcal{I}_c \times 4 \times 2}$, and $q^0 \in \mathbb{R}^{\mathcal{I}_c}$. Then for all $k \geq 0$ compute:

$$\begin{cases} u^{k+1} &= \text{prox}_{\tau H}(u^k - \tau(D^* \mathbf{p}^k + A^* q^k)) \\ \bar{u}^k &= 2u^{k+1} - u^k \\ \mathbf{p}^{k+1} &= \text{prox}_{\sigma F^*}(\mathbf{p}^k + \sigma D \bar{u}^k) \\ q^{k+1} &= \text{prox}_{\sigma G^*}(q^k + \sigma A \bar{u}^k) \end{cases} \quad (25)$$

It is shown in [22] that the algorithm converges as long as $\tau\sigma L^2 < 1$, where $L = \|(D, A)\|$ is the operator norm of the linear operator (D, A) . Moreover, its (partial) primal-dual gap converges with rate $O(\frac{1}{K})$, where K is the total number of iterations. In case G and/or F^* are strongly convex one can choose optimal (iteration dependent) step sizes such that the algorithm provides improved convergence rates, which in fact are optimal in the sense of lower bounds of first-order methods, see [22, 23, 25] for more details. Finally, the values of the pixel averages can be easily recovered as $u_a = Au$. In some applications, we observed that the central values $u_c = (u_i)_{i \in \mathcal{I}_c}$ can also be interesting as they can give sharper images and lower errors.

5. An error analysis

The aim of this section is to show for our non-conforming P1 approximation of the total variation error bounds which are not worse than standard P1, and which can even be improved if some (geometric) regularity is known. We address here the simple exemplar ‘‘ROF’’ [46] problem (3) for $g \in L^\infty(\Omega)$ (and $\tau = 1$), and in order not to deal with boundary issues, we assume that Ω is polyhedral so that the boundaries of the discretized domain and the original domain coincide. Hence we aim at comparing \bar{u} , solution of (3), with the solution \hat{u}_h^0 of

$$\min_{u_h \in P0} J_h^0(u_h) + \frac{1}{2} \int_{\Omega} (u_h - g)^2 dx \quad (26)$$

where we use the P0 variant of the Crouzeix-Raviart total variation (15). Our most striking result is that in case the solution of the dual problem of (3) has some regularity (which is known to be true in particular instances, but unknown in general, see Sec. 5.1.1), then the discrete energy differs from the continuous one from an error $O(h)$, while the L^2 distance between the discrete solution \hat{u}_h^0 and the solution \bar{u} is of order $O(\sqrt{h})$. This is the best one can hope for possibly discontinuous solutions, as in case \bar{u} has discontinuities, then it differs from its P0 projection by the same order of magnitude.

We require, for these error estimates to hold, a minimal regularity of the meshes. Basically, we must assume that there exists $\kappa \geq 1$ such that for any $T \in \mathcal{T}^h$ and any $z \in W^{1,\infty}(T; \mathbb{R}^d)$, denoting z^h the RT_0 projection of the vector field z in T ,

$$\|z^h\|_{L^\infty(T)} \leq \|z\|_{L^\infty(T)} + \kappa Lh \quad (27)$$

where L is the Lipschitz constant of z . Notice that it is enough to show that when $\operatorname{div} z^h = 0$ the constant vector z^h satisfies $|z^h| \leq c(\|z\|_{L^\infty(T)} + h\|\nabla z\|_{L^\infty(T)})$ and then apply this inequality to $z - \int_T z - \frac{x-x_c}{d} \int_T \operatorname{div} z$ to recover (27) with $\kappa = 4c$.

Such regularity is known to hold, obviously for 2D triangles with a minimal angle condition, but in fact also, in less obvious configurations. For instance, it is true for any d -dimensional simplex T which satisfies a ‘‘regular vertex property’’, that is, if the unit vectors along the edges at *one* vertex have a determinant bounded away from zero (by a constant c which will determine κ in (27)), see [1, Def. 2.1, Thm. 3.1] (where the proof is in 3D but can be extended). In 2D, it allows for instance for arbitrarily flat

triangles as long as one edge is of the order of the width of the element (and the two other remain large). Still in [1], it is shown for 3D tetraedra which satisfy the “maximum angle condition”, that is such that the maximum angle between faces and the maximum angle inside the faces are less than some $\bar{\psi} < \pi$.

5.1. Primal estimate

Given $u \in BV(\Omega)$, we consider u_h^0 and $u_h^c \in V(\mathcal{T}^h)$, respectively, its projection onto P0 and CR elements in a given mesh \mathcal{T}^h . This means that $u_h^0(T)$ is the average of u on $T \in \mathcal{T}^h$ while $u_h^c(x)$, $x \in \mathcal{N}(\mathcal{T}^h)$, is the average of u on the facet whose x is the center of mass. We also let $u_h^{0,c}$ the P0 function which coincides with $u_h^c(c_T)$ on each element T .

Lemma 5.1. *One has*

$$\|u_h^0 - u_h^{0,c}\|_{L^1} \leq \frac{h}{d} |Du|(\Omega) \quad (28)$$

(here d is the dimension, usually 2), and in particular $\|u_h^0 - u\|_{L^1} \leq ch|Du|(\Omega)$ and $\|u_h^{0,c} - u\|_{L^1} \leq ch|Du|(\Omega)$.

Proof. Let us consider a simple element T , with center c_T . One has

$$\begin{aligned} u_h^0(T) &= \frac{1}{|T|} \int_T u(x) dx = \frac{1}{d|T|} \int_T u(x) \operatorname{div}(x - c_T) dx \\ &= \frac{1}{d|T|} \int_{\partial T} u(x) (x - c_T) \cdot \nu_T d\mathcal{H}^{d-1} - \frac{1}{d|T|} \int_T (x - c_T) \cdot Du. \end{aligned} \quad (29)$$

On each facet of T , $(x - c_T) \cdot \nu_T$ is exactly the height of the simplex orthogonal to the facet, divided by $(d + 1)$ (3 in 2D). Using that the product of this height with the area of the facet is precisely $d|T|$, we deduce that the boundary integral in (29) is nothing but the average of the CR projection of u (as it is the average over the facets of the average on each facet of u), $u_h^c(c_T)$. Hence (28) follows. The other inequalities also follow, actually the first is classical (and easy to show) and the second follows from the first and (28). \square

Remark 5.2. *Observe that (29) is also true if Du in the last integral is replaced with $(Du - \mathbf{p})$ for \mathbf{p} a constant vector, since $\int_T (x - c_T) \cdot \mathbf{p} = 0$. So that we would also obtain a higher order error if u had more regularity: for instance if one had $D^2u \in L^1$, using $\int_T |Du - \frac{1}{T} \int_T Du| \leq ch \int_T |D^2u|$ one would obtain $\|u_h^0(T) - u_h^{0,c}\|_{L^1} \leq ch^2 \int_T |D^2u|$ and a second order error.*

It follows that if \bar{u} is the solution of (3), its projection $\bar{u}_h^{0,c}$ satisfies, using that $J_h^0(\bar{u}_h^{0,c}) \leq J_h(\bar{u}_h^c) \leq |D\bar{u}|(\Omega)$:

$$\begin{aligned} J_h^0(\bar{u}_h^{0,c}) + \frac{1}{2} \int_{\Omega} |\bar{u}_h^{0,c} - g|^2 dx &\leq \left(1 + 2\frac{h\|g\|_{\infty}}{d}\right) |D\bar{u}|(\Omega) + \frac{1}{2} \int_{\Omega} |\bar{u}_h^0 - g|^2 dx \\ &\leq (1 + Ch\|g\|_{\infty}) |D\bar{u}|(\Omega) + \frac{1}{2} \int_{\Omega} |\bar{u} - g|^2 dx \end{aligned}$$

Using the minimality of \hat{u}_h^0 and the strong convexity of the energy, we deduce:

$$(1 + Ch\|g\|_\infty) |D\bar{u}|(\Omega) + \frac{1}{2} \int_\Omega |\bar{u} - g|^2 dx - \left(J_h^0(\hat{u}_h^0) + \frac{1}{2} \int_\Omega |\hat{u}_h^0 - g|^2 dx \right) \geq \frac{1}{2} \int_\Omega |\bar{u}_h^{0,c} - \hat{u}_h^0|^2 dx. \quad (30)$$

We now need to show a similar estimate from above. We distinguish two cases.

5.1.1. First case: there exists a Lipschitz dual field

The Euler-Lagrange equation for (3) can be written in the following form: there exists $\bar{z} \in L^\infty(\Omega; B(0, 1))$ with

$$-\operatorname{div} \bar{z} + \bar{u} = g, \quad \bar{z} \cdot D\bar{u} = |Du|$$

and $\bar{z} \cdot \nu = 0$ on $\partial\Omega$ (unless one solves (3) with some boundary condition). Here the dot product $\bar{z} \cdot Du$ has to be understood in the sense of Anzelotti, see [4], observe also that \bar{z} is not unique (it can be modified outside of the support of $|Du|$). Convex duality shows that

$$\begin{aligned} \int_\Omega |D\bar{u}| + \frac{1}{2} \int_\Omega (\bar{u} - g)^2 dx &= \min_u \int_\Omega |Du| + \frac{1}{2} \int_\Omega (u - g)^2 dx \\ &= \sup_{\substack{z, \\ z \cdot \nu = 0 \text{ on } \partial\Omega}} - \int_\Omega \frac{1}{2} |\operatorname{div} z|^2 + g \operatorname{div} z \, dx = - \int_\Omega \frac{1}{2} |\operatorname{div} \bar{z}|^2 + g \operatorname{div} \bar{z} \, dx. \end{aligned} \quad (31)$$

Similarly thanks to (16), one has, introducing g_h the orthogonal projection of g on P0 functions,

$$\begin{aligned} J_h^0(\hat{u}_h^0) + \frac{1}{2} \int_\Omega (\hat{u}_h^0 - g_h)^2 dx &= \min_{u_h \in P0} J_h^0(u_h) + \frac{1}{2} \int_\Omega (u_h - g_h)^2 dx \\ &= \max \left\{ - \int_\Omega \frac{1}{2} |\operatorname{div} z_h|^2 + g_h \operatorname{div} z_h \, dx : z_h \in RT0_0(\mathcal{T}^h), |z_h(c_T)| \leq 1 \, \forall T \in \mathcal{T}^h \right\}. \end{aligned} \quad (32)$$

We denote \hat{z}_h^0 a maximizer of the dual problem, then, one has $\hat{u}_h = g_h + \operatorname{div} \hat{z}_h^0$.

We first assume that there exists a dual field \bar{z} which is L -Lipschitz. The existence of a Lipschitz field can be asserted in some situations. For instance, if g is the characteristic of a disk of radius R and Ω the whole plane, then $\bar{z}(x) = x/R$ for $|x| \leq R$ and $Rx/|x|^2$ for $|x| \geq R$ is a possible field, which is $(1/R)$ -Lipschitz. It is yet unclear in general what conditions on g and the domain are necessary for such a field to exist. In 2D, it could be that bounded is enough, in a convex domain (as, at least on the support of $|Du|$, the curvature of the level set, which is bounded by $2\|g\|_\infty$, is essentially controlling the gradient of \bar{z} , which is the normal to the level sets; however it is unclear if one can extend \bar{z} into a global Lipschitz field out of the support of $|Du|$). In general (and in higher dimension) no such regularity is known; we believe however that it could be expected if the level sets of g have some smoothness.

Under this assumption, let us project \bar{z} onto RT0 elements, defining a Raviart-Thomas field \bar{z}_h : this is simply done by computing the flux of \bar{z} through each edge of the triangulation (or facet in higher dimension). We observe that in this case, $\text{div } \bar{z}_h$ is the (L^2 -) orthogonal projection of $\text{div } z$ onto the P0 fields. In each T , thanks to the regularity condition (27), one has $\|\bar{z}_h(x)\| \leq 1 + \kappa Lh$, and in particular letting $\bar{z}'_h = \bar{z}_h/(1 + \kappa Lh)$, we find a field which is admissible in the dual problem (32). We first estimate:

$$\begin{aligned} - \int_{\Omega} \frac{1}{2} |\text{div } \bar{z}|^2 + g \text{div } \bar{z} \, dx &= -\frac{1}{2} \int_{\Omega} |\text{div } \bar{z} + g|^2 \, dx + \frac{1}{2} \int_{\Omega} |g|^2 \, dx \\ &\leq -\frac{1}{2} \int_{\Omega} |\text{div } \bar{z}_h + g_h|^2 \, dx + \frac{1}{2} \int_{\Omega} |g|^2 \, dx \\ &= - \int_{\Omega} \frac{1}{2} |\text{div } \bar{z}_h|^2 + g_h \text{div } \bar{z}_h \, dx + \frac{1}{2} \int_{\Omega} |g - g_h|^2 \, dx \quad (33) \end{aligned}$$

In the second line, we have used that the L^2 norm of a P0 projection is of course smaller than the L^2 norm of the initial function $\text{div } \bar{z} + g$, and in the third, that $\int |g|^2 \, dx - \int |g_h|^2 \, dx = \int |g - g_h|^2 \, dx$, being g_h the average of g on each element of the mesh. Then, we write, using (32):

$$\begin{aligned} - \int_{\Omega} \frac{1}{2} |\text{div } \bar{z}_h|^2 + g_h \text{div } \bar{z}_h \, dx &= -(1 + \kappa Lh) \int_{\Omega} \frac{1}{2} |\text{div } \bar{z}'_h|^2 + g_h \text{div } \bar{z}'_h \, dx - \frac{\kappa Lh(1 + \kappa Lh)}{2} \int_{\Omega} |\text{div } \bar{z}'_h|^2 \, dx \\ &\leq -(1 + \kappa Lh) \int_{\Omega} \frac{1}{2} |\text{div } \hat{z}_h^0|^2 + g_h \text{div } \hat{z}_h^0 \, dx \\ &= (1 + \kappa Lh) \left(J_h^0(\hat{u}_h^0) + \frac{1}{2} \int_{\Omega} (\hat{u}_h^0 - g_h)^2 \, dx \right). \quad (34) \end{aligned}$$

From (33)-(34), we deduce

$$\begin{aligned} \int_{\Omega} |D\bar{u}| + \frac{1}{2} \int_{\Omega} (\bar{u} - g)^2 \, dx &\leq (1 + \kappa Lh) \left(J_h^0(\hat{u}_h^0) + \frac{1}{2} \int_{\Omega} (\hat{u}_h^0 - g_h)^2 \, dx \right) + \frac{1}{2} \int_{\Omega} |g - g_h|^2 \, dx \\ &\leq (1 + \kappa Lh) \left(J_h^0(\hat{u}_h^0) + \frac{1}{2} \int_{\Omega} (\hat{u}_h^0 - g)^2 \, dx \right). \quad (35) \end{aligned}$$

Together with (30), we deduce that

$$\|\bar{u}_h^{0,c} - \hat{u}_h^0\|_{L^2(\Omega)} \leq C\sqrt{h} \quad (36)$$

with a constant C depending on L and the minimal energy. Observe that the same rate is proven in [8] without any assumption (such as here, on \bar{z}), however for a simpler (ℓ_1 -based, or anisotropic) total variation, which is easier to discretize on a square grid or structured mesh. Thanks to Lemma 5.1, (36) also holds with $\bar{u}_h^{0,c}$ replaced with \bar{u} .

5.2. Second case: simpler domain, g (slightly) regular

If we do not know whether a Lipschitz \bar{z} exists, it is a bit harder to show estimates, and these, in turn, are a bit weaker. The easiest way is to follow the strategy in [47]. To simplify (a lot) we assume our domain is periodic (that is, a torus). In this case, we introduce a parameter $\varepsilon > 0$ and a smoothing kernel, and replace \bar{z} with $\bar{z}_\varepsilon := \eta_\varepsilon * \bar{z}$ which is c/ε -Lipschitz (where $c = \int |\nabla \eta| dx$). We then project this smoothed vector field onto Raviart-Thomas fields: it gives a field $\bar{z}_{\varepsilon,h}$ with $|\bar{z}_{\varepsilon,h}| \leq 1 + ch/\varepsilon$ for some $c > 0$, thanks to (27). First observe that (34) still holds exactly as before, with κL replaced with c/ε and \bar{z}_h replaced with $\bar{z}_{\varepsilon,h}$. Next, we write that (denoting $\bar{u}_\varepsilon = \eta_\varepsilon * \bar{u}$, $g_\varepsilon = \eta_\varepsilon * g$):

$$\begin{aligned} - \int_{\Omega} g \operatorname{div} \bar{z} dx &= - \int_{\Omega} g \bar{u} dx + \int_{\Omega} (g)^2 dx = \\ &= - \int_{\Omega} g \bar{u}_\varepsilon dx + \int_{\Omega} g(\bar{u}_\varepsilon - \bar{u}) dx + \int_{\Omega} (g)^2 dx \\ &= - \int_{\Omega} g \operatorname{div} \bar{z}_\varepsilon dx + \int_{\Omega} g(\bar{u}_\varepsilon - \bar{u}) dx + \int_{\Omega} g(g - g_\varepsilon) dx. \end{aligned}$$

One has $\|\bar{u}_\varepsilon - \bar{u}\|_{L^1} \leq C\varepsilon$ where C depends on $|D\bar{u}|(\Omega)$, hence

$$\begin{aligned} - \int_{\Omega} \frac{1}{2} |\operatorname{div} \bar{z}|^2 + g \operatorname{div} \bar{z} dx &\leq - \int_{\Omega} \frac{1}{2} |\operatorname{div} \bar{z}_\varepsilon|^2 + g \operatorname{div} \bar{z}_\varepsilon dx + \|g\|_{L^\infty} (C\varepsilon + \|g - g_\varepsilon\|_{L^1}) \\ &\leq - \int_{\Omega} \frac{1}{2} |\operatorname{div} \bar{z}_{\varepsilon,h}|^2 + g_h \operatorname{div} \bar{z}_{\varepsilon,h} dx + \frac{1}{2} \int_{\Omega} (g - g_h)^2 dx + \|g\|_{L^\infty} (C\varepsilon + \|g - g_\varepsilon\|_{L^1}) \end{aligned}$$

where for the last estimate we have argued as in (33). Reasoning then as in (34)–(35), we find:

$$\begin{aligned} \int_{\Omega} |D\bar{u}| + \frac{1}{2} \int_{\Omega} (\bar{u} - g)^2 dx &\leq J_h^0(\hat{u}_h^0) + \frac{1}{2} \int_{\Omega} (\hat{u}_h^0 - g)^2 dx \\ &\quad + C \frac{h}{\varepsilon} + C\varepsilon + C\|g - g_\varepsilon\|_{L^1}, \quad (37) \end{aligned}$$

now C depending on the minimal energy, η , and $\|g\|_\infty$. The error estimate now depends on the initial smoothness of g and the speed at which it is approximated by g_ε , as for instance in [47] where a few cases are discussed. If, to simplify, we assume $g \in BV(\Omega)$, then the last two terms are of the same order and the optimal choice of ε is of order \sqrt{h} , yielding the estimate

$$\|\bar{u}_h^{0,c} - \hat{u}_h^0\|_{L^2(\Omega)} \leq Ch^{1/4}, \quad (38)$$

obviously weaker than (36).

6. Numerical experiments

6.1. Comparisons to the state-of-the-art

In order to compare our proposed ACR scheme to the state-of-the-art, we have chosen four different discretization schemes.

First, we compare to the most simple and most commonly used scheme, which is based on forward differences on a regular grid, as in (2) (see for instance [20]). This scheme suffers from an anisotropy, which is induced by the bias of the forward differences, as shown in Figure 1. In our experiments we will refer to this schemes as forward differences (FD).

Second, we compare to a upwind forward-backward differences (UFBD) scheme, which borrows ideas from the discretization of Hamilton-Jacobi equations [21]. This scheme adaptively selects between forward and backward differences, based on the sign of the difference and hence is more isotropic but it suffers from the fact that $J(u) \neq J(-u)$.

Third, we compare to an ℓ_1 -type anisotropic total variation that uses non-local finite differences in order to approximate the isotropic total variation [30]. We choose a relatively large neighborhood of 16 to obtain competitive results and hence refer to this scheme as anisotropic finite differences (AFD16). Furthermore, we weight the contributions of the ℓ_1 norms such that we obtain exact values of the total variation for the edges directions supported by the 16-neighborhood.

Finally, we compare with a discrete total variation recently proposed in [27], which is based on a staggered grid discretization of the dual field and constraints first introduced in [36]. The optimization is based on the infimal-convolution representation of Condat in [27]. The method is based on the dual formulation and uses a more sophisticated application of the pointwise constraint of the dual variable based on averaging. In the primal formulation, the scheme can be written as the infimal convolution of three different discrete derivative operators. We will refer to this method as CONDAT.

All algorithms have been implemented in Matlab and are minimized using the first-order primal-dual algorithm [22]. The fastest scheme is FD due to its simplicity. UFBD is slower by about a factor of 2 as it is based on forward and backward differences. ACR and AFD16 have a comparable computational complexity among each other but are slower than FD by about a factor 4. The slowest method is CONDAT which is about a factor of 6 slower than FD. In order to make sure that all algorithms have converged to a solution with sufficient accuracy, we run them for many iterations (> 10000).

6.2. Rotational invariance

In our first experiment, we demonstrate the rotational invariance of the proposed ACR scheme by recovering (inpainting) straight discontinuities of various orientations in the image domain $\Omega = [0, 1]^2$. For this, we consider the characteristic function $\chi_{\{x \cdot \nu > a\}}(x)$ as a boundary condition, where $\nu \in \mathbb{S}^1$ is its normal and $a = (\frac{1}{2}, \frac{1}{2}) \cdot \nu$ is set such that the discontinuity passes through the center of the image domain. As before, we discretize

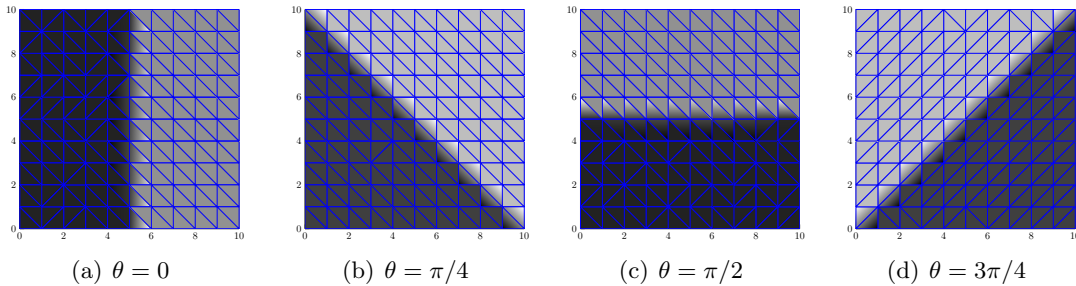


Figure 7: Recovery of a discontinuity with normal $\nu = (\cos \theta, \sin \theta)$ on a square grid of 10×10 pixels. We plot the image function which is piecewise affine on the triangles together with the triangulation shown in blue. Observe that the triangles are well aligned with the direction of discontinuity while in homogeneous regions the orientation of the triangles is ambiguous.

the domain Ω at scale $h = 1/N$ and identify the index set

$$\mathcal{B} = \{\mathbf{i} \in \mathcal{I} : \mathbf{i}h \in \partial\Omega\},$$

as the set of boundary points. At those points we compute the average values of the characteristic function along the edges $e_{\mathbf{i}}$ that is

$$g_{\mathbf{i}} = \frac{1}{|e_{\mathbf{i}}|} \int_{e_{\mathbf{i}}} \chi_{\{x \cdot \nu > a\}}(x) dx.$$

In order to use the values $g_{\mathbf{i}}$ as boundary condition we solve problem (23) using

$$G(v) = 0, \quad H(u) = \sum_{\mathbf{i} \in \mathcal{B}} \delta_{\{g_{\mathbf{i}}\}}(u_{\mathbf{i}}),$$

where $\delta_{\{c\}}(\cdot)$ denotes the indicator function of the singleton c .

We first demonstrate the property of our proposed ACR discretization scheme to adapt the triangulation to the direction of the discontinuity. For this we set $h = 1/10$, which yields a square grid of 10×10 pixels (squares). We set $\nu = (\cos \theta, \sin \theta)$, with $\theta \in \{0, \pi/4, \pi/2, 3\pi/4\}$ in order to recover discontinuities of the four main orientations. In Figure 7 we show the results for these four main orientations. We plot the continuous image function u , which is piecewise affine on the triangles, and the triangulation itself which is shown in blue. As expected the solution of the problem yields a straight discontinuity with the correct orientation. Observe that the triangulation is automatically adapted to match the orientation of the discontinuity. Moreover, one can also see that in the homogeneous regions the triangulation is ambiguous. Note that the value of the discrete total variation is equivalent to the true total variation, that is $J_h(u) = 10$ for $\pi \in \{0, \pi/2\}$ and $J_h(u) = 10\sqrt{2}$ for $\pi \in \{\pi/4, 3\pi/4\}$.

The aim of the second experiment is to evaluate the rotational invariance of the proposed discrete total variation and compare to other methods. For this we repeat the

θ	l	ACR	FD	UFBD	AFD16	CONDAT
$\pi/2$	100.00	100.00	100.00	100.00	100.00	100.00
$3\pi/8$	108.24	108.31	109.07	108.59	109.75	108.31
$\pi/4$	141.42	141.42	140.38	140.71	140.07	141.42

Table 1: This table compares the values of the discrete total variation $J_h(u)$ for the different schemes with the true value of the total variation which is equivalent to the length l of the discontinuity.

first experiment but now with orientations $\theta \in \{\pi/2, 3\pi/8, \pi/4\}$ and using a larger grid of 100×100 pixels and using $h = 1$. As the other approaches (FD, UFBD, AFD16, CONDAT) do not give explicit access to boundary points, we instead use a classical pixel-based boundary condition of sufficient width (5 pixels on each side). It is computed by first computing the discrete characteristic function on a 30 times finer grid and then downsampling the function to the desired resolution using averaging.

In Table 1, we compare the values of the discrete total variation computed by the different schemes. For $\theta = 0$ all schemes give the correct value of the discontinuity. As predicted by Proposition 3.3 the proposed ACR scheme is isotropic and hence very successful in recovering the correct value of the total variation for different orientations. Also CONDAT is very successful as it gives results of exactly the same quality. Note that for $\theta = 3\pi/8$, both ACR and CONDAT slightly overestimate the value of the total variation, see also Remark 4.1.

In Figure 8 we show the images corresponding to the experiments presented in Table (1). For the proposed ACR scheme we plot an image consisting of the center values u_c as they provide slightly sharper results. In general all schemes yield sharp discontinuities for $\theta = 0$, which is explained by the alignment of the discontinuity with the grid. FD yields very blurry results for $\theta = \{\pi/4, 3\pi/8\}$, because the forward differences would considerably overestimate the total variation of a sharp discontinuity (compare Figure 1). AFD16 works well for $\theta = \pi/4$ but produces strong artifacts for $\theta = 3\pi/8$. This is explained by the fact that only orientations which are supported by the neighborhood system can be recovered. UFBD yields good results for $\theta = \pi/4$ but gives slightly more blurry results for $\theta = 3\pi/8$. ACR yields good results for all orientations, with a small degree of blur for $\theta = 3\pi/8$. The sharpest transitions are recovered by CONDAT.

6.3. A segmentation problem

In the next experiment we are considering the following geometric minimization problem:

$$\min_u \lambda |Du|(\Omega) + \int_{\Omega} u(x)(|x| - R)dx, \quad \text{s.t. } 0 \leq u(x) \leq 1 \quad \forall x \in \Omega \quad (39)$$

where $R > 0$ is some parameter and we assume that $0 \in \Omega$. Observe that the level lines of the term $(|x| - R)$ are circles around the origin. Hence, for not too large $\lambda > 0$, the minimizer will be the characteristic function of a disk with radius $r > 0$. Though it can

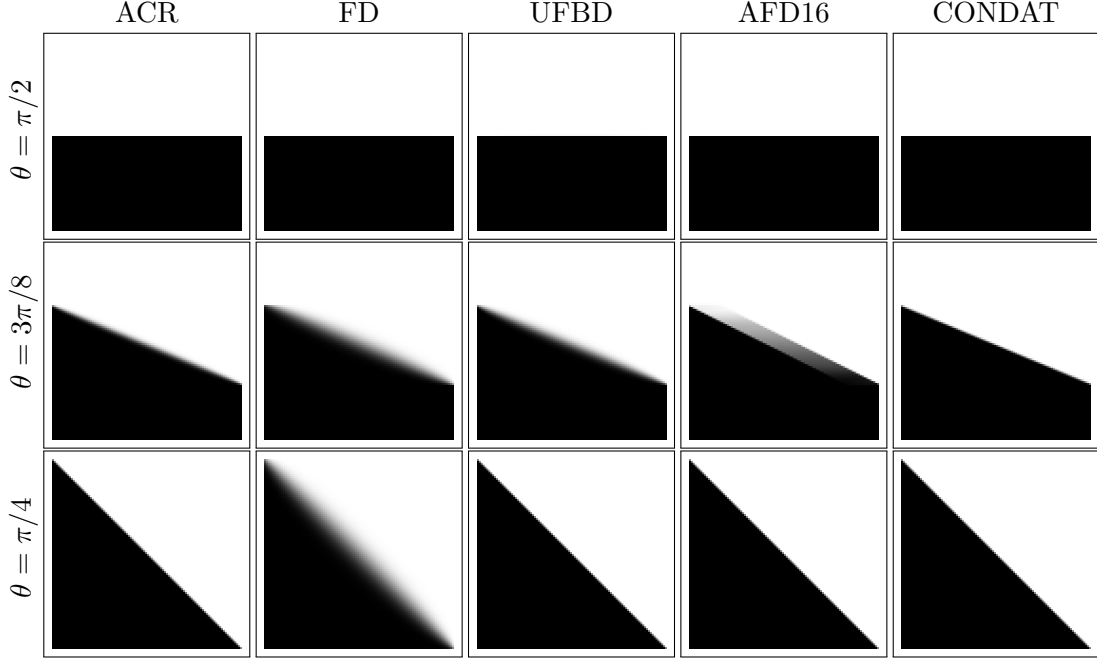


Figure 8: Images corresponding to the experiments of Table 1. Note the significantly more blurry results of FD and the “failure” of AFD16 to recover a straight line with angle $\theta = 3\pi/8$.

be recovered as the sublevel R of the function in [19, Eq. (39)], for $N = 2$ and $h = \lambda$, we now derive its expression for the reader’s convenience.

Using polar coordinates (ρ, θ) the variational problem can be re-written as

$$\min_{r \geq 0} 2\lambda r\pi + \int_0^{2\pi} \int_0^r \rho(\rho - R)d\rho d\theta = 2\pi \min_{r \geq 0} \lambda r + \frac{r^3}{3} - R\frac{r^2}{2}. \quad (40)$$

The minimal r , if positive, must satisfy $\lambda + r^2 - Rr = 0$, so that:

$$r = \frac{R + \sqrt{R^2 - 4\lambda}}{2}.$$

The disk of radius r solves the problem as long as its energy (40) is nonpositive, since otherwise it is energetically more favorable to choose $r = 0$. The value (40) is less or equal than zero as long as

$$\frac{3R - \sqrt{9R^2 - 48\lambda}}{4} \leq r \leq \frac{3R + \sqrt{9R^2 - 48\lambda}}{4}$$

Comparing this bound with the minimizing radius r , we see that we need to ensure

$$\frac{R + \sqrt{R^2 - 4\lambda}}{2} \leq \frac{3R + \sqrt{9R^2 - 48\lambda}}{4}.$$

λ	p	ACR	FD	UFBD	AFD16	CONDAT
25	150.52	150.46	151.79	149.19	151.38	150.31
50	143.31	143.29	143.64	141.74	143.38	143.39
100	125.66	125.28	124.72	122.56	124.07	125.83

Table 2: For a range of different regularization parameters λ , we compute the perimeter $p = 2\pi r$ of the true minimizing disk and compare it to the values of the discrete total variation obtained from different discretization schemes.

Solving for equality we find the upper bound on λ as

$$\lambda = \frac{3}{16}R^2,$$

which corresponds to a radius $r = \frac{3}{4}R$.

The aim of our experiment is now to numerically compute the solution of (39) using our proposed discretization and existing discretization schemes and compare it to the true solution. For comparison we use the value of the total variation which is equivalent of the perimeter $p = 2\pi r$ of the disk. We perform a numerical experiment on a grid of $N \times N$ pixels with $N = 100$, $h = 1$, approximating the image domain Ω and we set $R = N/4 = 25$. The data term and boundary conditions in the variational model (23) are given by

$$G(v) = \sum_{\mathbf{i} \in \mathcal{I}_c} v_{\mathbf{i}} w_{\mathbf{i}}, \quad H(u) = \sum_{\mathbf{i} \in \mathcal{I}} \delta_{[0,1]}(u_{\mathbf{i}})$$

with segmentation “weight” $w_{\mathbf{i}} = (|\mathbf{i}h| - R)$ and $\delta_{[0,1]}(\cdot)$ denotes the indicator function of the interval $[0, 1]$. Here, we shall assume that the values of the index set are properly shifted such that $0 \in \mathcal{I}_c$ is in the middle of the domain. From our above computation we know that the problem has a non-trivial minimizer as long as $\lambda \in [0, \frac{3}{16}R^2] \approx [0, 117.18]$.

For the other schemes (FD, UFBD, AFD16, CONDAT) we use exactly the same data fidelity term and the bound constraint is directly applied to the image pixels.

Table 2 compares the values of the discrete total variation to the true perimeter of the disk for different values of the parameter $\lambda \in \{25, 50, 100\}$. From the results we see that both the proposed ACR scheme and CONDAT approximate the true perimeter quite well. FD seems to overestimate the TV for smaller values of λ and underestimate it for larger values of λ . AFD16 and UFBD generally underestimate the TV for all values of λ .

In Figure 9 we provide the images corresponding to the experiments conducted in Table 2. We can observe that ACR provides quite isotropic solutions with slightly blurred interfaces. FD shows the well-known anisotropic smoothing behavior in the four quadrants. UFBD is more isotropic but also shows blurry interfaces. AFD16 provides very sharp interfaces but approximates the disc – as expected – by a 16-polygon. Visually, CONDAT provides the sharpest results.

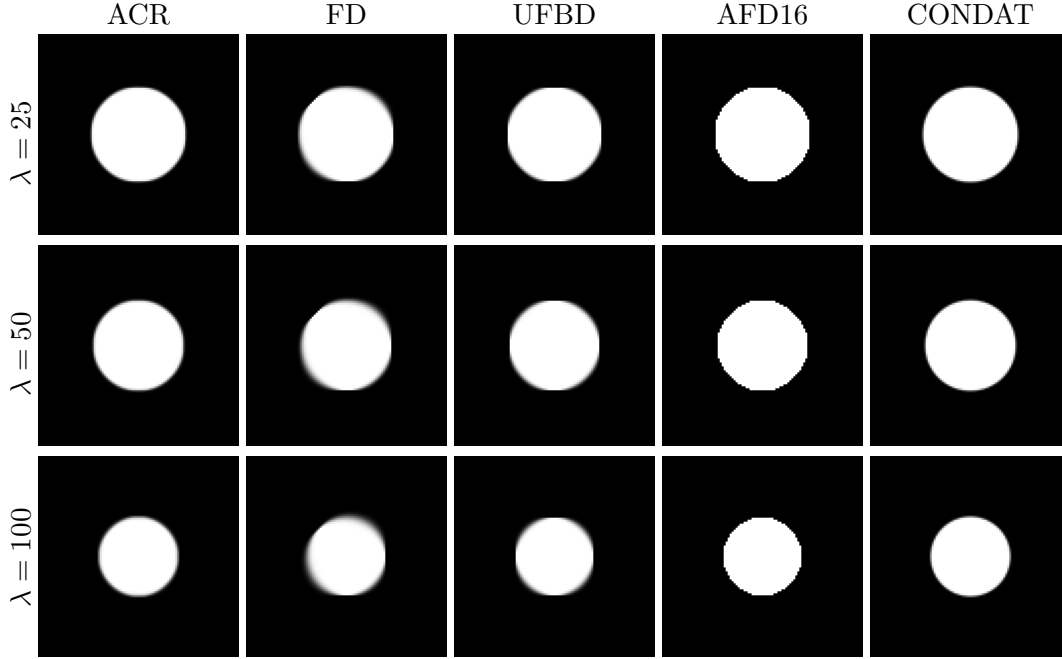


Figure 9: Image corresponding to the experiments in Table 2. Note the anisotropic behavior of FD and the polygonal behavior of AFD16.

6.4. Computing the ROF problem for a square

In the next example, we consider the problem of minimizing the ROF problem

$$\min_u \lambda |Du|(\Omega) + \frac{1}{2} \int_{\Omega} (u(x) - g(x))^2 dx,$$

where $g(x)$ is the characteristic function of a square of size L , that is $g(x) = \chi_{\{[-L/2, L/2]^2\}}(x)$. The image domain is set as $\Omega = [-L, L]^2$ and we are using Dirichlet zero boundary conditions. It is well-known that for $\lambda < \lambda^*$ with $\lambda^* = \frac{L}{2 + \sqrt{\pi}}$, the solution is given by a lower intensity square with rounded and blurred corners. The exact solution u_{λ} is given by an analytical formula, see [2, 27] for more details. Moreover, it can be shown that the value of the total variation of the solution u_{λ} is given by

$$TV(u_{\lambda}) = 4L \left(1 - \frac{\lambda}{\lambda^*}\right) - 2(4 - \pi)\lambda \log \frac{\lambda^*}{\lambda}.$$

Figure 10 plots the characteristic function of the square as well as exact solutions of the regularized squares for $\lambda \in \{2, 5, 10\}$.

In our experiments, we numerically compute the solutions of the regularized squares using different discretizations of the total variation and we compare it to the analytical solution. For this we set $L = 50$ and generate a grid of $2L \times 2L$ pixels to cover the image domain $\Omega = [-L, L]^2$. In order to obtain an accurate ground truth solution, we

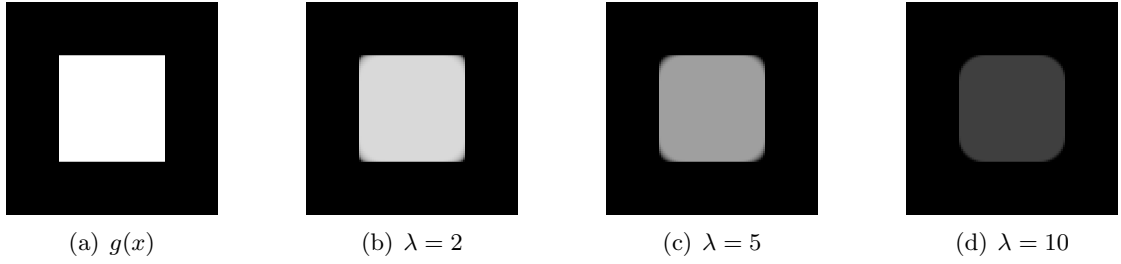


Figure 10: Original square $g(x)$, and ground truth solutions $u(x)$ for different values of λ .

λ	ACR	FD	UFBD	AFD16	CONDAT
2	3.08	6.21	9.35	3.50	8.93
5	3.54	10.95	20.58	5.30	9.14
10	2.82	10.67	26.69	6.79	5.03

Table 3: ℓ_1 -errors between the analytical solutions of the ROF problem and the discrete solutions for different values of the regularization parameter λ .

first compute the analytical solution on a 30 times finer grid (similar to [27]) and then downsample the image to the target resolution using averaging. For solving the ROF problem using the proposed ACR method, we use a quadratic data fidelity term defined on the average pixels:

$$G(v) = \sum_{\mathbf{i} \in \mathcal{I}_c} (v_{\mathbf{i}} - g_{\mathbf{i}})^2,$$

where $g_{\mathbf{i}}$ is the pixel-averaged discrete version of the function $g(x) = \chi_{\{[-L/2, L/2]^2\}}(x)$. For the other methods we use exactly the same data fidelity term.

In Table 3 we give the ℓ_1 errors between the pixel-averaged analytical solution and the discrete solutions computed by different discretization schemes of the total variation. For ACR we use the pixel center values u_c as they give sharper results.

One can see that the proposed ACR scheme gives the smallest ℓ_1 errors. Surprisingly, the AFD16 scheme gives also quite low error rates. The worst results are provided by the UFBD scheme. In Figure 11 we show the corresponding error images. One can clearly see that UFBD has a significantly larger global error. The errors at the corners varies between the different methods, but one can also see the strong anisotropic behavior of FD. Interestingly, the errors of CONDAT are mainly concentrated at the edges which is explained by the fact that the discretization scheme of CONDAT is based on pixel averages.

Table 4 finally compares the values of the total variation with respect to the true values of the total variation. ACR is slightly worse compared to CONDAT for $\lambda \in \{2, 5\}$ but gives the best results for $\lambda = 10$.

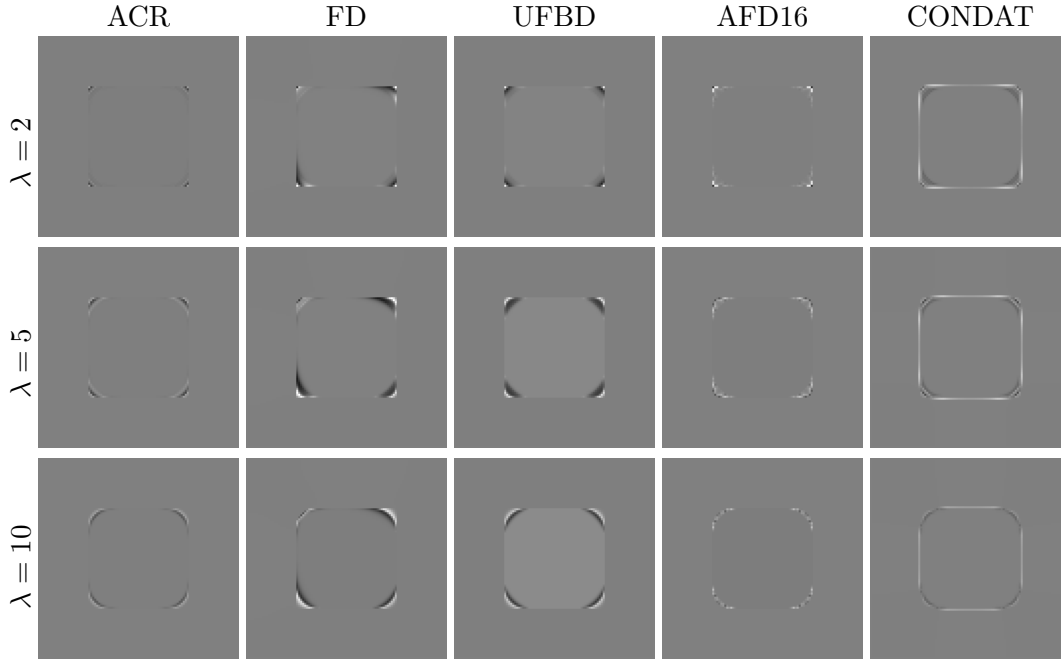


Figure 11: Error plots of the numerical solutions with respect to the analytical solution. For better comparison, the images are truncated to the range $[-0.1, 0.1]$ and gray corresponds to zero. Note the anisotropic behavior of FD and the significantly larger global error of UFBD.

λ	TV	ACR	FD	UFBD	AFD16	CONDAT
2	163.33	162.96	163.41	161.72	163.57	163.30
5	116.18	116.02	116.48	115.80	116.35	116.21
10	44.27	44.26	44.20	45.41	44.13	44.29

Table 4: Total variation of the different discretization schemes compared to the true total variation (TV) of the regularized square.

6.5. An example of image denoising

Finally, we present an examples where we remove the noise of a gray scale image. We use the same discrete ROF model as described in the previous section. Figure 12 shows the noisy clock image together with its ROF denoised version. We also provide detail views of the center pixel values, the average pixel values and the adaptive triangulation's. Note that the center pixel values yield slightly sharper results but can also contain some isolated pixels. The plots based on the triangulation nicely show how the mesh adapts to the structures. We also conducted comparisons with other existing discretization schemes, but we omit the results here because they were visually almost identical. The reason is that in case of image denoising the data term is dominating and hence differ-

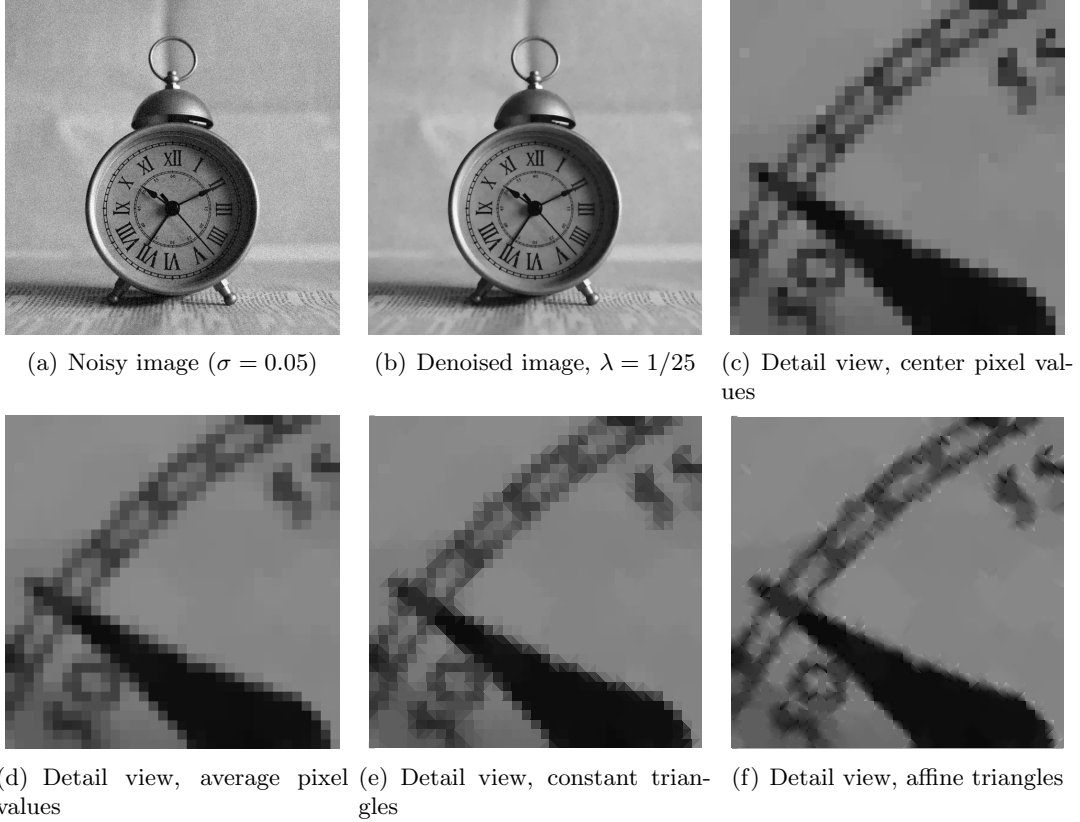


Figure 12: Image denoising using the proposed adaptive Crouzeix-Raviart discretization scheme. In (a), the noisy input image ($\sigma = 0.05$), in (b) the ROF denoised image. Image (c) shows the average pixel values and (d) the center pixel values for a detail of (b). Images (e), (f) show the average values (e) and the piecewise affine CR function (f) defined on the adaptive triangulation.

ences in the discretization of the regularizer only have a minor influence on the solution. However, we point out that for more geometric problems, as the problems we have presented in the previous sections, the discretization of the regularizer of course plays an important role.

References

- [1] Gabriel Acosta, Thomas Apel, Ricardo G. Durán, and Ariel L. Lombardi. Error estimates for Raviart-Thomas interpolation of any order on anisotropic tetrahedra. *Math. Comp.*, 80(273):141–163, 2011.
- [2] F. Alter, V. Caselles, and A. Chambolle. Evolution of characteristic functions of

- convex sets in the plane by the minimizing total variation flow. *Interfaces and Free Boundaries*, 7(1), 2005.
- [3] Luigi Ambrosio, Nicola Fusco, and Diego Pallara. *Functions of bounded variation and free discontinuity problems*. Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press, New York, 2000.
- [4] G. Anzellotti. Pairings between measures and bounded functions and compensated compactness. *Ann. Mat. Pura Appl. (4)*, 135:293–318 (1984), 1983.
- [5] Sören Bartels. Total variation minimization with finite elements: convergence and iterative solution. *SIAM J. Numer. Anal.*, 50(3):1162–1180, 2012.
- [6] Sören Bartels. Error control and adaptivity for a variational model problem defined on functions of bounded variation. *Math. Comp.*, 84(293):1217–1240, 2015.
- [7] 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.
- [8] Sören Bartels, Ricardo H. Nochetto, and Abner J. Salgado. A total variation diminishing interpolation operator and applications. *Math. Comp.*, 84(296):2569–2587, 2015.
- [9] Benjamin Berkels, Alexander Effland, and Martin Rumpf. A posteriori error control for the binary Mumford-Shah model. *Math. Comp.*, 86(306):1769–1791, 2017.
- [10] Y. Boykov and V. 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.
- [11] Y. Boykov, V. Kolmogorov, D. Cremers, and A. Delong. An integral solution to surface evolution PDEs via Geo-Cuts. In A. Leonardis, H. Bischof, and A. Pinz, editors, *European Conference on Computer Vision (ECCV)*, volume 3953 of *LNCS*, pages 409–422, Graz, Austria, May 2006. Springer.
- [12] Andrea Braides. *Γ -convergence for beginners*, volume 22 of *Oxford Lecture Series in Mathematics and its Applications*. Oxford University Press, Oxford, 2002.
- [13] Susanne C. Brenner. Forty years of the Crouzeix-Raviart element. *Numer. Methods Partial Differential Equations*, 31(2):367–396, 2015.
- [14] Susanne C. Brenner and L. Ridgway Scott. *The mathematical theory of finite element methods*, volume 15 of *Texts in Applied Mathematics*. Springer, New York, third edition, 2008.
- [15] 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.

- [16] Corentin Caillaud and Antonin Chambolle. Error estimates for finite differences approximations of the total variation. in preparation, 2020.
- [17] C. Carstensen and D. J. Liu. Nonconforming FEMs for an optimal design problem. *SIAM J. Numer. Anal.*, 53(2):874–894, 2015.
- [18] V. Caselles, A. Chambolle, and M. Novaga. The discontinuity set of solutions of the TV denoising problem and some extensions. *Multiscale Model. Simul.*, 6(3):879–894, 2007.
- [19] Vicent Caselles and Antonin Chambolle. Anisotropic curvature-driven flow of convex sets. *Nonlinear Anal.*, 65(8):1547–1577, 2006.
- [20] A. Chambolle. An algorithm for total variation minimization and applications. *J. Math. Imaging Vision*, 20(1-2):89–97, 2004. Special issue on mathematics and image analysis.
- [21] 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.
- [22] 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.
- [23] Antonin Chambolle and Thomas Pock. On the ergodic convergence rates of a first-order primal-dual algorithm. *Mathematical Programming*, pages 1–35, 2015. (online first).
- [24] 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.
- [25] Antonin Chambolle and Thomas Pock. An introduction to continuous optimization for imaging. *Acta Numer.*, 25:161–319, 2016.
- [26] Antonin Chambolle, Pauline Tan, and Samuel Vaiter. Accelerated alternating descent methods for Dykstra-like problems. *Journal of Mathematical Imaging and Vision*, Mar 2017.
- [27] Laurent Condat. Discrete total variation: new definition and minimization. *SIAM J. Imaging Sci.*, 10(3):1258–1290, 2017.
- [28] M. Crouzeix and P.-A. Raviart. Conforming and nonconforming finite element methods for solving the stationary Stokes equations. I. *Rev. Française Automat. Informat. Recherche Opérationnelle Sér. Rouge*, 7(R-3):33–75, 1973.

- [29] 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.
- [30] Jérôme Darbon and Marc Sigelle. Exact optimization of discrete constrained total variation minimization problems. In *Combinatorial image analysis*, volume 3322 of *Lecture Notes in Comput. Sci.*, pages 548–557. Springer, Berlin, 2004.
- [31] 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.
- [32] C. M. Elliott and S. A. Smitheman. Numerical analysis of the TV regularization and H^{-1} fidelity model for decomposing an image into cartoon plus texture. *IMA J. Numer. Anal.*, 29(3):651–689, 2009.
- [33] Xiaobing Feng and Andreas Prohl. Analysis of total variation flow and its finite element approximations. *M2AN Math. Model. Numer. Anal.*, 37(3):533–556, 2003.
- [34] Xiaobing Feng, Markus von Oehsen, and Andreas Prohl. Rate of convergence of regularization procedures and finite element approximations for the total variation flow. *Numer. Math.*, 100(3):441–456, 2005.
- [35] 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.
- [36] 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.
- [37] Qianying Hong, Ming-Jun Lai, Leopold Matamba Messi, and Jingyue Wang. Galerkin method with splines for total variation minimization. *J. Algorithms Comput. Technol.*, 13:16, 2019.
- [38] Clemens Kirisits, Christiane Pöschl, Elena Resmerita, and Otmar Scherzer. Finite-dimensional approximation of convex regularization via hexagonal pixel grids. *Appl. Anal.*, 94(3):612–636, 2015.
- [39] 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.
- [40] 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.
- [41] Christoph Ortner. Nonconforming finite-element discretization of convex variational problems. *IMA J. Numer. Anal.*, 31(3):847–864, 2011.

- [42] 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.
- [43] P.-A. Raviart and J. M. Thomas. A mixed finite element method for 2nd order elliptic problems. pages 292–315. Lecture Notes in Math., Vol. 606, 1977.
- [44] S. 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.
- [45] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. “GrabCut”: Interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.*, 23(3):309–314, August 2004.
- [46] L. Rudin, S. J. Osher, and E. 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)].
- [47] 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.
- [48] Xianmin Xu and Duvan Henao. An efficient numerical method for cavitation in nonlinear elasticity. *Math. Models Methods Appl. Sci.*, 21(8):1733–1760, 2011.

A. Proof of the duality Theorem 1

The goal of this section is to provide a proof of the duality theorem, Theorem 1. We prove a more general result for the CR approximation of Sobolev semi-norms.

A.1. Almost constant Crouzeix-Raviart functions

To simplify we drop the scale parameter h which is useless in this section. We consider a mesh \mathcal{T} of d -dimensional simplices in a polygonal domain $\Omega \subset \mathbb{R}^d$. We denote $V(\mathcal{T})$ the nonconforming P1 functions (Crouzeix-Raviart), defined in (10). We then let

$$V^0(\mathcal{T}) = \left\{ u \in CR(\mathcal{T}) : \int_T u(x)dx = u(c_T) = 0 \quad \forall T \in \mathcal{T} \right\}.$$

It is the space of the functions which are 0 on average in each simplex (in other words, the average of the middle values of the facets vanishes, equivalently the value at the center of each simplex is zero).

We define $P0(\mathcal{T})^d \approx (\mathbb{R}^d)^\mathcal{T}$ as usual as the space of “P0” functions which are constant on each $T \in \mathcal{T}$. Endowed with the topology of $L^2(\Omega; \mathbb{R}^d)$, it is a Euclidean space with the weighted scalar product: for $p, q \in P0(\mathcal{T})$,

$$\int_{\Omega} p(x)q(x)dx = \sum_{T \in \mathcal{T}} |T| p_T \cdot q_T.$$

Then we consider the gradients

$$GV^0(\mathcal{T}) = \{Du : u \in V^0(\mathcal{T})\} \subset P0(\mathcal{T}).$$

We want to characterize this space and its orthogonal. In order to do this, we consider the space $RT0(\mathcal{T})$ of the first order Raviart-Thomas vector fields subject to the mesh \mathcal{T} (cf Section 2.3, these are defined by their fluxes through the edges of the elements $T \in \mathcal{T}$). As before we also let $RT0_0(\mathcal{T}) \subset RT0(\mathcal{T})$ the RT0 fields with zero flux through $\partial\Omega$.

We know that, cf Lemma 2.4:

$$\{Du : u \in V(\mathcal{T})\} = \left\{ p \in P0(\mathcal{T})^d : \int_{\Omega} p(x) \cdot \sigma(x) dx = 0 \ \forall \sigma \in RT0_0(\mathcal{T}), \operatorname{div} \sigma = 0 \right\}.$$

More generally, if $u \in V(\mathcal{T})$ and $\sigma \in RT0_0(\mathcal{T})$, one has thanks to (11):

$$\int_{\Omega} \sigma(x) \cdot Du(x) dx = - \int_{\Omega} \operatorname{div} \sigma(x) u(x) dx.$$

As Du and $\operatorname{div} \sigma$ are constant on each triangle, this can also be written:

$$\sum_{T \in \mathcal{T}} |T| \sigma(c_T) \cdot Du(T) = - \sum_{T \in \mathcal{T}} |T| \operatorname{div} \sigma(T) u(c_T) \quad (41)$$

where c_T is the center of mass of the simplex T (so that given any affine function $a(x)$, $\int_T a(x) dx = |T| a(c_T)$).

Hence in particular, for any $\mathbf{p} \in GV^0(\mathcal{T})$ and $\sigma \in RT0_0(\mathcal{T})$, $\int_{\Omega} \sigma \cdot \mathbf{p} dx = 0$. A natural question is whether this is an if and only if; that is, if the orthogonal of $\{(\sigma(c_T))_{T \in \mathcal{T}} \in P0(\mathcal{T}) : \sigma \in RT0_0(\mathcal{T})\}$ is $GV^0(\mathcal{T})$.

Assume $\mathbf{p} \in P0(\mathcal{T})^d$ is such that $\int_{\Omega} \sigma \cdot \mathbf{p} dx = 0$ for all $\sigma \in RT0_0(\mathcal{T})$. Then in particular it is orthogonal to fields with zero divergence and there exists $u \in V(\mathcal{T})$ with $Du = \mathbf{p}$, thanks to Lemma 2.4. In particular because of (41) we have

$$0 = - \sum_{T \in \mathcal{T}} |T| \operatorname{div} \sigma(T) u(c_T)$$

for all Raviart-Thomas field σ (vanishing on $\partial\Omega$). Now, given any inner facet $S \subset \partial T \cap \partial T'$, $T, T' \in \mathcal{T}$, $T \neq T'$. we can introduce the field σ with flux 1 through S from T to T' and zero through all other facets. Then the formula becomes

$$0 = u(c_{T'}) - u(c_T).$$

This shows that u must take the same value in all the centers of the vertices. As u is defined up to a constant (in each connected component of Ω) we can assume this value is zero, and we have shown that in $P0(\mathcal{T})^d$,

$$\{(\sigma(c_T))_{T \in \mathcal{T}} : \sigma \in RT0_0(\mathcal{T})\}^{\perp} = GV^0(\mathcal{T}). \quad (42)$$

In particular, as a consequence, also

$$GV^0(\mathcal{T})^{\perp} = \{(\sigma(c_T))_{T \in \mathcal{T}} : \sigma \in RT0_0(\mathcal{T})\}. \quad (43)$$

A.2. Duality for discrete Sobolev semi-norms

Let us introduce, for $u \in V(\mathcal{T})$, $p \in [1, \infty)$,

$$J_p(u) = \frac{1}{p} \int_{\Omega} |Du|^p dx \quad (44)$$

and, for $\bar{u} \in P0(\mathcal{T})$

$$J_p^0(\bar{u}) = \min \{ J_p(u) : u \in V(\mathcal{T}), u(c_T) = \bar{u}_T \forall T \in \mathcal{T} \}. \quad (45)$$

By a slight abuse of notation we also let $J_p^0(u) = J_p^0((u(c_T))_T)$ for $u \in V(\mathcal{T})$. Then:

Theorem 2. For all $p \in (1, \infty)$ and any $u \in V(\mathcal{T})$,

$$J_p^0(u) = \sup \left\{ \int_{\Omega} \sigma(x) \cdot Du(x) dx - \frac{1}{p'} \sum_{T \in \mathcal{T}} |T| |\sigma(c_T)|^{p'} : \sigma \in RT0_0(\mathcal{T}) \right\}, \quad (46)$$

where $p' = p/(p-1)$, and for $p = 1$,

$$J_p^0(u) = \sup \left\{ \int_{\Omega} \sigma(x) \cdot Du(x) dx : \sigma \in RT0_0(\mathcal{T}), |\sigma(c_T)| \leq 1 \right\}. \quad (47)$$

In particular, Theorem 1 corresponds to the particular case $p = 1$.

Proof. We consider the case $p > 1$. The case $p = 1$ is then recovered as a limit problem. The “ \geq ” inequality is obvious. To show the reverse, we assume that u is a minimizer in (45). Then, for any $v \in V^0(\mathcal{T})$ (cf Sec. A.1), $u + v$ is admissible in the problem and one has $J_p(u + v) \geq J_p(u)$. Hence, taking the derivative $\lim_{t \downarrow 0} (J_p(u + tv) - J_p(u))/t$, it follows that

$$\int_{\Omega} |Du|^{p-2} Du \cdot Dv dx = 0$$

for all $v \in V^0(\mathcal{T})$. That is, the field $(|Du(T)|^{p-2} Du(T))_{T \in \mathcal{T}} \in P0(\mathcal{T})$ is orthogonal to $GV^0(\mathcal{T})$, hence thanks to (43), there exists $\sigma \in RT0_0(\mathcal{T})$ such that $\sigma(c_T) = |Du(T)|^{p-2} Du(T)$ for all $T \in \mathcal{T}$. Clearly, $|\sigma(c_T)|^{p'} = |Du(T)|^p$ and the conclusion easily follows. The case $p = 1$ can be recovered as follows: one builds, for $p > 1$, a field $\sigma_p \in RT0_0(\mathcal{T})$ optimal in (46), and letting then $p \rightarrow 1$, one checks that it converges to a field which is optimal in (47). \square

Remark A.1. It is quite easy to derive, as a more general result, that given $f : \mathbb{R}^d \rightarrow \mathbb{R}$ a convex, lower semicontinuous function and any $\bar{u} \in P0(\mathcal{T})$ one has

$$\begin{aligned} & \inf \left\{ \int_{\Omega} f(\nabla u) dx : u \in V(\mathcal{T}), u(c_T) = \bar{u}_T \forall T \in \mathcal{T} \right\} \\ & = \sup \left\{ - \int_{\Omega} \bar{u} \operatorname{div} \sigma dx - \sum_{T \in \mathcal{T}} |T| f^*(\sigma(c_T)) : \sigma \in RT0_0(\mathcal{T}) \right\}. \end{aligned}$$

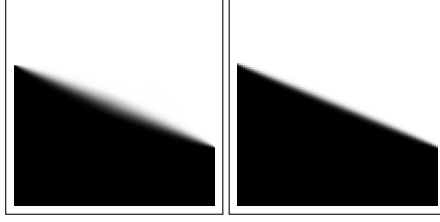


Figure 13: Left: the variant (49), right: the ACR result of Fig. 8.

B. A variant with one node per pixel

For imaging application, one drawback of our approach could be the need to introduce more nodes in the representation than the number of pixels.

Given a (grey-level) $n \times m$ image $(u_{i,j})_{i=1,\dots,n}^{j=1,\dots,m}$ (to simplify, we assume that the scale $h = 1$ in this section), even if one rotates the grid by 45° and considers the pixels (i, j) as centers of edges of larger squares (for instance, $(1, 1), (1, 2), (2, 1), (2, 2)$ would be the centers of the edges of the square $[(3/2, 1/2), (5/2, 3/2), (3/2, 5/2), (1/2, 3/2)]$), one still needs to introduce an additional node in the center of each square (in the above example, at $(3/2, 3/2)$) and introduce fictitious values $u_{i+1/2, j+1/2}$ (i, j both even or both odd) at these nodes. On average, this increases the dimension of the problems by roughly 50%.

Unfortunately, it seems there is no simple strategy to eliminate this additional node. To illustrate this issue, let us first concentrate on one square. We consider the four vertices $\{0, 1\}^2$ as the middle points of the edges of the square (of area 2)

$$C = [(\frac{1}{2}, -\frac{1}{2}), (\frac{3}{2}, \frac{1}{2}), (\frac{1}{2}, \frac{3}{2}), (-\frac{1}{2}, \frac{1}{2})]$$

and a fifth vertex in $(1/2, 1/2)$ in the middle, which is the middle of both the vertical and horizontal edges cutting the square into two halves. Then, given the values $u_{\alpha, \beta} := u(\alpha, \beta)$, $(\alpha, \beta) \in \{0, 1\}^2$, and c the value at the center, the Crouzeix-Raviart total variation inside the square is

$$\max \left\{ \sqrt{2} \sqrt{(u_{00} - c)^2 + (u_{10} - c)^2} + \sqrt{2} \sqrt{(u_{11} - c)^2 + (u_{01} - c)^2}, \right. \\ \left. \sqrt{2} \sqrt{(u_{00} - c)^2 + (u_{01} - c)^2} + \sqrt{2} \sqrt{(u_{11} - c)^2 + (u_{10} - c)^2} \right\}.$$

(Each gradient norm is multiplied by the area 1 of the corresponding triangle, and we have used that the distance between a vertex of the cube and the center is $1/\sqrt{2}$.) A possibility to eliminate c is to minimize this quantity with respect to c . In the “inpainting” problems of Fig. 8, this would give the same results (since this is precisely what is done automatically by the minimization in this case). Unfortunately we have, at this point, no idea of how to solve this problem explicitly. It means that to compute the “proximity” operator of the corresponding energy on a whole image, we need to solve subproblems which keep the additional central variable.

A simpler possibility is to first optimize with respect to the value c and then, afterwards, pick the best orientation. In that case, one needs to solve

$$\sqrt{2} \max \left\{ \min_c \sqrt{(u_{00} - c)^2 + (u_{10} - c)^2} + \sqrt{(u_{11} - c)^2 + (u_{01} - c)^2}, \right. \\ \left. \min_c \sqrt{(u_{00} - c)^2 + (u_{01} - c)^2} + \sqrt{(u_{11} - c)^2 + (u_{10} - c)^2}, \right\}.$$

A careful analysis shows that this value is given by the function

$$J^4((u_{00}, u_{10}, u_{01}, u_{11})) := \\ \sqrt{2} \max \left\{ \sqrt{(u_{11} - u_{00})^2 + (u_{10} - u_{01})^2}, \sqrt{(u_{01} - u_{00})^2 + (u_{11} - u_{10})^2}, \right. \\ \left. \sqrt{(u_{10} - u_{00})^2 + (u_{11} - u_{01})^2} \right\}. \quad (48)$$

One can use (48) to define, given u defined by its pixel values $(u_{i,j})_{i=1,\dots,n}^{j=1,\dots,m}$, a discrete total variation as

$$J(u) := \sum_{(i,j) \text{ even}} J^4((u_{i,j}, u_{i+1,j}, u_{i,j+1}, u_{i+1,j+1})) \\ + \sum_{(i,j) \text{ odd}} J^4((u_{i,j}, u_{i+1,j}, u_{i,j+1}, u_{i+1,j+1})). \quad (49)$$

We remark this is a variant of the energy defined in [24] (see also [26] for a theoretical study), which can be optimized by an efficient alternating descent method as soon as one knows how to solve explicitly the subproblems given by the proximity operator of J^4 , on each square.

Unfortunately, our implementation shows that it does not perform as well as the ACR technique introduced in this paper. Figure 13 compares this to the ACR result in Fig. 8: we obtain a very diffusive solution, with practically no improvement over a non-optimized Crouzeix Raviart implementation.

On the other hand, as is expected, this approximation (which in any case is still based on a hidden, underlying Crouzeix-Raviart discretization), yields to a quite precise approximation of the energy and is a reasonable regularizer for standard inverse problems, *cf.* Fig. 14.

C. The proximity operator of (22)

We describe in this Section how to implement the proximity operator of the function f in (22). The problem we need to solve is as follows, given $\bar{\xi} = (\bar{\xi}_{mn})_{m=1,\dots,4}^{n=1,2} \in \mathbb{R}^{4 \times 2}$ and $\tau > 0$:

$$\min_{\xi = (\xi_{mn})_{m=1,\dots,4}^{n=1,2} \in \mathbb{R}^{4 \times 2}} f(\xi) + \frac{1}{2\tau} \|\xi - \bar{\xi}\|^2. \quad (50)$$



Figure 14: Left: original “clock” image, middle: with a Gaussian blur of std. dev. 1.5 and a 1% Gaussian noise, right: TV-regularised deblurred image with (49).

We call $\text{prox}_{\tau f}(\bar{\xi})$ the solution of (50). We recall that the prox of the convex conjugate f^* is also easily recovered, once (50) is solved, using Moreau’s identity:

$$x = \text{prox}_{\tau f}(x) + \tau \text{prox}_{\frac{1}{\tau} f^*}\left(\frac{x}{\tau}\right).$$

To solve (50) we first make the following obvious observation: denoting

$$\begin{aligned} x_1 &= \sqrt{\xi_{1,1}^2 + \xi_{2,1}^2}, & (\xi_{1,1}, \xi_{2,1})^T &= x_1 \eta_1, \\ x_2 &= \sqrt{\xi_{3,1}^2 + \xi_{4,1}^2}, & (\xi_{3,1}, \xi_{4,1})^T &= x_2 \eta_2, \\ x_3 &= \sqrt{\xi_{1,2}^2 + \xi_{2,2}^2}, & (\xi_{1,2}, \xi_{2,2})^T &= x_3 \eta_3, \\ x_4 &= \sqrt{\xi_{3,2}^2 + \xi_{4,2}^2}, & (\xi_{3,2}, \xi_{4,2})^T &= x_4 \eta_4, \end{aligned}$$

(and the same for $\bar{\xi}$), it is equivalent to solve:

$$\min_{(x_i) \geq 0, (\eta_i)} \max\{|x_1| + |x_2|, |x_3| + |x_4|\} + \frac{1}{2\tau} \sum_{i=1}^4 |x_i \eta_i - \bar{x}_i \bar{\eta}_i|^2.$$

We obtain at the minimum that $\eta_i = \bar{\eta}_i$, $i = 1, \dots, 4$ and the problem boils down to

$$\min_{x=(x_i)_{i=1}^4 \in \mathbb{R}_+^4} \max\{|x_1| + |x_2|, |x_3| + |x_4|\} + \frac{1}{2\tau} |x - \bar{x}|^2$$

where $|x - \bar{x}|^2 = \sum_{i=1}^4 |x_i - \bar{x}_i|^2$. Remark that here, $\bar{x}_i \geq 0$ and it is equivalent to look for $x \in \mathbb{R}_+^4$ or in \mathbb{R}^4 .

We now explain how to solve this 4-dimensional convex problem. We can rewrite it as

$$\min_x \max_{\substack{\mu_{12} + \mu_{34} = 1 \\ \mu_{12} \geq 0, \mu_{34} \geq 0}} \mu_{12}(|x_1| + |x_2|) + \mu_{34}(|x_3| + |x_4|) + \frac{1}{2\tau} |x - \bar{x}|^2$$

and then we exchange min and max. We obtain 4 problems of the form

$$\min_{x_1} \mu_{12}|x_1| + \frac{1}{2\tau}|x_1 - \bar{x}_1|^2.$$

This is well known to be solved by $x_1 = (\bar{x}_1 - \tau\mu_{12})^+$ and with value

$$\mu_{12}(\bar{x}_1 - \tau\mu_{12})^+ + \frac{1}{2\tau} \begin{cases} |\bar{x}_1|^2 & \text{if } \bar{x}_1 \leq \tau\mu_{12} \\ |\tau\mu_{12}|^2 & \text{else.} \end{cases}$$

When $\bar{x}_1 \leq \tau\mu_{12}$, this is $|\bar{x}_1|^2/(2\tau)$, otherwise

$$\mu_{12}\bar{x}_1 - \frac{\tau}{2}|\mu_{12}|^2 = \frac{1}{2\tau}|\bar{x}_1|^2 - \frac{1}{2\tau}|\bar{x}_1 - \tau\mu_{12}|^2.$$

We end up with the dual problem

$$\max_{\substack{\mu_{12} + \mu_{34} = 1 \\ \mu_{12} \geq 0, \mu_{34} \geq 0}} \frac{1}{2\tau} \left(\sum_{i=1}^4 |\bar{x}_i|^2 - (|\bar{x}_1 - \tau\mu_{12}|^2 + |\bar{x}_2 - \tau\mu_{12}|^2 + |\bar{x}_3 - \tau\mu_{34}|^2 + |\bar{x}_4 - \tau\mu_{34}|^2) \right),$$

whose optimality reads, if $0 < \mu_{12} < 1$,

$$(\bar{x}_1 - \tau\mu_{12})^+ + (\bar{x}_2 - \tau\mu_{12})^+ = (\bar{x}_3 - \tau\mu_{34})^+ + (\bar{x}_4 - \tau\mu_{34})^+$$

with $\mu_{34} = 1 - \mu_{12}$.

Without loss of generality, assume that $\bar{x}_2 \geq \bar{x}_1$ and $\bar{x}_4 \geq \bar{x}_3$. We recast the problem as

$$\min_{0 \leq \mu \leq 1} |(\bar{x}_1 - \tau\mu)^+|^2 + |(\bar{x}_2 - \tau\mu)^+|^2 + |(\bar{x}_3 - \tau + \tau\mu)^+|^2 + |(\bar{x}_4 - \tau + \tau\mu)^+|^2$$

by letting $\mu := \mu_{12}$ and $\mu_{34} = 1 - \mu$.

By convexity of the objective, $\mu \in [0, 1]$ is optimal if and only if:

$$\begin{cases} (\bar{x}_1 - \tau\mu)^+ + (\bar{x}_2 - \tau\mu)^+ - (\bar{x}_3 - \tau + \tau\mu)^+ - (\bar{x}_4 - \tau + \tau\mu)^+ \leq 0 & \text{if } \mu < 1; \\ (\bar{x}_1 - \tau\mu)^+ + (\bar{x}_2 - \tau\mu)^+ - (\bar{x}_3 - \tau + \tau\mu)^+ - (\bar{x}_4 - \tau + \tau\mu)^+ \geq 0 & \text{if } \mu > 0. \end{cases} \quad (51)$$

Hence, one sees that if one knows which term are positive in the above sums, μ is found by solving the above equations with “= 0” instead of “ $\geq / \leq 0$ ” and then projecting the value onto the interval $[0, 1]$. For instance, if all values are positive,

$$\mu = \left(0 \vee \frac{\bar{x}_1 + \bar{x}_2 - \bar{x}_3 - \bar{x}_4 + 2\tau}{4\tau} \right) \wedge 1. \quad (52)$$

Whenever $\mu \in (0, 1)$, of course, (51) reads

$$(\bar{x}_1 - \tau\mu)^+ + (\bar{x}_2 - \tau\mu)^+ = (\bar{x}_3 - \tau + \tau\mu)^+ + (\bar{x}_4 - \tau + \tau\mu)^+. \quad (53)$$

Hence, the problem is solved by exhaustion of the following cases:

1. if $\bar{x}_2 + \bar{x}_4 \leq \tau$, then clearly one can find $\mu \in [0, 1]$ such that all terms of the sums in (51) are zero, hence the solution is $x_1 = x_2 = x_3 = x_4 = 0$.
2. if $\bar{x}_2 + \bar{x}_4 > \tau$ then:
 - a) either *both* $\bar{x}_2 - \tau\mu > 0$ and $\bar{x}_4 - \tau + \tau\mu > 0$,
 - b) or one side of (53) is zero so that one must be in a case of strict inequality in (51), and $\mu \in \{0, 1\}$.

The second case 2b can be first easily eliminated by checking whether $\mu = 0$ or $\mu = 1$ is a solution of the optimality condition: one has

$$\begin{aligned}\bar{x}_1 + \bar{x}_2 &\leq (\bar{x}_3 - \tau)^+ + (\bar{x}_4 - \tau)^+ \Leftrightarrow \mu = 0, \\ (\bar{x}_1 - \tau)^+ + (\bar{x}_2 - \tau)^+ &\geq \bar{x}_3 + \bar{x}_4 \Leftrightarrow \mu = 1.\end{aligned}$$

3. Otherwise, we must be in the first case 2a, where $\bar{x}_2 - \tau\mu > 0$ and $\bar{x}_4 - \tau + \tau\mu > 0$, equality (53) holds, and which is then split into four possible cases:
 - a) μ given by (52), and $\bar{x}_1 \geq \tau\mu$, $\bar{x}_3 \geq \tau(1 - \mu)$, then $x_1 = \bar{x}_1 - \tau\mu$, $x_2 = \bar{x}_2 - \tau\mu$, $x_3 = \bar{x}_3 - \tau(1 - \mu)$, $x_4 = \bar{x}_4 - \tau(1 - \mu)$;
 - b) $\mu = \frac{\bar{x}_1 + \bar{x}_2 - \bar{x}_4 + \tau}{3\tau}$ and $\bar{x}_1 \geq \tau\mu$, $\bar{x}_3 \leq \tau(1 - \mu)$, $\bar{x}_3 \geq \tau(1 - \mu)$, then $x_1 = \bar{x}_1 - \tau\mu$, $x_2 = \bar{x}_2 - \tau\mu$, $x_3 = 0$, $x_4 = \bar{x}_4 - \tau(1 - \mu)$;
 - c) $\mu = \frac{\bar{x}_2 - \bar{x}_3 - \bar{x}_4 + 2\tau}{3\tau}$ and $\bar{x}_1 \leq \tau\mu$, $\bar{x}_2 \geq \tau\mu$, $\bar{x}_3 \geq \tau(1 - \mu)$, then $x_1 = 0$, $x_2 = \bar{x}_2 - \tau\mu$, $x_3 = \bar{x}_3 - \tau(1 - \mu)$, $x_4 = \bar{x}_4 - \tau(1 - \mu)$;
 - d) $\mu = \frac{\bar{x}_2 - \bar{x}_4 + \tau}{2\tau}$ if all the previous cases fail to hold, and then $x_1 = x_3 = 0$, $x_2 = \bar{x}_2 - \tau\mu$, $x_4 = \bar{x}_4 - \tau(1 - \mu)$.