



HAL
open science

Error estimates for finite differences approximations of the total variation

Corentin Caillaud, Antonin Chambolle

► **To cite this version:**

Corentin Caillaud, Antonin Chambolle. Error estimates for finite differences approximations of the total variation. 2020. hal-02539136v1

HAL Id: hal-02539136

<https://hal.science/hal-02539136v1>

Preprint submitted on 9 Apr 2020 (v1), last revised 7 Oct 2021 (v2)

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.

Error estimates for finite differences approximations of the total variation

Corentin Caillaud Antonin Chambolle

CMAP, Ecole Polytechnique, CNRS, Institut Polytechnique de Paris, Palaiseau, France

April 8, 2020

We present a convergence rate analysis of the Rudin-Osher-Fatemi (ROF) denoising problem for two different discretizations of the total variation. The first discretization is the well-known isotropic total variation that suffers from a blurring effect in a special diagonal direction. We prove that in the setting corresponding to this direction, the discrete ROF energy converges to the continuous one in $O(h^{2/3})$. The second total variation is based on Raviart-Thomas fields and achieves a $O(h)$ convergence rate for the same quantity under some standard hypotheses.

Keywords: total variation, finite differences, finite elements, error bounds, convergence rates, image denoising

MSC (2020): 35A21 35A35 65D18 65N06 65N15 65N30

1 Introduction

Since its introduction by Rudin, Osher and Fatemi in [15] the use of total variation as a regularizer for denoising and inverse problems has proven to be effective in removing noise while preserving sharp edges. In the continuous setting, the denoising model consists in solving the “ROF” problem:

$$\bar{u} = \arg \min_{u \in BV \cap L^2(\Omega)} \frac{1}{2\lambda} \|u - g\|_{L^2}^2 + \text{TV}(u) =: E(u) \quad (1)$$

where $\Omega = [0, 1] \times [0, 1]$ is the domain of our noisy image g (for which we will suppose $g \in L^\infty(\Omega)$ as well as $g \in BV(\Omega)$ when needed) and $\lambda > 0$ is a regularizing parameter. Here TV stands for the continuous total variation given by $\text{TV}(u) = \int_\Omega |\nabla u|$ when u is regular, and with $|\cdot|$ denoting the euclidean norm in \mathbb{R}^2 .

To be more precise, we consider both Neumann and Dirichlet boundary conditions to this setting. This will result in two different variants of (1): in the Neumann setting

$$\begin{aligned} \bar{u}_N &= \arg \min_{u \in BV \cap L^2(\Omega)} \frac{1}{2\lambda} \|u - g\|_{L^2}^2 + \text{TV}_N(u) =: E_N(u) \text{ where} & (2) \\ \text{TV}_N(u) &= \sup \left\{ - \int_{\Omega} u \operatorname{div} \phi ; \phi \in \mathcal{C}_c^1(\Omega, \mathbb{R}^2) \text{ s.t. } \|\phi\|_{\infty} \leq 1 \right\} \end{aligned}$$

where $\mathcal{C}_c^1(\Omega, \mathbb{R}^2)$ is the space of continuously differentiable and compactly supported fields from Ω to \mathbb{R}^2 , and $\|\phi\|_{\infty} = \sup_{x \in \Omega} |\phi(x)|$. In the Dirichlet setting, we add the constraint that $u = b$ on $\partial\Omega$ for some $b \in BV \cap L^{\infty}(\partial\Omega)$ (naturally, one usually takes $b = g|_{\partial\Omega}$), and replace TV_N by

$$\text{TV}(u) = \sup \left\{ - \int_{\Omega} u \operatorname{div} \phi + \int_{\partial\Omega} u \langle \phi | \vec{n} \rangle ; \phi \in \mathcal{C}^1(\Omega, \mathbb{R}^2) \text{ s.t. } \|\phi\|_{\infty} \leq 1 \right\}$$

where \vec{n} denotes the outer normal unit vector. Equivalently, we formulate the Dirichlet problem as:

$$\begin{aligned} \bar{u}_D &= \arg \min_{u \in BV \cap L^2(\Omega)} \frac{1}{2\lambda} \|u - g\|_{L^2}^2 + \text{TV}_D(u) =: E_D(u) \text{ where} & (3) \\ \text{TV}_D(u) &= \sup \left\{ - \int_{\Omega} u \operatorname{div} \phi + \int_{\partial\Omega} b \langle \phi | \vec{n} \rangle ; \phi \in \mathcal{C}^1(\Omega, \mathbb{R}^2) \text{ s.t. } \|\phi\|_{\infty} \leq 1 \right\} \end{aligned}$$

In the following, we will denote for $B \in \{N, D\}$ the optimal value of the continuous problems $\bar{E}_B = E_B(\bar{u}_B)$. When no subscript (N or D) is used, it means our statement is valid under both boundary conditions.

In practice, Ω is discretized into $N \times N$ square pixels of size $h = 1/N$, namely $\Omega = \cup_{1 \leq i, j \leq N} C_{i,j}$ with $C_{i,j} = [(i-1)h, ih] \times [(j-1)h, jh]$. Images are now elements of $P0 = \{u : \Omega \rightarrow \mathbb{R} \text{ s.t. } \forall 1 \leq i, j \leq N, \exists u_{i,j} \in \mathbb{R} \text{ s.t. } u = u_{i,j} \text{ in } C_{i,j}\}$. One introduces the projection of the continuous image $g^h = \Pi_{P0}(g)$ given by $(g^h)_C = \frac{1}{h^2} \int_C g$ for every square pixel C , and the discrete counterpart of (1) is the following:

$$\bar{u}^h = \arg \min_{u^h \in P0} \frac{1}{2\lambda} \|u^h - g^h\|_{L^2}^2 + \text{TV}^h(u^h) =: E^h(u^h) \quad (4)$$

where TV^h is some discretization of the total variation defined on $P0$. In the Dirichlet setting, TV^h can involve the discretization b^h of b given by $(b^h)_e = \frac{1}{h} \int_e b$ for every boundary edge e . This article deals with the study of the convergence rate of $\bar{E}^h := E^h(\bar{u}^h)$ towards \bar{E} for two different discretizations TV^h .

A widely used choice for TV^h is the so called ‘‘isotropic’’ total variation which discretizes the expression $\text{TV}(u) = \int_{\Omega} |\nabla u|$ using a finite difference operator D . It is given by

$$\text{TV}_i^h(u^h) = h \sum_{1 \leq i, j \leq N} |(Du^h)_{i,j}| \text{ where } (Du^h)_{i,j} = \begin{pmatrix} u_{i+1,j}^h - u_{i,j}^h \\ u_{i,j+1}^h - u_{i,j}^h \end{pmatrix} \quad (5)$$

(with either $u_{N+1,j}^h = b_{N+\frac{1}{2},j}^h$, $u_{i,N+1}^h = b_{i,N+\frac{1}{2}}^h$ in the Dirichlet boundary conditions or $u_{N+1,j}^h - u_{N,j}^h = u_{i,N+1}^h - u_{i,N}^h = 0$ in the Neumann boundary conditions). The term “isotropic” refers to the behavior of this functional as the mesh size h tends to zero. One can indeed show (see [8] where this is proven for a more complicated TV) that the functional $u \mapsto \text{TV}_i^h(u^h)$ if $u = u^h \in P_0, +\infty$ otherwise Γ -converges to TV, so that the minimizers \bar{u}^h converge (for instance in L^2) to \bar{u} , the minimizer of (1). This convergence leads to saying that TV_i^h inherits of the isotropy of TV for denoising problems such as ROF. We recall below the standard example of this isotropy of the continuous total variation: the denoising of the characteristic of a half plane in the Dirichlet setting.

Given a direction $\nu \in \mathbb{R}^2$ with $|\nu| = 1$, take $g = g_\nu$ defined by $g_\nu(x) = 1$ if $\langle x|\nu \rangle \geq a$ and $g_\nu(x) = 0$ otherwise where a is some fixed real number (for instance $a = \langle (1/2, 1/2)|\nu \rangle$). Then, problem (3) with boundary condition $b = g_\nu|_{\partial\Omega}$ has solution $\bar{u}_D = g_\nu$, no matter the orientation of ν . This comes from the following important fact¹:

Claim. Fix $\nu \in \mathbb{R}^2$ with $|\nu| = 1$. When using the boundary condition $b = g_\nu|_{\partial\Omega}$, the value of $\text{TV}_D(g_\nu)$ is reached for $\phi \equiv \nu$ so that $\text{TV}_D(g_\nu) = \int_{\partial\Omega} g_\nu \langle \nu|\bar{n} \rangle$.

Our claim is indeed a direct consequence: if $u \in BV \cap L^2(\Omega)$ is such that $u|_{\partial\Omega} = g_\nu|_{\partial\Omega}$, taking the admissible field $\phi \equiv \nu$ gives

$$\text{TV}_D(u) \geq \int_{\partial\Omega} g_\nu \langle \nu|\bar{n} \rangle = \text{TV}_D(g_\nu)$$

and the result follows.

However, this convergence result does not guarantee the isotropy of the discrete isotropic TV itself. In fact $\text{TV}_i^h(g_\nu^h)$ can be quite far from the length of the continuous line $\text{TV}(g_\nu)$. What is worse is that the value of $\text{TV}_i^h(g_\nu^h)$ actually depends on the orientation of ν . The case of the 45° diagonal is eloquent: as noted for instance in [7], the choice of the finite difference operator D induces a difference of roughly 40% between the main diagonal, that is $\nu = \frac{1}{\sqrt{2}}(1, 1)$ and its flipped version that is $\nu = \frac{1}{\sqrt{2}}(-1, 1)$:

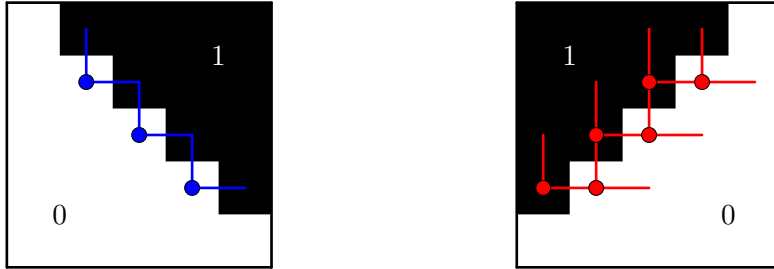


Figure 1: On the left image $\text{TV}_i^h \simeq N\sqrt{2}$ while on the right $\text{TV}_i^h \simeq 2N$

This differentiation breaks the isotropy of TV_i^h for a fixed $h > 0$ leading to artefacts depending on the direction in denoising problems such as the denoising of a circle: the

¹Which just follows from $-\int_{\Omega} g_\nu \text{div} \phi + \int_{\partial\Omega} g_\nu \langle \phi|\bar{n} \rangle = \int_{\{x \text{ s.t. } \langle x|\nu \rangle = a\}} \langle \nu|\phi \rangle$

edges oriented along the more penalized diagonal are blurred (see Figure 2). Going back to the case $g = g_\nu$, even if one always has $\bar{u}_h \rightarrow g_\nu$, the speed of this convergence may vary with ν . We take again the example of the two mirror diagonals for which denoising with TV_i^h for different step sizes h are shown Figure 2. One notices that the denoising is achieved correctly for the 135° diagonal \triangleleft (which we will now call consequently the “good” diagonal) whereas one needs to take h very small before obtaining a sharp looking discontinuity with the other diagonal \triangleleft (the “bad” one).

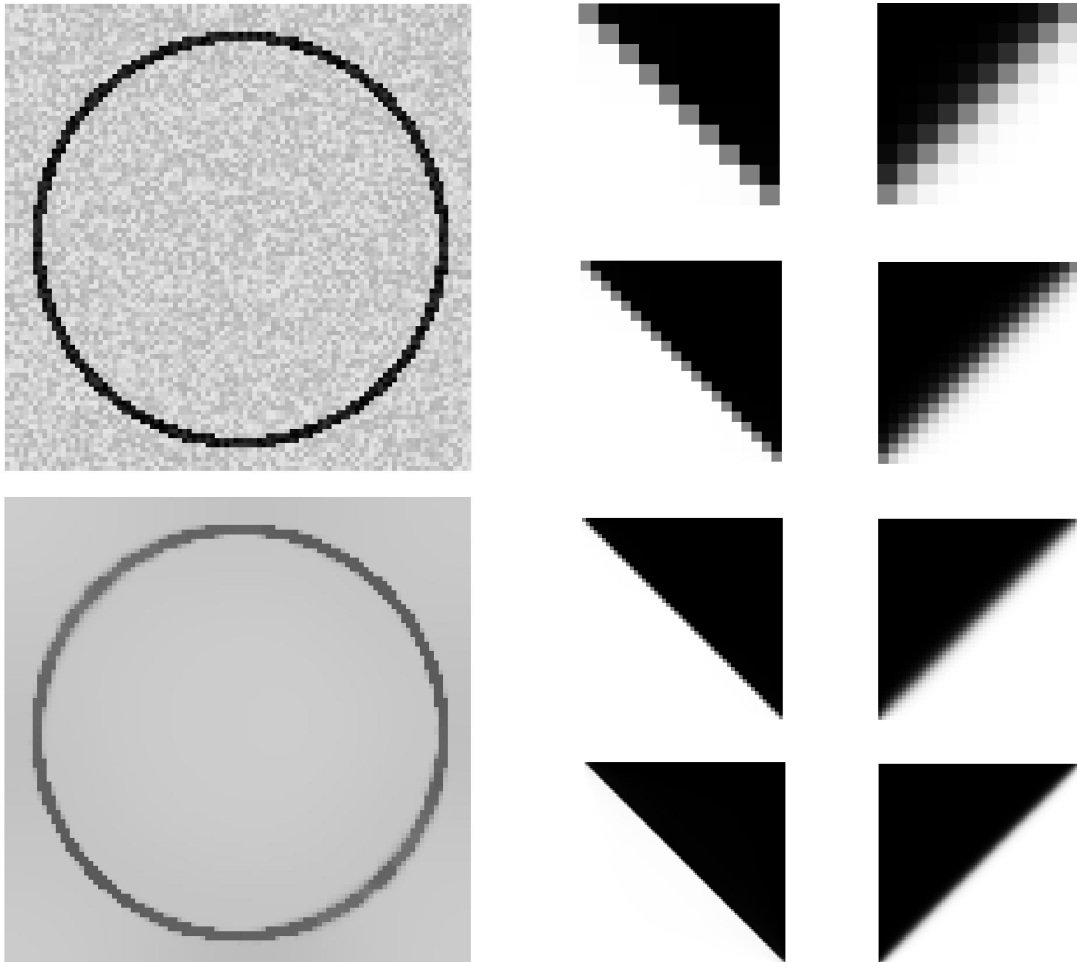


Figure 2: Denoising with TV_i^h : noisy and denoised circle with Neumann b.c., good (2nd col.) and bad (3rd col.) diagonals with Dirichlet b.c. and $N = 10, 20, 50, 100$.

The main purpose of this article is to study the error made by the isotropic total variation in the “bad” diagonal denoising problem. To this end, we estimate the convergence of the optimal discrete energy of problem (4) towards the optimal continuous energy in (1). Up to a slight change of the domain ($\Omega = \Omega_{per}$ that we will define later), we show

that it is of order $O(h^{2/3})$. More precisely we will prove the following theorem:

Theorem 1. *On an appropriate domain $\Omega = \Omega_{per}$ we have:*

1. For $\nu = \frac{1}{\sqrt{2}}(-1, 1)$ the denoising is exact, meaning that $\bar{u}^h = \Pi_{P_0}(\bar{u})$.
2. For $\nu = \frac{1}{\sqrt{2}}(1, 1)$, $\exists \underline{h}, c, c' > 0$ depending only on λ such that

$$\forall h \leq \underline{h}, \quad ch^{2/3} \leq \bar{E}^h - \bar{E} \leq c'h^{2/3}$$

These rates ought to be compared with results obtained by Lucier and co-authors in [12] and [16]. In [12] the authors give a bound of type $|\bar{E}^h - \bar{E}| \leq c\sqrt{h}$ (as well as $\|\bar{u}^h - \bar{u}\|^2 \leq c\sqrt{h}$) for a so called central-difference discretization of the ROF model meaning that they use the following discrete total variation $\text{TV}^h = \text{TV}_c^h$ with

$$\text{TV}_c^h(u^h) = h \sum_{i,j} \sqrt{\left(\frac{u_{i+1,j}^h - u_{i-1,j}^h}{2}\right)^2 + \left(\frac{u_{i,j+1}^h - u_{i,j-1}^h}{2}\right)^2}$$

In [16], errors in $h^{\frac{\alpha}{\alpha+1}}$ are given, where α is the Lipschitz order of g , and the discrete total variation at stake is an average of the four possible isotropic total variations obtained by the finite difference approximations of the gradient: forward/forward (which is TV_i^h), forward/backward, backward/forward and backward/backward.

The second part of this paper is also to be compared with these rates. We establish in our context the convergence rate (valid under some hypothesis) $|\bar{E}^h - \bar{E}| \leq ch$ for another discrete total variation $\text{TV}^h = \text{TV}_{RT}^h$. A similar error is obtained in [7] (see also [2]) for a non-conforming P1 finite-elements based approximation of the total variation. The idea behind the TV_{RT}^h total variation is to gain isotropy in the discretization of the continuous TV by allowing any direction ν to be an admissible discrete field ϕ . We propose to mimic the dual definition of the continuous total variation in the Dirichlet setting TV_D using Raviart-Thomas fields ($RT0$) [14] which are piecewise affine fields including the constant fields (precise definitions will be given later) and define for either continuous or discrete function and boundary term u, b :

$$\text{TV}_{RT,D}^h(u) = \sup \left\{ - \int_{\Omega} u \operatorname{div} \phi + \int_{\partial\Omega} b \langle \phi | \vec{n} \rangle; \phi \in RT0 \text{ s.t. } \|\phi\|_{\infty} \leq 1 \right\} \quad (6)$$

The fact that $\nu \in RT0$ allows one to do as above to show that this total variation is isotropic in the sense that when taking $b = g_{\nu}$ one has $\bar{u} = g_{\nu}$ for any ν in the following ROF model (mixing a discrete TV term to a continuous L^2 term):

$$\bar{u} = \arg \min_{\substack{u \in BV \cap L^2(\Omega) \\ u|_{\partial\Omega} = g_{\nu}|_{\partial\Omega}}} \frac{1}{2\lambda} \|u - g_{\nu}\|_{L^2}^2 + \text{TV}_{RT,D}^h(u)$$

In the discrete setting, we finally define our problem as:

$$\bar{u}^h = \arg \min_{u^h \in P0} \frac{1}{2\lambda} \|u^h - g_\nu^h\|_2^2 + \text{TV}_{RT}^h(u^h)$$

where TV_{RT}^h stands for $\text{TV}_{RT,D}^h$ given by (6) in the Dirichlet setting, and is replaced by

$$\text{TV}_{RT,N}^h(u^h) = \sup \left\{ - \int_{\Omega} u^h \operatorname{div} \phi, \phi \in RT0_0 \text{ s.t. } \|\phi\|_{\infty} \leq 1 \right\}$$

in the Neumann setting where we enforce the $RT0$ fields to vanish on the boundary obtaining the $RT0_0$ fields.

This total variation performs well on the denoising problem.

The paper is organised as follows: Sections 2,3 and 4 are devoted to TV_i^h denoising and prove theorem 1. Section 5 presents the Raviart-Thomas total variation and gives the $O(h)$ convergence rate for this setting. The last section presents the implementation of the algorithms, and shows numerical results comparing on test images the total variations TV_i^h , TV_{RT}^h and a state-of-the-art variant initially proposed in [11], and analysed and implemented by Condat [9] which seems to perform better, but for which we do not have error bounds up to now (nor even consistency, although this is addressed in a work in progress).

2 Reduction to a 1D TV denoising problem

To study the orientation dependent error of the isotropic TV, we introduce the following experiment. Placing ourselves in a well-chosen periodic domain $\Omega = \Omega_{per}$, we reduce the 2D TV_i^h denoising problem in the case of a diagonal image $g = g_\nu$ with $\nu = \frac{1}{\sqrt{2}}(-1, 1)$ to a 1D problem. In the following, we will denote respectively TV and tv the 2D and 1D total variations. The first point of theorem 1, which is the case $\nu = \frac{1}{\sqrt{2}}(1, 1)$, will be quickly obtained. We next present some general results about the case $\nu = \frac{1}{\sqrt{2}}(-1, 1)$ that will be useful to prove the second point of theorem 1 in the following sections.

2.1 The domain Ω_{per}

We actually do not consider the ROF model (4) on a square domain, but on a periodic strip oriented along the diagonal at stake, see the drawing below in which each square pixel is of size $h = 1/N$ and where the (green) dotted lines are to be identified. For $\nu = \frac{1}{\sqrt{2}}(-1, 1)$, we now work with a variable $u_{i,j}^h$ defined for $(i, j) \in \mathbb{Z}^2$ such that $-N \leq i - j \leq N$; $0 \leq i + j \leq D$ and satisfying $u_{i+D,j+D}^h = u_{i,j}^h$ for any (i, j) . Making the change of variables $n = i - j$; $d = \lfloor \frac{i+j}{2} \rfloor$, one checks that our domain can be represented by

$$\Omega_{per} = \{(n, d), -N \leq n \leq N, d \in \mathbb{Z}/D\mathbb{Z}\}$$

Our source term $g^h : \Omega_{per} \rightarrow \mathbb{R}$ is given by $g^h(n, d) = 1$ for $n > 0$, $g^h(n, d) = 0$ for $n < 0$ and $g^h(0, d) = 1/2$, for all $d \in \mathbb{Z}/D\mathbb{Z}$.

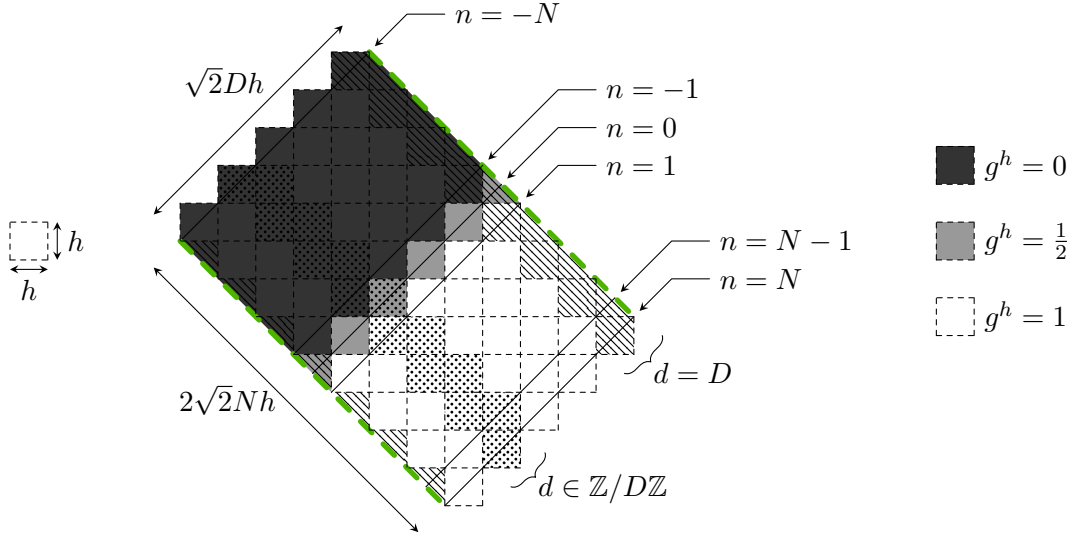


Figure 3: Setting for the lower error bound estimate.

Then the problem (4) is to solve

$$\bar{u}^h = \arg \min_{u^h: \Omega_{per} \rightarrow \mathbb{R}} \frac{h^2}{2\lambda} \sum_{(n,d) \in \Omega_{per}} |u^h(n,d) - g^h(n,d)|^2 + \text{TV}_i^h(u) := E^h(u^h)$$

where TV_i^h stands for the isotropic TV on this particular domain. If $n = i - j$ and $d = \lfloor \frac{i+j}{2} \rfloor$ so that $u^h(n,d)$ codes for the value of $u_{i,j}^h$, then one finds that $u_{i+1,j}^h$ (respectively $u_{i,j+1}^h$) is represented by $u^h(n+1, d')$ (respectively $u^h(n-1, d')$) with $d' = d$ for n even and $d' = d+1$ for n odd. Following (5), this leads to the following expression of the isotropic total variation:

$$\text{TV}_i^h(u^h) = h \sum_{\substack{d \in \mathbb{Z}/D\mathbb{Z} \\ n \text{ even}}} \left| \begin{array}{l} u^h(n+1, d) - u^h(n, d) \\ u^h(n-1, d) - u^h(n, d) \end{array} \right| + h \sum_{\substack{d \in \mathbb{Z}/D\mathbb{Z} \\ n \text{ odd}}} \left| \begin{array}{l} u^h(n+1, d+1) - u^h(n, d) \\ u^h(n-1, d+1) - u^h(n, d) \end{array} \right|$$

We will first study the case of Dirichlet boundary conditions meaning that we impose (both on the definition of TV_i^h and on the optimization problem) that for all $d \in \mathbb{Z}/D\mathbb{Z}$:

$$u^h(N+1, d) = u^h(N, d) = g^h(N, d) = 1 \quad ; \quad u^h(-N-1, d) = u^h(-N, d) = g^h(-N, d) = 0$$

Later on we will deduce from the Dirichlet setting the same rate for the Neumann boundary conditions:

$$u^h(N+1, d) = u^h(N, d) \quad ; \quad u^h(-N-1, d) = u^h(-N, d)$$

The benefit of this periodic setting is to reduce the problem from 2D to 1D as at the optimum one has $\bar{u}^h(n, d) = \bar{u}^h(n, d')$ for all n and $d, d' \in \mathbb{Z}/D\mathbb{Z}$. Indeed, as all

the terms in the objective are invariant when changing d to $d + 1$, the shifted image $\tilde{u}^h : (n, d) \mapsto \bar{u}^h(n, d + 1)$ has the same energy E^h , hence $\tilde{u}^h = \bar{u}^h$ by uniqueness of the optimizer.

We keep the letter u for this now 1D variable, and divide our energy by a factor $\sqrt{2}Dh$ which is the width of our 2D domain. The problem then rewrites as:

$$\bar{u}^h = \arg \min_{\substack{u^h \in \mathbb{R}^{2N+1} \\ \text{s.t. BC}}} E^h(u^h) := \frac{h}{2\sqrt{2}\lambda} \|u^h - g^h\|_2^2 + \text{tv}_i^h(u^h) \quad (7)$$

where we defined

$$\begin{cases} \|u^h - g^h\|_2^2 = \sum_{n=-N}^N (u_n^h - g_n^h)^2 \\ \text{tv}_i^h(u^h) = \frac{1}{\sqrt{2}} \sum_{n=-N}^N \sqrt{(u_{n+1}^h - u_n^h)^2 + (u_n^h - u_{n-1}^h)^2} \end{cases}$$

with $g_n^h = 0$ for $n < 0$, $g_n^h = 1$ for $n > 0$ and $g_0^h = 1/2$ and where BC stands for the following boundary conditions:

$$\begin{cases} u_{N+1}^h = u_N^h = 1 \text{ and } u_{-N-1}^h = u_{-N}^h = 0 \text{ for Dirichlet} \\ u_{N+1}^h = u_N^h \text{ and } u_{-N-1}^h = u_{-N}^h \text{ for Neumann} \end{cases}$$

This problem is therefore a 1D signal denoising relying on a biased 1D total variation $\frac{1}{\sqrt{2}} \sum_n \sqrt{(u_{n+1}^h - u_n^h)^2 + (u_n^h - u_{n-1}^h)^2}$. This bias is responsible for the bad behavior of TV_i^h on this diagonal.

As a comparison, with the other diagonal, we define similarly Ω_{per} to be oriented along this diagonal. One can check that this leads to a 1D denoising with the classical 1D discrete total variation $\text{tv}^h(u^h) = \sum_n |u_{n+1}^h - u_n^h|$. As a consequence, the denoising is exact: $\bar{u}^h = g^h$. Indeed, the problem (in the Dirichlet setting) is to minimize $\|u^h - g^h\|_2^2 + c \text{tv}^h(u^h)$ for some constant $c > 0$ and under the constraint that $u_{N+1}^h = u_N^h = 1$ and $u_{-N}^h = 0$. This constraint gives $\text{tv}^h(u^h) \geq \left| \sum_{n=-N}^N u_{n+1}^h - u_n^h \right| = 1 = \text{tv}^h(g^h)$, hence we obtain the first point of our main theorem 1.

2.2 Solution of the continuous limit problem

In this section we investigate the continuous 1D denoising problem obtained when passing to the limit $h \rightarrow 0$ in problem (7). Assuming u^h is the discretization of a smooth function u defined on $[-1, 1]$, we write:

$$\begin{aligned} E^h(u^h) &= \frac{1}{N\sqrt{2}} \sum_{n=-N}^N \frac{1}{2\lambda} (u(nh) - g_n^h)^2 \\ &\quad + \sqrt{\left(\frac{u(nh+h) - u(nh)}{h} \right)^2 + \left(\frac{u(nh) - u(nh-h)}{h} \right)^2} \end{aligned}$$

and we see that this converges as $h \rightarrow 0$ to

$$E(u) = \int_{-1}^1 \frac{1}{2\sqrt{2\lambda}}(u - g)^2 + |u'| \quad (8)$$

with $\int_{-1}^1 |u'| =: \text{tv}(u)$ being the continuous 1D total variation. It is easily shown that (8) is also the Gamma-limit of the discrete problem, so that the minimizers \bar{u}^h of (7) will converge to the minimizer of (8).

For the Dirichlet setting, we enforce the constraint $u = g$ at the boundary of the domain i.e. $u(-1) = 0$ and $u(1) = 1$. In that situation, for any admissible u we have:

$$\int_{-1}^1 |u'| \geq \left| \int_{-1}^1 u' \right| = |u(1) - u(-1)| = 1 = \int_{-1}^1 |g'|$$

which directly shows that the energy (8) is minimal for $u = g$ with value $\bar{E}_D = 1$.

In the Neumann setting however, no boundary condition is required. To find the solution, one can write the optimality conditions given by duality theory (see [4]):

$$\text{tv}(u) = - \int_{-1}^1 u z' \quad \text{and} \quad \frac{1}{\sqrt{2\lambda}}(u - g) - z' = 0$$

for some function z such that $|z| \leq 1$ and $z(-1) = z(1) = 0$. If these equations are met for some couple (u, z) then u is optimal in problem (8). We search for u of the form $u = u_a$ for some $a \in \mathbb{R}$ with $u_a(x) = a$ if $x \in (-1, 0)$ and $u_a(x) = 1 - a$ if $x \in (0, 1)$. This leads to taking $z(x) = \frac{a}{\sqrt{2\lambda}}(x+1)$ if $x \in (-1, 0)$ and $z(x) = \frac{a}{\sqrt{2\lambda}}(1-x)$ if $x \in (0, 1)$. Then one must try to fulfill the equations $\text{tv}(u_a) = - \int_{-1}^1 u_a z'$ that is $|1 - 2a| = \frac{1}{\sqrt{2\lambda}}a(1 - 2a)$ and $|z| \leq 1$ that is $|a| \leq \sqrt{2\lambda}$. These two equations on a always give rise to a unique solution: if $\lambda \leq \lambda^* := \frac{\sqrt{2}}{4}$ then u_a is optimal with $a = a_{opt} := \sqrt{2\lambda}$ and the minimal energy is $\bar{E}_N^{\leq} := 1 - \sqrt{2\lambda}$. If $\lambda > \lambda^*$ then u_a is optimal with $a = \frac{1}{2}$ and the minimal energy is $\bar{E}_N^{>} := \frac{1}{4\sqrt{2\lambda}}$. In the following, we will see that in the case $\lambda > \lambda^*$ the discrete problem is exact ($\bar{u}^h \equiv \frac{1}{2}$), therefore we will always place ourselves in the case $\lambda \leq \lambda^*$, and we denote $\bar{E}_N := \bar{E}_N^{\leq} = 1 - \sqrt{2\lambda}$.

2.3 Form of the solution

Before turning to the proof of the $O(h^{2/3})$ bounds, we make some general remarks on the form of the solution of (7):

Proposition 1. *The solution \bar{u}^h of problem (7) (either with Dirichlet or Neumann boundary conditions) satisfies:*

1. $\forall n, \bar{u}_{-n}^h = 1 - \bar{u}_n^h$, in particular $\bar{u}_0^h = \frac{1}{2}$.
2. $\forall n > 0, 1 \geq \bar{u}_n^h \geq \frac{1}{2}$, hence $\forall n < 0, 0 \leq \bar{u}_n^h \leq \frac{1}{2}$.

3. \bar{u}^h is non-decreasing: $\forall n, \bar{u}_{n+1}^h \geq \bar{u}_n^h$.

Proof. For the first point, the symmetry of g^h and tv_i^h yields that $\tilde{u}_n^h = 1 - \bar{u}_{-n}^h$ satisfies $E^h(\tilde{u}^h) = E^h(\bar{u}^h)$. By uniqueness of the minimizer, $\tilde{u}^h = \bar{u}^h$.

For the second point, the truncated variable $\hat{u}_n^h = \max(g_n^h, \min(\bar{u}_n^h, \frac{1}{2}))$ satisfies $|\hat{u}_n^h - g_n^h| \leq |\bar{u}_n^h - g_n^h|$ and $|\hat{u}_{n+1}^h - \hat{u}_n^h| \leq |\bar{u}_{n+1}^h - \bar{u}_n^h|$ for any n , hence $E^h(\hat{u}^h) \leq E^h(\bar{u}^h)$ and $\bar{u}^h = \hat{u}^h$.

For the third point, consider the staircase version of \bar{u}^h given by: $\check{u}_n^h = \max\{\bar{u}_k^h, 0 \leq k \leq n\}$ if $n > 0$, $\check{u}_0^h = \frac{1}{2}$ and $\check{u}_n^h = \min\{\bar{u}_k^h, n \leq k \leq 0\}$ if $n < 0$. As $\bar{u}_n^h \in [0, 1]$ we have $|\check{u}_n^h - g_n^h| \leq |\bar{u}_n^h - g_n^h|$, and again $|\check{u}_{n+1}^h - \check{u}_n^h| \leq |\bar{u}_{n+1}^h - \bar{u}_n^h|$ for any n , hence $E^h(\check{u}^h) \leq E^h(\bar{u}^h)$ and $\bar{u}^h = \check{u}^h$. \square

Proposition 2. We denote $\lambda^* = \frac{\sqrt{2}}{4}$. The solution \bar{u}^h of problem (7) is such that:

1. With Dirichlet boundary conditions, $\bar{u}_1^h > \frac{1}{2}$ for any λ .
2. With Neumann boundary conditions, $\bar{u}^h \equiv \frac{1}{2}$ for any $\lambda \geq \lambda^*$ and $\bar{u}_1^h > \frac{1}{2}$ for any $\lambda < \lambda^*$ for some λ_h^* such that $|\lambda_h^* - \lambda^*| \leq ch^{1/3}$ for some constant $c > 0$. In particular, for any $\lambda < \lambda^*$ one has $\bar{u}_1^h > \frac{1}{2}$ for h small enough.

Proof. For $u \in \mathbb{R}^{2N+1}$ satisfying the three properties of Proposition 1 and such that $u_1 = \frac{1}{2}$, we define $k \in \{1, \dots, N\}$ such that $u_{-1} = u_0 = \dots = u_k = \frac{1}{2}$ and $u_{k+1} > \frac{1}{2}$. Suppose first that $k \leq N - 2$ then the energy of u can be written

$$E^h(u) = \frac{h}{2\sqrt{2}\lambda}(u_k - 1)^2 + \frac{1}{\sqrt{2}}|u_k - \frac{1}{2}| + \frac{1}{\sqrt{2}}\sqrt{(u_{k+1} - u_k)^2 + (u_k - \frac{1}{2})^2} + \frac{1}{\sqrt{2}}\sqrt{(u_{k+2} - u_{k+1})^2 + (u_{k+1} - u_k)^2} + R(u)$$

where $R(u)$ does not depend on u_k . As $u_{k+1} > \frac{1}{2}$, we have the following derivatives or subgradients:

$$\begin{aligned} \frac{\partial}{\partial u_k} \left(\sqrt{(u_{k+1} - u_k)^2 + (u_k - \frac{1}{2})^2} \right) \Big|_{u_k = \frac{1}{2}} &= \left(\frac{(u_k - u_{k+1}) + (u_k - \frac{1}{2})}{\sqrt{(u_{k+1} - u_k)^2 + (u_k - \frac{1}{2})^2}} \right) \Big|_{u_k = \frac{1}{2}} = -1 \\ \frac{\partial}{\partial u_k} \left(\sqrt{(u_{k+2} - u_{k+1})^2 + (u_{k+1} - u_k)^2} \right) \Big|_{u_k = \frac{1}{2}} &= \frac{\frac{1}{2} - u_{k+1}}{\sqrt{(\frac{1}{2} - u_{k+1})^2 + (u_{k+2} - u_{k+1})^2}} = d < 0 \\ \frac{\partial}{\partial u_k} ((u_k - 1)^2) \Big|_{u_k = \frac{1}{2}} &= -1 \text{ and } \frac{\partial}{\partial u_k} (|u_k - \frac{1}{2}|) \Big|_{u_k = \frac{1}{2}} = [-1, 1] \end{aligned}$$

Finally $\frac{\partial E^h}{\partial u_k} \Big|_{u_k = \frac{1}{2}} = -\frac{h}{2\sqrt{2}\lambda} + \frac{1}{\sqrt{2}}[-1, 1] - \frac{1}{\sqrt{2}}(1 - d) \subset \mathbb{R}_*^-$ so that $0 \notin \frac{\partial E^h}{\partial u_k} \Big|_{u_k = \frac{1}{2}}$ hence u is not optimal. For $k = N - 1$ the same reasoning is correct in the Dirichlet setting

noting that $u_{k+2} = 1$ whereas in the Neumann setting it is changed to

$$E^h(u) = \frac{h}{2\sqrt{2}\lambda}(u_k - 1)^2 + \frac{1}{\sqrt{2}}|u_k - \frac{1}{2}| \\ + \frac{1}{\sqrt{2}}\sqrt{(u_{k+1} - u_k)^2 + (u_k - \frac{1}{2})^2} + \frac{1}{\sqrt{2}}|u_{k+1} - u_k| + R(u)$$

for which one computes $\frac{\partial E^h}{\partial u_k}|_{u_k=\frac{1}{2}} = -\frac{h}{2\sqrt{2}\lambda} + \frac{1}{\sqrt{2}}[-1, 1] - \frac{2}{\sqrt{2}} \subset \mathbb{R}_*^-$ and gets the same conclusion. This concludes the proof in the Dirichlet setting as in this case $k < N$.

In the Neumann setting, the case $k = N$ corresponds to our alternative $\bar{u}^h \equiv \frac{1}{2}$ so that we only have to exhibit an admissible u^h such that $E^h(u^h) < E(\frac{1}{2})$ to prove that $\bar{u}_1^h > \frac{1}{2}$. We postpone this construction to Section 3.3 where, provided that $\lambda < \lambda^*$, we will explicitly build a u^h such that $E^h(u^h) \leq 1 - \lambda\sqrt{2} + ch^{2/3}$ for some constant $c > 0$. In comparison the energy of the constant $u^h \equiv \frac{1}{2}$ is $E^h(\frac{1}{2}) = \frac{h}{2\sqrt{2}\lambda} \times 2N \times (\frac{1}{2})^2 = \frac{\sqrt{2}}{8\lambda}$. The conclusion comes from studying when $1 - \lambda\sqrt{2} + ch^{2/3} < \frac{\sqrt{2}}{8\lambda}$.

Finally, suppose now that $\lambda \geq \lambda^*$, we want to prove that $\bar{u}^h \equiv \frac{1}{2}$. For any $u \in \mathbb{R}^{2N+1}$ satisfying the three properties of Proposition 1, denoting $a = u_{-N} \in [0, \frac{1}{2}]$ we form the following estimate. On one hand, as u is non-decreasing, the L^2 term $\|u - g^h\|^2$ is bounded below by $\|u^a - g^h\|^2$ where $u_n^a = a$ for $n < 0$, $u_0^a = \frac{1}{2}$ and $u_n^a = 1 - a$ for $n > 0$. On the other hand, we write that $\sqrt{(u_n - u_{n+1})^2 + (u_n - u_{n-1})^2} \geq \sqrt{2}|u_{n+1} - u_{n-1}| = \sqrt{2}(u_{n+1} - u_{n-1})$. We obtain:

$$E^h(u) \geq \frac{h}{2\sqrt{2}\lambda} \times 2Na^2 + \frac{1}{2}(u_{N+1} + u_N - u_{-N-1} - u_{-N}) = \frac{\sqrt{2}}{2\lambda}a^2 + 1 - 2a$$

As $\lambda \geq \lambda^* = \frac{\sqrt{2}}{4}$, minimizing this quantity over $a \in [0, \frac{1}{2}]$ leads to taking $a = \frac{1}{2}$, and we get $E^h(u) \geq \frac{\sqrt{2}}{8\lambda} = E^h(\frac{1}{2})$, hence $\bar{u}^h \equiv \frac{1}{2}$. \square

3 Upper bound for the primal energy

In this section we prove the upper bound of the point 2 of theorem 1, that is: $\exists \underline{h}, c > 0$ such that

$$\forall h \leq \underline{h}, \bar{E}^h - \bar{E} \leq ch^{2/3}$$

We first focus on the Dirichlet case and later on will present the modifications needed for Neumann boundary conditions. As no reference to the continuous problem will appear in this section (except from its value \bar{E}) we drop the exponent h and denote the variables u^h, g^h simply by $u, g \in \mathbb{R}^{2N+1}$. Recall that the primal problem in the Dirichlet

setting is:

$$\bar{u} = \arg \min_{\substack{(u_n)_{-2N \leq n \leq 2N} \\ u_{2N+1} = u_{2N} = 1 \\ u_{-2N-1} = u_{-2N} = 0}} \frac{h}{2\sqrt{2}\lambda} \|u - g\|_2^2 + \text{tv}_i^h(u) := E^h(u)$$

$$\text{with } \text{tv}_i^h(u) = \frac{1}{\sqrt{2}} \sum_{n=-2N}^{2N} \sqrt{(u_{n+1} - u_n)^2 + (u_n - u_{n-1})^2}$$

and where $g_n = 0$ for $n < 0$, $g_n = 1$ for $n > 0$ and $g_0 = 1/2$. The limit continuous energy is $\bar{E} = \bar{E}_D = 1$. In the following we build an admissible u of a particular form to establish an upper bound estimate of the type

$$\bar{E}^h \leq E^h(u) \leq \bar{E} + ch^\theta,$$

for some $0 < \theta < 1$.

3.1 General construction

The idea is to take a function u such that $u - g$ has a compact support of vanishing size but containing a number of points going to infinity. This is achieved by taking u_n , for $-N \leq n \leq N$, of the form (remember that $N = \frac{1}{h}$):

$$u_n = f\left(\frac{n}{N_\alpha}\right) \text{ with } N_\alpha = \lceil h^{-\alpha} \rceil \text{ and } 0 < \alpha < 1$$

where f is some continuous function increasing from $f(x) = 0$ for $x \leq -1$ to $f(x) = 1$ for $x \geq 1$. We also suppose in all what follows that f satisfies $f(-x) = 1 - f(x)$ for any $x \in \mathbb{R}$ to fulfill the conclusions of Proposition 1.

As $u = g$ is constant for $|n| \geq N_\alpha$, one only has to consider what is happening in the transition phase, that is for $|n| < N_\alpha$ for the L^2 terms, and for $|n| \leq N_\alpha$ for the tv terms. To understand what is at stake, let us first try with the piecewise affine function

$$f(x) = \begin{cases} 0 & \text{if } x < -1 \\ \frac{x+1}{2} & \text{if } -1 \leq x \leq 1 \\ 1 & \text{if } x > 1 \end{cases}$$

First compute the fidelity term:

$$\begin{aligned} \frac{h}{2} \|u - g\|_2^2 &= h \sum_1^{N_\alpha-1} (f(\frac{n}{N_\alpha}) - 1)^2 \\ &= \frac{h}{4N_\alpha^2} \sum_1^{N_\alpha-1} n^2 \\ &= \frac{hN_\alpha}{12} - \frac{h}{8} + \frac{h}{24N_\alpha} \end{aligned}$$

and then the tv term:

$$\begin{aligned}
\text{tv}_i^h(u) &= \frac{1}{\sqrt{2}} \sum_{-N_\alpha+1}^{N_\alpha-1} \sqrt{\left(\frac{n+1}{2N_\alpha} - \frac{n}{2N_\alpha}\right)^2 + \left(\frac{n}{2N_\alpha} - \frac{n-1}{2N_\alpha}\right)^2} \\
&\quad + \frac{1}{\sqrt{2}} \left| 1 - \frac{1}{2} \left(\frac{N_\alpha-1}{N_\alpha} + 1 \right) \right| + \frac{1}{\sqrt{2}} \left| \frac{1}{2} \left(\frac{-N_\alpha+1}{N_\alpha} + 1 \right) \right| \\
&= \frac{1}{\sqrt{2}} \left((2N_\alpha - 1) \times \sqrt{2 \times \frac{1}{4N_\alpha^2} + \frac{1}{N_\alpha}} \right) \\
&= 1 + \frac{\sqrt{2} - 1}{2N_\alpha}
\end{aligned}$$

Note that, the limit energy appears in the above expression as $1 = \bar{E}$. This finally leads to

$$\begin{aligned}
E^h(u) - \bar{E} &= \frac{\sqrt{2} - 1}{2N_\alpha} + \frac{hN_\alpha}{12\sqrt{2}\lambda} - \frac{h}{8\sqrt{2}\lambda} + \frac{h}{24\sqrt{2}\lambda N_\alpha} \\
&\leq \frac{\sqrt{2} - 1}{2} h^\alpha + \frac{h(h^{-\alpha} + 1)}{12\sqrt{2}\lambda} - \frac{h}{8\sqrt{2}\lambda} + \frac{h^{\alpha+1}}{24\sqrt{2}\lambda} \\
&\leq \frac{\sqrt{2} - 1}{2} h^\alpha + \frac{h^{1-\alpha}}{12\sqrt{2}\lambda} + \frac{h^{\alpha+1}}{24\sqrt{2}\lambda}
\end{aligned}$$

The optimal choice of α is then to make the two dominant terms in h^α and $h^{1-\alpha}$ of the same order, hence $\alpha = 1/2$. We conclude that, for any $c > \frac{\sqrt{2}-1}{2} + \frac{1}{12\sqrt{2}\lambda}$, one has for h small enough

$$E^h(u) - \bar{E} \leq c\sqrt{h}$$

In the following we show that with a cubic function f , realising a smoother transition, this procedure leads to the better result: there exist constants $c > 0$ and $\underline{h} > 0$ depending only on λ such that:

$$\forall h \leq \underline{h}, \quad E^h(u) - \bar{E} \leq ch^{2/3} \quad (9)$$

3.2 Analysis for an appropriate function f

In fact for any function regular enough (C^1) f , when $h \rightarrow 0$ we have: u^h converges to g in L^2 so $h\|u^h - g^h\|_2^2 \rightarrow 0$, and $\text{tv}_i^h(u) \rightarrow \text{tv}(g) = 1$. So $E(u) \rightarrow \bar{E}$. We want to estimate the speed of this convergence.

The L^2 term is easy to estimate:

$$\begin{aligned} \frac{h}{2} \|u - g\|_2^2 &= h \sum_{n=1}^{N_\alpha-1} \left(f\left(\frac{n}{N_\alpha}\right) - 1\right)^2 \\ &= h N_\alpha \frac{1}{N_\alpha} \sum_1^{N_\alpha-1} \left(f\left(\frac{n}{N_\alpha}\right) - 1\right)^2 \\ &\sim h^{1-\alpha} \int_0^1 (f - 1)^2 \text{ when } N_\alpha \rightarrow \infty \end{aligned}$$

hence for any $c_1 > \frac{1}{\sqrt{2\lambda}} \int_0^1 (f - 1)^2$, we have for h small enough:

$$\frac{h}{2\sqrt{2\lambda}} \|u - g\|_2^2 \leq c_1 h^{1-\alpha} \quad (10)$$

Manipulations on the total variation term are trickier, it is given by:

$$\begin{aligned} \text{tv}_i^h(u) &= \frac{1}{\sqrt{2}} \sum_{n=-N_\alpha+1}^{N_\alpha-1} \sqrt{\left(f\left(\frac{n+1}{N_\alpha}\right) - f\left(\frac{n}{N_\alpha}\right)\right)^2 + \left(f\left(\frac{n}{N_\alpha}\right) - f\left(\frac{n-1}{N_\alpha}\right)\right)^2} \\ &\quad + \frac{1}{\sqrt{2}} \left|1 - f\left(\frac{N_\alpha-1}{N_\alpha}\right)\right| + \frac{1}{\sqrt{2}} \left|f\left(\frac{-N_\alpha+1}{N_\alpha}\right)\right| \end{aligned}$$

The boundary terms simplify into

$$\frac{1}{\sqrt{2}} \left|1 - f\left(\frac{N_\alpha-1}{N_\alpha}\right)\right| + \frac{1}{\sqrt{2}} \left|f\left(\frac{-N_\alpha+1}{N_\alpha}\right)\right| = \sqrt{2} \left(1 - f\left(1 - \frac{1}{N_\alpha}\right)\right)$$

For the middle terms, we use the following lemma with $u_n = f\left(\frac{n}{N_\alpha}\right)$:

Lemma 1. *If (u_n) is an increasing sequence, then for any n :*

$$\frac{1}{\sqrt{2}} \sqrt{(u_{n+1} - u_n)^2 + (u_n - u_{n-1})^2} \leq \frac{1}{2} (u_{n+1} - u_{n-1}) + d_n$$

with

$$d_n = \frac{1}{4} (u_{n+1} - u_{n-1}) (2u_n - u_{n+1} - u_{n-1}) \left(\frac{1}{u_{n+1} - u_n} - \frac{1}{u_n - u_{n-1}} \right) \quad (11)$$

$$= \frac{(u_{n+1} - u_{n-1}) (2u_n - u_{n+1} - u_{n-1})^2}{4(u_{n+1} - u_n)(u_n - u_{n-1})} \quad (12)$$

Proof. Denote $A = \sqrt{(u_{n+1} - u_n)^2 + (u_n - u_{n-1})^2}$ the quantity we want to estimate. Using $\sqrt{x+h} \leq \sqrt{x} + \frac{1}{2\sqrt{x}}h$ we get:

$$\begin{aligned} A &= \sqrt{2(u_{n+1} - u_n)^2 + (u_n - u_{n-1})^2 - (u_{n+1} - u_n)^2} \\ &= \sqrt{2(u_{n+1} - u_n)^2 + (u_{n+1} - u_{n-1})(2u_n - u_{n+1} - u_{n-1})} \\ &\leq \sqrt{2}(u_{n+1} - u_n) + \frac{1}{2\sqrt{2}(u_{n+1} - u_n)} (u_{n+1} - u_{n-1})(2u_n - u_{n+1} - u_{n-1}) \end{aligned}$$

And similarly

$$\begin{aligned} A &= \sqrt{2(u_n - u_{n-1})^2 + (u_{n+1} - u_n)^2 - (u_n - u_{n-1})^2} \\ &\leq \sqrt{2}(u_n - u_{n-1}) - \frac{1}{2\sqrt{2}(u_n - u_{n-1})}(u_{n+1} - u_{n-1})(2u_n - u_{n+1} - u_{n-1}) \end{aligned}$$

The result is obtained as the average of these two estimates. \square

The term in $\frac{1}{2}(u_{n+1} - u_{n-1}) = \frac{1}{2}(f(\frac{n+1}{N_\alpha}) - f(\frac{n-1}{N_\alpha}))$ is responsible for the convergence towards 1 as

$$\begin{aligned} \sum_{-N_\alpha+1}^{N_\alpha-1} \frac{1}{2} \left(f\left(\frac{n+1}{N_\alpha}\right) - f\left(\frac{n-1}{N_\alpha}\right) \right) &= \frac{1}{2} \left(f(1) + f\left(1 - \frac{1}{N_\alpha}\right) - f(-1) - f\left(-1 + \frac{1}{N_\alpha}\right) \right) \\ &= f\left(1 - \frac{1}{N_\alpha}\right) \end{aligned}$$

For the term in d_n note that the symmetry of f gives $d_0 = 0$ and $d_{-n} = d_n$ so that the sum is reduced to $n \in [1, N_\alpha - 1]$ and we get the expression:

$$\text{tv}_i^h(u) \leq 1 + (\sqrt{2} - 1)\left(1 - f\left(1 - \frac{1}{N_\alpha}\right)\right) + \sum_1^{N_\alpha-1} d_n \quad (13)$$

Next we pursue our analysis for a particular function f given by

$$f(t) = \begin{cases} 0 & \text{if } t \leq -1 \\ \frac{1}{2}(1+t)^3 & \text{if } -1 \leq t \leq 0 \\ 1 - \frac{1}{2}(1-t)^3 & \text{if } 0 \leq t \leq 1 \\ 1 & \text{if } t \geq 1 \end{cases} \quad (14)$$

Then the term $1 - f\left(1 - \frac{1}{N_\alpha}\right)$ equals $\frac{1}{N_\alpha^3}$ while $\sum_1^{N_\alpha-1} d_n$ is bounded by $\frac{1}{N_\alpha^2}$:

Claim. For the choice of f given by (14), one has

$$\sum_1^{N_\alpha-1} d_n \leq \frac{6}{N_\alpha^2}$$

Proof. Let us denote

$$\begin{aligned} \Delta_+ &:= f\left(\frac{n+1}{N_\alpha}\right) - f\left(\frac{n}{N_\alpha}\right) = \frac{1}{2} \left(\left(1 - \frac{n}{N_\alpha}\right)^3 - \left(1 - \frac{n+1}{N_\alpha}\right)^3 \right) \\ &= \frac{1}{2} \left(3\left(1 - \frac{n}{N_\alpha}\right)^2 \frac{1}{N_\alpha} - 3\left(1 - \frac{n}{N_\alpha}\right) \frac{1}{N_\alpha^2} + \frac{1}{N_\alpha^3} \right) \\ &= \frac{3}{2N_\alpha} \left(\left(1 - \frac{n}{N_\alpha}\right)^2 - \left(1 - \frac{n}{N_\alpha}\right) \frac{1}{N_\alpha} + \frac{1}{3N_\alpha^2} \right) \end{aligned}$$

Similarly (that is, taking $n \leftarrow n - 1$),

$$\Delta_- := f\left(\frac{n}{N_\alpha}\right) - f\left(\frac{n-1}{N_\alpha}\right) = \frac{3}{2N_\alpha} \left(\left(1 - \frac{n}{N_\alpha}\right)^2 + \left(1 - \frac{n}{N_\alpha}\right) \frac{1}{N_\alpha} + \frac{1}{3N_\alpha^2} \right)$$

so that

$$\begin{aligned}\Delta_+ + \Delta_- &= f\left(\frac{n+1}{N_\alpha}\right) - f\left(\frac{n-1}{N_\alpha}\right) = \frac{3}{N_\alpha} \left(\frac{1}{3N_\alpha^2} + \left(1 - \frac{n}{N_\alpha}\right)^2 \right) \\ \Delta_- - \Delta_+ &= 2f\left(\frac{n}{N_\alpha}\right) - f\left(\frac{n+1}{N_\alpha}\right) - f\left(\frac{n-1}{N_\alpha}\right) = \frac{3}{N_\alpha^2} \left(1 - \frac{n}{N_\alpha}\right)\end{aligned}$$

and

$$\begin{aligned}\Delta_+ \times \Delta_- &= \left(f\left(\frac{n+1}{N_\alpha}\right) - f\left(\frac{n}{N_\alpha}\right)\right) \left(f\left(\frac{n}{N_\alpha}\right) - f\left(\frac{n-1}{N_\alpha}\right)\right) \\ &= \frac{9}{4N_\alpha^2} \left(\left(1 - \frac{n}{N_\alpha}\right)^2 + \frac{1}{3N_\alpha^2} - \left(1 - \frac{n}{N_\alpha}\right) \frac{1}{N_\alpha} \right) \\ &\quad \times \left(\left(1 - \frac{n}{N_\alpha}\right)^2 + \frac{1}{3N_\alpha^2} + \left(1 - \frac{n}{N_\alpha}\right) \frac{1}{N_\alpha} \right) \\ &= \frac{9}{4N_\alpha^2} \left(\left(\left(1 - \frac{n}{N_\alpha}\right)^2 + \frac{1}{3N_\alpha^2} \right)^2 - \left(\left(1 - \frac{n}{N_\alpha}\right) \frac{1}{N_\alpha} \right)^2 \right)\end{aligned}$$

We can now estimate d_n thanks to expression (12):

$$\begin{aligned}d_n &= \frac{1}{4} \times \frac{\left(\frac{3}{N_\alpha} \left(\frac{1}{3N_\alpha^2} + \left(1 - \frac{n}{N_\alpha}\right)^2 \right) \right) \times \left(\frac{3}{N_\alpha^2} \left(1 - \frac{n}{N_\alpha}\right) \right)^2}{\frac{9}{4N_\alpha^2} \left(\left(\left(1 - \frac{n}{N_\alpha}\right)^2 + \frac{1}{3N_\alpha^2} \right)^2 - \left(\left(1 - \frac{n}{N_\alpha}\right) \frac{1}{N_\alpha} \right)^2 \right)} \\ &= \frac{3}{N_\alpha^3} \left(1 - \frac{n}{N_\alpha}\right)^2 \times \frac{\frac{1}{3N_\alpha^2} + \left(1 - \frac{n}{N_\alpha}\right)^2}{\left(\left(1 - \frac{n}{N_\alpha}\right)^2 + \frac{1}{3N_\alpha^2} \right)^2 - \left(1 - \frac{n}{N_\alpha}\right)^2 \frac{1}{N_\alpha^2}}\end{aligned}$$

Then as $n \leq N_\alpha - 1$ we can use

$$\left(\left(1 - \frac{n}{N_\alpha}\right)^2 + \frac{1}{3N_\alpha^2} \right)^2 - \left(1 - \frac{n}{N_\alpha}\right)^2 \frac{1}{N_\alpha^2} \geq \left(1 - \frac{n}{N_\alpha}\right)^4 - \frac{1}{3N_\alpha^2} \left(1 - \frac{n}{N_\alpha}\right)^2 > 0$$

and make the variable change $n \leftarrow N - n$ to get:

$$\sum_{n=1}^{N_\alpha-1} d_n \leq \frac{3}{N_\alpha^3} \sum_{n=1}^{N_\alpha-1} n^2 \frac{\frac{1}{3N_\alpha^2} + n^2}{n^4 - \frac{1}{3N_\alpha^2} n^2} \leq \frac{3}{N_\alpha^3} \sum_{n=1}^{N_\alpha-1} 2 \leq \frac{6}{N_\alpha^2}$$

because $\frac{1}{3N_\alpha^2} + n^2 \leq 2(n^2 - \frac{1}{3N_\alpha^2})$. □

This concludes our proof of the upper bound inequality for the main theorem 1. Indeed, when combining the tv estimate (13) with the L^2 estimate (10), we finally are able to state the following: for any $c_1 > \frac{1}{\sqrt{2\lambda}} \int_0^1 (f-1)^2 = \frac{1}{28\sqrt{2\lambda}}$ and $c_2 > 6$, there exists $\underline{h} > 0$ such that

$$\forall h \leq \underline{h}, \quad E^h(u) \leq \bar{E} + c_1 h^{1-\alpha} + c_2 h^{2\alpha} \quad (15)$$

Taking $\alpha = 1/3$ then proves our result (9). More precisely, given c_1, c_2 and $h > 0$, the best α in (15) must satisfy $-c_1 h^{1-\alpha} \log h + 2c_2 h^{2\alpha} \log h = 0$ which leads to $\alpha = \frac{1}{3} - \frac{\log(2c_2/c_1)}{3 \log h}$ and gives the upper bound $E^h(u) \leq \bar{E} + ch^{2/3}$ with $c = (2^{1/3} + 2^{-2/3}) c_1^{2/3} c_2^{1/3}$ (note that c varies in $\lambda^{-2/3}$).

3.3 Upper bound for Neumann boundary conditions

In this section we adjust the admissible variable u from the previous section to explain why the upper bound result (9) remains valid for Neumann boundary conditions. In the following, c denotes a constant depending only on λ that can change from line to line.

Remember from Section 2.2 that with the Neumann boundary conditions, the limit continuous value of the energy is changed to $\bar{E} = \bar{E}_N = 1 - \sqrt{2}\lambda$ when $\lambda \leq \lambda^* = \frac{\sqrt{2}}{4}$. Because of the form of this continuous solution, it is natural to consider, for u the cubic transition in the Dirichlet setting of the previous section, the variable v given by

$$\forall -N \leq n \leq N, v_n = \frac{1}{2} + \mu(u_n - \frac{1}{2})$$

Here $\mu \in (0, 1)$ is a shrinking parameter that we adjust so that $v_N = 1 - a_{opt} = 1 - \sqrt{2}\lambda$: as $u_N = 1$ this corresponds to taking $\mu = 1 - 2\sqrt{2}\lambda$.

We write $v_n = f_\mu(\frac{n}{N_\alpha})$ for the function $f_\mu = \frac{1}{2} + \mu(f - \frac{1}{2})$ which is such that $f_\mu(x) = \frac{1+\mu}{2} = 1 - \sqrt{2}\lambda$ for $x \geq 1$. This leads to splitting the fidelity term into:

$$\frac{h}{2} \|v - g\|_2^2 = h \sum_{n=1}^{N_\alpha} (v_n - 1)^2 + h \sum_{n=N_\alpha+1}^N (v_n - 1)^2$$

Then on one hand when $N_\alpha \rightarrow \infty$,

$$h \sum_{n=1}^{N_\alpha} (v_n - 1)^2 \sim h^{1-\alpha} \int_0^1 (f_\mu - 1)^2 \quad \text{so} \quad h \sum_{n=1}^{N_\alpha} (v_n - 1)^2 \leq ch^{1-\alpha}$$

and on the other hand

$$h \sum_{n=N_\alpha+1}^N (v_n - 1)^2 = h(N - N_\alpha) \times 2\lambda^2 \leq 2\lambda^2$$

For the tv term, we have

$$\begin{aligned} \text{tv}_i^h(v) &= \mu \text{tv}_i^h(u) = (1 - 2\sqrt{2}\lambda) \text{tv}_i^h(u) \\ &\leq (1 - 2\sqrt{2}\lambda)(1 + ch^{2\alpha}) \\ &\leq 1 - 2\sqrt{2}\lambda + ch^{2\alpha} \end{aligned}$$

so finally

$$\begin{aligned} E^h(v) &= \frac{h}{2\sqrt{2}\lambda} \|v - g\|_2^2 + \text{tv}_i^h(v) \\ &\leq \sqrt{2}\lambda + ch^{1-\alpha} + 1 - 2\sqrt{2}\lambda + ch^{2\alpha} \\ &\leq \bar{E} + ch^{2/3} \end{aligned}$$

when taking $\alpha = 1/3$.

4 Lower bound estimate

In this section we now prove the lower bound of the point 2 of theorem 1, that is: $\exists \underline{h}, c > 0$ such that

$$\forall h \leq \underline{h}, ch^{2/3} \leq \overline{E}^h - \overline{E}$$

Symmetrically to what we did in the previous section, we will obtain this bound by proposing an admissible solution, but for the dual problem.

4.1 Dual problem

Writing that

$$\sqrt{(u_{n+1} - u_n)^2 + (u_n - u_{n-1})^2} = \max_{\substack{p_n^2 + q_n^2 \leq 1 \\ -N \leq n \leq N}} q_n(u_{n+1} - u_n) + p_n(u_n - u_{n-1})$$

we obtain the following dual problem of (7):

$$\begin{aligned} & \max_{\substack{p_n^2 + q_n^2 \leq 1 \\ -N \leq n \leq N}} \min_{u \in \mathbb{R}^{2N+1}} \sum_{n=-N}^N \frac{h}{2\lambda} (u_n - g_n)^2 + q_n(u_{n+1} - u_n) + p_n(u_n - u_{n-1}) \\ &= \max_{\substack{p_n^2 + q_n^2 \leq 1 \\ -N \leq n \leq N}} \min_{u \in \mathbb{R}^{2N+1}} \frac{1}{\sqrt{2}} \left\{ \sum_{n=-N}^N \frac{h}{2\lambda} (u_n - g_n)^2 + \sum_{n=-N+1}^{N-1} u_n (q_{n-1} - q_n + p_n - p_{n+1}) \right. \\ & \quad \left. + u_N (q_{N-1} - q_N + p_N) + u_{-N} (-q_{-N} + p_{-N} - p_{-N+1}) \right. \\ & \quad \left. + u_{N+1} q_N - u_{-N-1} p_{-N} \right\} \end{aligned}$$

From this point on, we focus exclusively on Dirichlet boundary conditions, that is $u_N = u_{N+1} = 1$; $u_{-N} = u_{-N-1} = 0$. See Section 4.5 for Neumann boundary conditions.

For $|n| < N$, we find that $u_n = g_n - \frac{\lambda}{h}(q_{n-1} - q_n + p_n - p_{n+1})$, and the value of the dual problem is consequently (after simplification using the value of g_n):

$$\max_{\substack{p_n^2 + q_n^2 \leq 1 \\ -N \leq n \leq N}} \frac{1}{\sqrt{2}} \left\{ \frac{1}{2}(q_{-1} + q_0 + p_0 + p_1) - \frac{\lambda}{2h} \sum_{n=-N+1}^{N-1} (q_{n-1} - q_n + p_n - p_{n+1})^2 \right\}$$

Now we make two more simplifications before turning to an evaluation of the convergence rate of this quantity. First, one easily checks that the objective is concave and invariant by the change $(q_n, p_n) \rightarrow (p_{-n}, q_{-n})$: as a consequence, one can find a solution satisfying $q_n = p_{-n}$ for all n .

Second, duality indicates that at the optimum one should have for all n : $\sqrt{(u_{n+1} - u_n)^2 + (u_n - u_{n-1})^2} = q_n(u_{n+1} - u_n) + p_n(u_n - u_{n-1})$. For $n = 0$ this gives, thanks to Proposition 2 that $\sqrt{2}|u_1 - u_0| = (q_0 + p_0)(u_1 - u_0)$ so that $q_0 = p_0 = \frac{\sqrt{2}}{2}$. Simplifying the term $(q_{n-1} - q_n + p_n - p_{n+1})^2$ which is invariant by $n \rightarrow -n$ and vanishes at $n = 0$, we finally get

$$\overline{E}^h = \max \frac{1}{2} + \frac{1}{\sqrt{2}} p_1 - \frac{\lambda}{\sqrt{2}h} \sum_{n=1}^{N-1} (p_{-n+1} - p_{-n} + p_n - p_{n+1})^2 \quad (16)$$

with the constraint that $p_n^2 + p_{-n}^2 \leq 1$ for all $1 \leq n \leq N$ and that $p_0 = \frac{\sqrt{2}}{2}$.

4.2 Change of variables

We are interested in the evaluation of the convergence rate of the value of the problem (16) towards its continuous limit $\bar{E} = \bar{E}_D = 1$. First let us notice that taking $p_n \equiv \sqrt{2}/2$ gives $\bar{E}^h \geq \bar{E}$. Consequently, we expect the optimal value of p to be close to $\sqrt{2}/2$ for N large. Together with the symmetry regarding $n \rightarrow -n$ of the objective, this leads us to proposing the following change of variable: for $0 \leq n \leq N$

$$s_n = \frac{1}{\sqrt{2}}(p_n + p_{-n}) - 1 ; r_n = \frac{1}{\sqrt{2}}(p_n - p_{-n})$$

for which we calculate

$$p_{-n+1} - p_{-n} + p_n - p_{n+1} = \frac{1}{\sqrt{2}}(s_{n-1} - s_{n+1} + 2r_n - r_{n-1} - r_{n+1})$$

$$p_n^2 + p_{-n}^2 \leq 1 \iff s_n^2 + 2s_n + r_n^2 \leq 0$$

and it gives rise to

$$\bar{E}^h - \bar{E} = \max_{\substack{(s_n, r_n)_{0 \leq n \leq N} \\ s_0 = r_0 = 0 \\ s_n^2 + 2s_n + r_n^2 \leq 0}} \frac{1}{\sqrt{2}} \left\{ s_1 + r_1 - \frac{\lambda}{2h} \sum_{n=1}^{N-1} (s_{n-1} - s_{n+1} + 2r_n - r_{n-1} - r_{n+1})^2 \right\}$$

We would like to show that $\bar{E}^h - \bar{E} \geq cN^{-\alpha}$ for some exponent $0 < \alpha < 1$. If we introduce $\tau = 1/N^\beta$ for some $\beta \in (0, \alpha)$ and $\sigma_n = N^\alpha s_n$, $\rho_n = N^{\alpha-\beta} r_n$, then we can force the appearance of first and second discrete derivatives for σ and ρ as

$$(\bar{E}^h - \bar{E})N^\alpha = \max_{\substack{(\sigma_n, \rho_n) \\ 0 \leq n \leq N}} \frac{1}{\sqrt{2}} \left\{ \sigma_1 + \frac{\rho_1}{\tau} - \frac{\lambda}{2} N^{1-\alpha-\beta} \tau \sum_{n=1}^{N-1} \left(\frac{\sigma_{n-1} - \sigma_{n+1}}{\tau} + \frac{2\rho_n - \rho_{n-1} - \rho_{n+1}}{\tau^2} \right)^2 \right\}$$

along with the constraints $\sigma_0 = \rho_0 = 0$ and $N^{-\alpha} \sigma_n^2 + 2\sigma_n + N^{2\beta-\alpha} \rho_n^2 \leq 0$.

If $1 - \alpha - \beta = 0$, we find that as $N \rightarrow \infty$, the limiting energy in the variational problem should be of the form,

$$\max \frac{1}{\sqrt{2}} \left\{ \rho'(0) - \frac{\lambda}{2} \int_0^\infty |2\sigma' + \rho''|^2 \right\}$$

for functions $\sigma, \rho : [0, \infty) \rightarrow \mathbb{R}$ with $\sigma(0) = \rho(0) = 0$. The constraint, on the other hand, becomes

$$\begin{cases} \rho^2 = 0 & \text{if } 2\beta - \alpha > 0 \\ 2\sigma + \rho^2 \leq 0 & \text{if } 2\beta - \alpha = 0 \Leftrightarrow \beta = 1/3, \alpha = 2/3 \\ 2\sigma \leq 0 & \text{if } 2\beta - \alpha < 0. \end{cases}$$

In the first case, which is when $\alpha < 2/3$, we may expect that the discrete energy goes to zero, and we expect that $\overline{E}^h - \overline{E} = o(N^{-\alpha})$ as $N \rightarrow \infty$. In the third case, the continuous problem has value $+\infty$ and we expect that $N^\alpha(\overline{E}^h - \overline{E}) \rightarrow \infty$ for $\alpha > 2/3$. We would like to show that in the second case, that is $\alpha = 2/3$, the limiting problem has a positive value c so that $\overline{E}^h - \overline{E} \geq cN^{-2/3}$ for sufficiently large N . Consequently we deal with the problem

$$\max_{(\sigma, \rho) \in S} \frac{1}{\sqrt{2}} \left\{ \rho'(0) - \frac{\lambda}{2} \int_0^\infty (2\sigma' + \rho'')^2 \right\} =: D(\sigma, \rho) \quad (17)$$

where S is the set of couples of functions $\sigma, \rho : [0, \infty) \rightarrow \mathbb{R}$ such that: $\sigma(0) = \rho(0) = 0$, $2\sigma + \rho^2 \leq 0$, ρ admits a right derivative at 0 and the distributional derivative $2\sigma' + \rho''$ is in $L^2(0, \infty)$.

Our strategy is now the following: in Section 4.3 we prove that Problem (17) has a positive value and investigate the form of the solution (σ, ρ) . Then in Section 4.4 we explain how to discretize it in order to get the positivity, for h small enough, of our discrete problem:

$$(\overline{E}^h - \overline{E})h^{-2/3} = \max_{\substack{(\sigma_n, \rho_n)_{0 \leq n \leq N} \\ \sigma_0 = \rho_0 = 0 \\ N^{-2/3}\sigma_n^2 + 2\sigma_n + \rho_n^2 \leq 0}} D^h(\sigma, \rho) \quad (18)$$

$$\text{where } D^h(\sigma, \rho) := \frac{1}{\sqrt{2}} \left\{ \sigma_1 + \frac{\rho_1}{\tau} - \frac{\lambda}{2} \tau \sum_{n=1}^{N-1} \left(\frac{\sigma_{n-1} - \sigma_{n+1}}{\tau} + \frac{2\rho_n - \rho_{n-1} - \rho_{n+1}}{\tau^2} \right)^2 \right\}$$

4.3 Study of the limit problem

First, the change of variable $\hat{\sigma}(t) = \lambda^{-2/3}\sigma(t\lambda^{-1/3})$, $\hat{\rho}(t) = \lambda^{-1/3}\rho(t\lambda^{-1/3})$ shows that (adding the parameter λ to the arguments of D)

$$\max D(\sigma, \rho, \lambda) = \lambda^{-2/3} \max D(\sigma, \rho, 1)$$

Consequently we suppose $\lambda = 1$ in all of the following.

4.3.1 A dual of the dual

To understand the solution of problem (17), we derive a dual of it writing

$$-\frac{1}{2} \int_0^\infty (2\sigma' + \rho'')^2 = \inf_{\psi} \int_0^\infty (2\sigma' + \rho'')\psi + \frac{1}{2} \int_0^\infty \psi^2$$

where the infimum lies on $\psi \in \mathcal{C}_c^\infty([0, \infty))$, the set of the restrictions to $[0, \infty)$ of smooth functions with compact support in \mathbb{R} . (Note that if σ, ρ are regular enough one has at the optimum $\psi = -(2\sigma' + \rho'')$). Integrating by parts and using that $\sigma(0) = \rho(0) = 0$ for any $(\sigma, \rho) \in S$, we obtain the dual problem

$$\frac{1}{\sqrt{2}} \inf_{\psi} \frac{1}{2} \int_0^\infty \psi^2 + \sup_{(\sigma, \rho) \in S} (1 - \psi(0))\rho'(0) + \int_0^\infty (\rho\psi'' - 2\sigma\psi')$$

First, taking for ρ a bounded smooth function with $|\rho'(0)|$ as large as we want, we see that one must have $\psi(0) = 1$. Second, we relax the constraint $(\sigma, \rho) \in S$ in the remaining integral into just $2\sigma + \rho^2 \leq 0$ (we will show below that strong duality with problem (17) actually occurs) to get:

$$\frac{1}{\sqrt{2}} \inf_{\psi(0)=1} \frac{1}{2} \int_0^\infty |\psi|^2 + \int_0^\infty H(\psi', \psi'')$$

where the function H is defined for $x, y \in \mathbb{R}$ by

$$H(x, y) = \sup_{2\sigma + \rho^2 \leq 0} -2\sigma x + \rho y = \begin{cases} +\infty & \text{if } x > 0 \text{ or } x = 0, y \neq 0, \text{ (via } \rho = 0, \sigma = -N) \\ 0 & \text{if } (x, y) = (0, 0), \\ \frac{y^2}{4|x|} & \text{if } x < 0 \text{ (via } \rho = -y/2x, \sigma = -\rho^2/2). \end{cases}$$

Observe that necessarily $\psi' \leq 0$. Denoting $\phi = \sqrt{-\psi'}$ gives $\phi' = -\psi''/(2\sqrt{-\psi'})$ so that $H(\psi', \psi'') = |\phi'|^2$. Then, one has $\psi(x) = 1 - \int_0^x \phi(t)^2 dt$. In particular as ψ^2 is integrable, one must have $\int_0^\infty \phi(t)^2 dt = 1$ and $\psi(x) = \int_x^\infty \phi(t)^2 dt$. Hence the dual problem can be rewritten (extending the search of ϕ to $H^1(0, \infty)$ by density)

$$\frac{1}{\sqrt{2}} \inf_{(\phi, \psi) \in S'} \left\{ \frac{1}{2} \int_0^\infty |\psi|^2 + \int_0^\infty |\phi'|^2 \right\} \quad (19)$$

where $S' = \{(\phi, \psi) \text{ such that } \phi \in H^1(0, \infty), \|\phi\|_2^2 = 1 \text{ and } \psi(x) = \int_x^\infty \phi(t)^2 dt\}$.

It turns out this problem has a positive value:

Proposition 3. *Problem (19) has a minimizer $(\psi, \phi) \in W^{2,1}(0, \infty) \times H^1(0, \infty)$.*

Proof. Consider a minimizing sequence (ϕ_n, ψ_n) : as ϕ_n is bounded in $H^1(0, +\infty)$, up to a subsequence it converges to some ϕ , moreover the convergence is strong in $L^2(0, T)$ for any $T > 0$, and $\int_0^\infty \phi^2 \leq 1$. We also assume that ψ_n converges, weakly in $L^2(0, +\infty)$, to some ψ . In addition, $\psi_n(x) = 1 - \int_0^x \phi_n^2 \rightarrow 1 - \int_0^x \phi^2 =: \tilde{\psi}(x)$ for any $x \geq 0$, and one even has $|\psi_n(x) - \tilde{\psi}(x)| = |\int_0^x (\phi_n - \phi)(\phi_n + \phi)| \leq 2\|\phi_n - \phi\|_{L^2(0, x)}$ hence the convergence is locally uniform. Consequently, it must be that $\tilde{\psi} = \psi$. As $\int_0^\infty |\psi|^2 < +\infty$, we deduce that ψ (which is nonincreasing) goes to 0 at ∞ , hence $\int_0^\infty \phi^2 = 1$. It follows that (ψ, ϕ) is a minimizer of (19). \square

To recover the positive value of problem (17), we now need to show that strong duality holds. To do that we first prove some properties of the minimizer (ψ, ϕ) .

Proposition 4. *The minimizer (ψ, ϕ) of problem (19) satisfies:*

1. $\psi, \phi \in C^\infty([0, \infty)) \cap L^2(0, \infty)$.
2. $\phi'(0) = 0$ and $\phi'' = k\phi$ where $k(t) = \int_0^t \psi - A$ with $A = \|\phi'\|_2^2 + \|\psi\|_2^2$ satisfies $k' = \psi$.
3. $\phi \geq 0$, $\phi(0) > 0$, ϕ is nonincreasing and tends to zero at infinity.

Proof. One has $\psi' = -\phi^2 \in L^1(0, \infty)$ and $\psi'' = -2\phi\phi' \in L^1(0, \infty)$ (hence $\psi \in W^{2,1}(0, +\infty)$ and is at least C^1). Moreover, if (ψ, ϕ) is a minimizer, so is $(\psi, |\phi|)$. The solution of (19) being unique, one has $\phi \geq 0$.

From this solution (ψ, ϕ) , let us form for $\varepsilon \in \mathbb{R}$ and for a test function η

$$\phi_\varepsilon = \frac{\phi + \varepsilon\eta}{\|\phi + \varepsilon\eta\|_2} ; \quad \psi_\varepsilon(x) = \int_x^\infty \phi_\varepsilon^2$$

Then $(\phi_\varepsilon, \psi_\varepsilon)$ are admissible in the dual of the dual problem and one computes:

$$\begin{aligned} \phi_\varepsilon^2 &= \phi^2 + 2\varepsilon\eta\phi - 2\varepsilon\phi^2 \int_0^\infty \phi\eta + O(\varepsilon^2) \\ \psi_\varepsilon^2(x) &= \psi^2(x) + 4\varepsilon\psi(x) \int_x^\infty \phi\eta - 4\varepsilon\psi^2(x) \int_0^\infty \phi\eta + O(\varepsilon^2) \\ \phi'_\varepsilon &= \phi' + \varepsilon\eta' - \varepsilon\phi' \int_0^\infty \phi\eta + O(\varepsilon^2) \end{aligned}$$

so that, after noting that $\int_0^\infty \psi(x) \int_x^\infty \phi\eta \, dx = \int_0^\infty \phi\eta\nu$ with $\nu(t) = \int_0^t \psi$ one has

$$\begin{aligned} \int_0^\infty |\phi'_\varepsilon|^2 &= \int_0^\infty |\phi'|^2 - 2\varepsilon \int_0^\infty |\phi'|^2 \int_0^\infty \phi\eta + 2\varepsilon \int_0^\infty \phi'\eta' + O(\varepsilon^2) \\ \int_0^\infty |\psi_\varepsilon|^2 &= \int_0^\infty |\psi|^2 - 4\varepsilon \int_0^\infty |\psi|^2 \int_0^\infty \phi\eta + 4\varepsilon \int_0^\infty \phi\eta\nu + O(\varepsilon^2) \end{aligned}$$

Now the optimality of (ψ, ϕ) in problem (19) leads to

$$\int_0^\infty \phi\eta\nu - \int_0^\infty |\psi|^2 \int_0^\infty \phi\eta + \int_0^\infty \phi'\eta' - \int_0^\infty |\phi'|^2 \int_0^\infty \phi\eta = 0 \quad (20)$$

First, as this relation holds for any $\eta \in \mathcal{C}_c^\infty(0, \infty)$, we have $\phi'' = k\phi$ (with $k = \nu - A$ where $A = \|\psi\|^2 + \|\phi'\|^2$) in the weak sense. However this relation induces the regularity of ϕ and ψ which are finally C^∞ . What is more is that, re-evaluating the relation (20) with now $\eta \in \mathcal{C}_c^\infty([0, \infty))$, we also deduce that $\phi'(0) = 0$.

To finish with, one must have $\phi(0) > 0$ as otherwise ϕ would be zero everywhere as solution of $\phi'' = k\phi$, $\phi'(0) = \phi(0) = 0$. And for its monotonicity, note that $\phi'' = k\phi$ has the sign of k which is nonincreasing since $k' = \psi \geq 0$. Hence ϕ'' is first nonpositive (starting at $\phi''(0) = -A\phi(0) \leq 0$) then possibly nonnegative. As a consequence, ϕ' is first nonincreasing, and hence nonpositive since $\phi'(0) = 0$, then can become nondecreasing. But even in that case, ϕ' has to remain nonpositive otherwise one has $\phi'(t) \geq c > 0$ for t large enough so $\phi(t) \geq ct + c'$ which contradicts the fact that ϕ^2 is integrable. This concludes the proof. \square

In the following we show that strong duality holds between problems (17) and (19). To do so we divide our study in two cases: either $\phi > 0$ on \mathbb{R}^+ (the ‘‘positive’’ case), or $\phi > 0$ on $[0, a)$ and $\phi = 0$ on $[a, +\infty[$ for some $a > 0$ (the ‘‘compact support’’ case). Note that numerical experiments seem to show we actually are in the ‘‘compact support’’ case, see Figure 4.

4.3.2 Strong duality holds

In the “positive” case, recalling how the dual problem was obtained, one defines $\sigma = -\rho^2/2$ and $\rho = -\phi'/\phi$ and then checks that $2\sigma + \rho^2 \leq 0$, $\sigma(0) = \rho(0) = 0$, $\rho'(0) = A$ and $2\sigma' + \rho'' = -\psi$ so that

$$\frac{1}{\sqrt{2}} \left\{ \rho'(0) - \frac{1}{2} \int_0^\infty |2\sigma' + \rho''|^2 \right\} = \frac{1}{\sqrt{2}} \left\{ \int_0^\infty \phi'^2 + \frac{1}{2} \int_0^\infty \psi^2 \right\}$$

and strong duality holds.

In the “compact support” case, one still defines $\rho = -\phi'/\phi$ and $\sigma = -\rho^2/2$ on $[0, a)$. Then one has to decide what to do on $[a, +\infty)$. First, for $t < a$:

$$\rho(t) = -\frac{\phi'(t)}{\phi(t)} = \frac{1}{\phi(t)} \int_t^a \phi''(s) ds = \int_t^a \frac{\phi(s)}{\phi(t)} k(s) ds$$

As ϕ is nonincreasing, $\frac{\phi(s)}{\phi(t)} \leq 1$ in the above integral and we deduce

$$|\rho(t)| \leq \int_t^a k(s) ds \rightarrow 0 \quad \text{when } t \rightarrow a$$

and also $\sigma(t) = -\rho(t)^2/2 \rightarrow 0$ when $t \rightarrow a$. The first guess would then consist in extending σ and ρ by continuity one could set $\sigma = \rho = 0$ on $[a, +\infty)$.

This would actually lead to a discontinuous ρ' . Indeed ρ is differentiable in a^+ with $\rho'(a^+) = 0$; furthermore $\rho'(t) = \rho^2(t) + k(t)$ for $t \in (0, a)$ with $\rho(t) \rightarrow 0$ and $k(t) \rightarrow k(a)$ when $t \rightarrow a$, hence ρ is differentiable in a^- with $\rho'(a^-) = k(a)$. Anyway ρ' is discontinuous at a (and \mathcal{C}^∞ elsewhere), so ρ'' has a dirac mass at a . Whereas $\sigma = -\rho^2/2$ on $(0, a)$ as well as on $[a, +\infty)$ is continuous and has derivative $\sigma' = -\rho'\rho$ also continuous at a as $\rho(a) = 0$. Finally $2\sigma' + \rho'' \notin L^2$.

This is why one should not take $\sigma = 0$ but rather $\sigma = -k(a)/2$ on $(a, +\infty)$ and still $\rho = 0$. This is correct as $k(a) > 0$: indeed this comes again from the fact that $\phi'' = k\phi$: if $k(a) < 0$ then, as $\phi > 0$ on $[0, a)$ and k is nondecreasing, one would obtain that ϕ' is (strictly) decreasing on $[0, a)$. Starting with $\phi'(0) = 0$ we obtain that $\phi'(a) < 0$, but $\phi = 0$ on $[a, +\infty)$ so one should have $\phi'(a) = 0$. With this setting, $2\sigma + \rho'$ is continuous at a and two dirac masses compensate each other so that $2\sigma' + \rho'' \in L^2$. And, just as before, $2\sigma' + \rho'' = -\psi$ so strong duality holds.

4.4 Back to the discrete problem

Recall we denoted

$$D(\sigma, \rho) = \frac{1}{\sqrt{2}} \left\{ \rho'(0) - \frac{1}{2} \int_0^\infty |2\sigma' + \rho''|^2 \right\}$$

$$D^h(\sigma, \rho) = \frac{1}{\sqrt{2}} \left\{ \sigma_1 + \frac{1}{\tau} \rho_1 - \frac{\tau}{2} \sum_{n=1}^{N-1} \left(\frac{\sigma_{n-1} - \sigma_{n+1}}{\tau} + \frac{2\rho_n - \rho_{n-1} - \rho_{n+1}}{\tau^2} \right)^2 \right\}$$

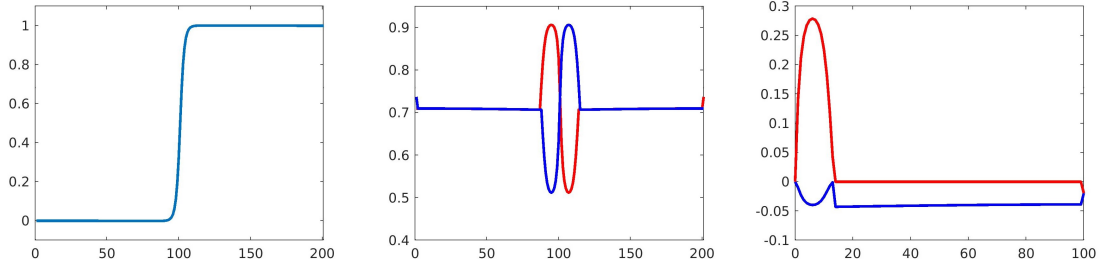


Figure 4: Primal solution u (left), dual solutions p and q (center), corresponding σ (blue) and ρ (red) (right) in the Dirichlet setting with $N = 100$.

the objectives of the continuous and discrete problems respectively (recall that $\tau = N^{-1/3}$). The constraints on $\sigma, \rho : \mathbb{R}^+ \rightarrow \mathbb{R}$ and $\boldsymbol{\sigma}, \boldsymbol{\rho} \in \mathbb{R}^{N+1}$ are

$$\begin{aligned} \sigma(0) = \rho(0) = 0 \text{ and } 2\sigma + \rho^2 \leq 0 \text{ on } \mathbb{R}^+ \\ \boldsymbol{\sigma}_0 = \boldsymbol{\rho}_0 = 0 \text{ and } \forall n \geq 1, N^{-2/3}\boldsymbol{\sigma}_n^2 + 2\boldsymbol{\sigma}_n + \boldsymbol{\rho}_n^2 \leq 0 \end{aligned}$$

Given an admissible (σ, ρ) of the continuous problem with $D(\sigma, \rho) > 0$ we chose the following discretization: set $\boldsymbol{\sigma}_0 = 0$ and $\forall n \geq 1, \boldsymbol{\sigma}_n = \sigma(\tau n) - \tau$ and $\boldsymbol{\rho}_n = \rho(\tau n)$ for all n . Then - provided σ is bounded - $(\boldsymbol{\sigma}, \boldsymbol{\rho})$ is indeed admissible in the discrete problem as $\boldsymbol{\sigma}_0 = \boldsymbol{\rho}_0 = 0$ and

$$\begin{aligned} N^{-2/3}\boldsymbol{\sigma}_n^2 + 2\boldsymbol{\sigma}_n + \boldsymbol{\rho}_n^2 &= N^{-2/3}(\sigma(\tau n) - \tau)^2 - 2\tau + 2\sigma(\tau n) + \rho(\tau n)^2 \\ &\leq N^{-2/3}(\sigma(\tau n) - \tau)^2 - 2N^{-1/3} \end{aligned}$$

with this quantity being nonpositive as soon as $|\sigma(\tau n) - N^{-1/3}| \leq \sqrt{2}N^{1/6}$ which is true for N sufficiently large when σ is bounded.

Therefore we just need to check that with this discretization $D^h(\boldsymbol{\sigma}, \boldsymbol{\rho})$ converges to $D(\sigma, \rho)$ when $N \rightarrow \infty$ as expected in the first place. First note that $\boldsymbol{\sigma}_1 = \sigma(\tau) - \tau \rightarrow \sigma(0) = 0$ (as long as σ is continuous) and that $\frac{1}{\tau}\boldsymbol{\rho}_1 = \frac{\rho(\tau) - \rho(0)}{\tau} \rightarrow \rho'(0)$. As a result we focus next on the convergence of the Riemann sum towards the desired integral.

Second, we can in fact take $\boldsymbol{\sigma}$ to be $\boldsymbol{\sigma}_n = \sigma(\tau n)$. Indeed this only affects the first term of the sum adding:

$$-\frac{\tau}{2} \left| \frac{\sigma(2\tau) - \tau}{\tau} + \frac{\rho(2\tau) - 2\rho(\tau) + \rho(0)}{\tau^2} \right|^2 + \frac{\tau}{2} \left| \frac{\sigma(2\tau) - \sigma(0)}{\tau} + \frac{\rho(2\tau) - 2\rho(\tau) + \rho(0)}{\tau^2} \right|^2$$

with $\frac{\rho(2\tau) - 2\rho(\tau) + \rho(0)}{\tau^2} \rightarrow \rho''(0)$, $\frac{\sigma(2\tau) - \sigma(0)}{\tau} \rightarrow 2\sigma'(0)$ and $\frac{\sigma(2\tau) - \tau}{\tau} \rightarrow 2\sigma'(0) - 1$, so that this quantity tends to zero when $\tau \rightarrow 0$.

To ensure the convergence of the sum, we will need additional regularity on σ and ρ . In the compact support case, we find a new couple (σ, ρ) , more regular and still satisfying $D(\sigma, \rho) > 0$ whereas in the positive case we stick with the (σ, ρ) defined above but show they decrease exponentially fast.

4.4.1 Compact support case

Recall that in this case we have $\sigma, \rho : \mathbb{R}^+ \rightarrow \mathbb{R}$ satisfying $D(\sigma, \rho) > 0$ with $\rho = \sigma' = 0$ on $(a, +\infty)$ and σ, ρ of class \mathcal{C}^∞ on $[0, \infty) \setminus \{a\}$. We extend σ and ρ to \mathbb{R}^- by 0 and regularize them into \mathcal{C}^∞ functions on $[0, \infty)$ while keeping their admissibility in problem (17) as well as the compactness of their support and the value of $\rho'(0)$.

To this end, we first regularize by convolution with a function $\eta \in \mathcal{C}_c^\infty(\mathbb{R})$ with $\eta \geq 0$, $\int_0^\infty \eta = 1$, and $\eta(x) = 0$ for any $x \notin (0, 1)$: we obtain functions $\rho_\varepsilon = \int_{\mathbb{R}} \rho(\cdot + \varepsilon t) \eta(t) dt$ and $\sigma_\varepsilon = \int_{\mathbb{R}} \sigma(\cdot + \varepsilon t) \eta(t) dt$ which are \mathcal{C}^∞ on $[0, \infty)$ and satisfy $\rho_\varepsilon = \sigma'_\varepsilon = 0$ on (a, ∞) as well as $2\sigma_\varepsilon + \rho_\varepsilon^2 \leq 0$ since this constraint is convex, that is $C = \{(s, r) \in \mathbb{R}^2 : 2s + r^2 \leq 0\}$ is a convex set.

However, we lost the values of $\rho(0), \sigma(0)$ and more importantly of $\rho'(0)$ which appears in problem (17). To this end, take $\nu \in \mathcal{C}^\infty$ a plateau function such that $\nu = 1$ on $(-\infty, \frac{a}{3})$ and $\nu = 0$ on $(\frac{2a}{3}, +\infty)$, and set $\hat{\sigma}_\varepsilon = \nu\sigma + (1 - \nu)\sigma_\varepsilon$, $\hat{\rho}_\varepsilon = \nu\rho + (1 - \nu)\rho_\varepsilon$. As σ and ρ are \mathcal{C}^∞ on $[0, +\infty)$ except in a which is avoided, $\hat{\sigma}_\varepsilon$ and $\hat{\rho}_\varepsilon$ are \mathcal{C}^∞ on $[0, +\infty)$, and as $\hat{\rho}_\varepsilon = \rho$, $\hat{\sigma}_\varepsilon = \sigma$ near 0 we keep $\hat{\sigma}_\varepsilon(0) = \hat{\rho}_\varepsilon(0) = 0$ and $\hat{\rho}'_\varepsilon(0) = \rho'(0)$. Furthermore, the constraint $2\hat{\sigma}_\varepsilon + \hat{\rho}_\varepsilon^2 \leq 0$ is still fulfilled by convexity. Finally one checks that:

$$2\hat{\sigma}'_\varepsilon + \hat{\rho}''_\varepsilon = 2\sigma'_\varepsilon + \rho''_\varepsilon + \{2(\sigma' - \sigma'_\varepsilon) + (\rho'' - \rho''_\varepsilon)\}\nu + \{(\sigma - \sigma_\varepsilon) + 2(\rho' - \rho'_\varepsilon)\}\nu' + \{\rho - \rho_\varepsilon\}\nu''$$

so that when ε goes to 0:

- $2\sigma'_\varepsilon + \rho''_\varepsilon$ converges to $2\sigma' + \rho''$ in $L^2(0, \infty)$.
- σ', ρ'' are continuous on $[0, \frac{2a}{3}]$ hence $2(\sigma' - \sigma'_\varepsilon) + (\rho'' - \rho''_\varepsilon)$ converges to 0 uniformly on $[0, \frac{2a}{3}]$. As $\nu = 0$ on $(\frac{2a}{3}, +\infty)$ this implies that $\{2(\sigma' - \sigma'_\varepsilon) + (\rho'' - \rho''_\varepsilon)\}\nu$ converges to 0 in $L^2(0, \infty)$.
- $\nu' = \nu'' = 0$ outside of $[\frac{a}{3}, \frac{2a}{3}]$ where σ, σ' and ρ' are continuous hence $\{(\sigma - \sigma_\varepsilon) + 2(\rho' - \rho'_\varepsilon)\}\nu' + \{\rho - \rho_\varepsilon\}\nu''$ converges to 0 uniformly hence in $L^2(0, \infty)$.

To conclude, $D(\hat{\sigma}_\varepsilon, \hat{\rho}_\varepsilon) \rightarrow D(\sigma, \rho)$. This shows that one can find (σ, ρ) admissible in the continuous problem such that $D(\sigma, \rho) > 0$ and σ, ρ are \mathcal{C}^∞ on $[0, +\infty)$, with ρ and σ' having compact supports. In particular all the functions $\sigma, \sigma', \sigma'', \rho, \rho', \rho''$ and ρ''' can be uniformly bounded by some constant $M > 0$.

Then to estimate convergence of $D^h(\sigma, \rho)$ towards $D(\sigma, \rho)$ we can truncate the Riemann sum at $n = \lfloor \frac{a}{\tau} \rfloor$ where the supports of σ' and ρ are included in $[0, a]$. Doing so it is easy to show that

$$\tau \sum_{n=1}^{N-1} \left| \frac{\sigma_{n+1} - \sigma_{n-1}}{\tau} + \frac{\rho_{n+1} - 2\rho_n + \rho_{n-1}}{\tau^2} \right|^2 = \tau \sum_{n=1}^{\lfloor \frac{a}{\tau} \rfloor} |2\sigma'(\tau n) + \rho''(\tau n)|^2 + O(\tau)$$

And we conclude saying that as $(2\sigma' + \rho'')^2$ is Riemann integrable one has

$$\tau \sum_{n=1}^{\lfloor \frac{a}{\tau} \rfloor} |2\sigma'(\tau n) + \rho''(\tau n)|^2 \rightarrow \int_0^a (2\sigma' + \rho'')^2 = \int_0^\infty (2\sigma' + \rho'')^2$$

hence the desired convergence.

4.4.2 Positive case

Recall that in this case we have $\sigma, \rho : \mathbb{R}^+ \rightarrow \mathbb{R}$ satisfying $D(\sigma, \rho) > 0$ with $\sigma = -\rho^2/2$ and $\rho = -\phi'/\phi$ for some $\phi > 0$ C^∞ on \mathbb{R}^+ . We also had that $\phi' \leq 0$ and $\phi'' = k\phi$ with $k(t) = \int_0^t s\phi^2(s)ds + t\psi(t) - A$ nondecreasing. Therefore ρ satisfies on \mathbb{R}^+

$$\rho' = -\frac{\phi''}{\phi} + \frac{\phi'^2}{\phi^2} = \rho^2 - k$$

This relation allows us to show that the derivatives of ρ tends to 0 exponentially fast, which will compensate the non compactness of their support. It is important to note that the key argument in the following proofs is that this relation holds on the whole \mathbb{R}^+ : in the case of compact support it only holds on $[0, a)$ and one cannot obtain the same conclusions (especially, in the compact support case, we cannot have $\rho'(t) \geq 0$ for all $t \geq 0$ as shown below). Our analysis begins with the two following lemmas that derive from easy manipulations and antidifferentiation and for which we only give sketches of the proofs.

Lemma 2. *Let $\rho, k : \mathbb{R}^+ \rightarrow \mathbb{R}$ be C^1 functions such that for all $t \geq 0$, $\rho'(t) = \rho^2(t) - k(t)$, $\rho(t) \geq 0$ and $k'(t) \geq 0$. Then for all $t \geq 0$, $\rho'(t) \geq 0$.*

Proof. Suppose $\rho'(t) = -r < 0$ for some $t \geq 0$, then one can prove that ρ is nonincreasing on (t, ∞) . But then so is $\rho' = \rho^2 - k$ as $\rho, k' \geq 0$. Consequently, $\rho'(s) \leq -r$ for any $s \geq t$ which cannot stand with the hypothesis that $\rho \geq 0$. \square

Lemma 3. *Let $t_1 \in \mathbb{R}$ and let $\rho : [t_1, +\infty[\rightarrow \mathbb{R}^+$ be a C^1 function. There is no $L \in \mathbb{R}$ such that $\forall t \geq t_1$*

$$\rho^2(t) - L \neq 0 \quad \text{and} \quad \frac{\rho'(t)}{\rho^2(t) - L} \geq 1$$

Proof. The case $L = 0$ is clear. Otherwise, one integrates $\frac{\rho'}{\rho^2 - L}$ as $\log\left(\frac{\rho - \sqrt{L}}{\rho + \sqrt{L}}\right)$ if $L > 0$ or as $\arctan\left(\frac{\rho}{\sqrt{-L}}\right)$ if $L < 0$. In either cases, taking the limit at infinity leads to a contradiction. \square

Thanks to the first lemma, ρ is nonnegative and nondecreasing (and not zero everywhere), so $\rho(t) \rightarrow R \in (0, +\infty]$ when $t \rightarrow \infty$. In particular there exists $c > 0$ and $t_0 > 0$ such that $\forall t \geq t_0$, $-\frac{\phi'(t)}{\phi(t)} = \rho(t) \geq c > 0$ which leads to $\phi(t) \leq \phi(t_0) \exp(-c(t - t_0))$. As a consequence, $k(t) = \int_0^t s\phi^2(s)ds + t \int_t^\infty \phi^2(s)ds - A$ is bounded and increasing so converges to some $L \in \mathbb{R}$ and the convergence is exponential since :

$$L - k(t) = \int_t^\infty (s - t)\phi^2(s)ds \leq M \exp(-2ct) \text{ for some } M > 0$$

Next we must have $R < +\infty$. Indeed, otherwise we would have a $t_1 > 0$ such that $\forall t \geq t_1$, $\rho'(t) = \rho^2(t) - k(t) \geq \rho^2(t) - L > 0$ hence $\frac{\rho'(t)}{\rho^2(t) - L} \geq 1$ which is not possible according to the second lemma.

Hence $R^2 \leq L$, while since $\rho' = \rho^2 - k$ remains nonnegative and converges to $R^2 - L$, $R^2 = L$ and finally,

$$\forall t \geq 0, \rho'(t) = \rho^2(t) - L + L - k(t) \leq L - k(t) \leq M \exp(-2ct)$$

As a consequence, $\sigma', \rho'', \sigma''$ and ρ''' decrease exponentially to zero. Indeed:

- $\sigma' = -\rho'\rho$ with ρ bounded.
- $\rho'' = 2\sigma'\sigma - \psi$ with $\sigma = -\rho^2/2$ bounded and ψ decreasing exponentially to zero (as $\psi(t) = \int_t^\infty \phi^2$ with ϕ decreasing exponentially).
- $\sigma'' = -\rho'^2 - \rho''\rho$.
- $\rho''' = 2\rho''\rho + 2\rho'^2 + \phi^2$.

Then we get the following estimate for our discretization: write for $1 \leq n \leq N-1$

$$\frac{\sigma_{n+1} - \sigma_{n-1}}{\tau} = 2\sigma'(\tau n + \eta_n) \quad \text{and} \quad \frac{\rho_{n+1} - 2\rho_n + \rho_{n-1}}{\tau^2} = \rho''(\tau n + \tilde{\eta}_n)$$

for some $\eta_n, \tilde{\eta}_n \in (-\tau, \tau)$, so that we have:

$$\left| \left| \frac{\sigma_{n+1} - \sigma_{n-1}}{\tau} + \frac{\rho_{n+1} - 2\rho_n + \rho_{n-1}}{\tau^2} \right|^2 - \tau \sum_{n=1}^{N-1} |2\sigma'(\tau n) - \rho''(\tau n)|^2 \right| = \Delta_n^- \times \Delta_n^+$$

with

$$\begin{aligned} \Delta_n^- &:= |2\sigma'(\tau n + \eta_n) - 2\sigma'(\tau n) + \rho''(\tau n + \tilde{\eta}_n) - \rho''(\tau n)| \\ &\leq 2\tau \times (2\|\sigma''\|_{\infty, (\tau n - \tau, \tau n + \tau)} + \|\rho'''\|_{\infty, (\tau n - \tau, \tau n + \tau)}) \\ &\leq \tau M \exp(-c(\tau n - \tau)) \\ \Delta_n^+ &:= |2\sigma'(\tau n + \eta_n) + 2\sigma'(\tau n) + \rho''(\tau n + \tilde{\eta}_n) + \rho''(\tau n)| \\ &\leq 4\|\sigma'\|_{\infty, (\tau n - \tau, \tau n + \tau)} + 2\|\rho'''\|_{\infty, (\tau n - \tau, \tau n + \tau)} \\ &\leq \tau M \exp(-c(\tau n - \tau)) \end{aligned}$$

for some constants $M, c > 0$ and finally one can write (for other constants $M, c > 0$):

$$\begin{aligned} &\left| \tau \sum_{n=1}^{N-1} \left| \frac{\sigma_{n+1} - \sigma_{n-1}}{\tau} + \frac{\rho_{n+1} - 2\rho_n + \rho_{n-1}}{\tau^2} \right|^2 - \tau \sum_{n=1}^{N-1} |2\sigma'(\tau n) - \rho''(\tau n)|^2 \right| \\ &\leq \tau^2 \sum_{n=1}^{N-1} M \exp(-c(\tau n - \tau)) \\ &\leq M\tau^2 \sum_{n=0}^{\infty} \exp(-c\tau)^n = M \frac{\tau^2}{1 - \exp(-c\tau)} \sim M \frac{\tau^2}{c\tau} \rightarrow 0 \text{ as } N \rightarrow \infty \end{aligned}$$

To conclude (i.e. to obtain $D^h(\boldsymbol{\sigma}, \boldsymbol{\rho}) \rightarrow D(\boldsymbol{\sigma}, \boldsymbol{\rho})$), we state that

$$\tau \sum_{n=1}^{N-1} (2\sigma'(\tau n) + \rho''(\tau n))^2 \rightarrow \int_0^\infty (2\sigma' + \rho'')^2 \text{ as } N \rightarrow \infty$$

This comes from taking $f = (2\sigma' + \rho'')^2 = \psi^2$ – which is indeed nonincreasing as $\psi' = -\phi^2 \leq 0$ and $\psi \geq 0$ – in the following easy result:

Proposition 5. *Let $f : \mathbb{R}^+ \rightarrow \mathbb{R}$ be a continuous and nonincreasing function such that $\int_0^\infty f$ converges. Let $a > b > 0$ and $c_1, c_2, c_3 \in \mathbb{R}$ constants. Then*

$$\frac{1}{N^b} \sum_{l=\lfloor c_1 \rfloor}^{\lfloor c_2 N^a + c_3 \rfloor} f\left(\frac{l}{N^b}\right) \rightarrow \int_0^\infty f \text{ when } N \rightarrow \infty$$

4.5 Modifications for Neumann boundary conditions

Dealing with Neumann boundary conditions takes us back to the 1D problem (7), where we know take $u_{N+1} = u_N$ and $u_{-N-1} = u_{-N}$. We also suppose $\lambda < \lambda^*$ so that $\bar{u} \neq \frac{1}{2}$. Thanks to Proposition 2, we can suppose $p_0 = q_0 = \sqrt{2}/2$ in the dual problem (16), and one checks that it is changed into

$$\bar{E}^h = \max_{\substack{p_n^2 + p_{-n}^2 \leq 1 \\ -N \leq n \leq N \\ p_0 = \sqrt{2}/2}} \frac{1}{2} + \frac{1}{\sqrt{2}} p_1 - \frac{\lambda}{h\sqrt{2}} \sum_{n=1}^{N-1} (p_{-n+1} - p_{-n} + p_n - p_{n+1})^2 - \frac{\lambda}{h\sqrt{2}} (p_{-N+1} + p_N)^2$$

Remember from Section 2.2 that the limit value when $h = \frac{1}{N} \rightarrow 0$ is $\bar{E} = \bar{E}_N = 1 - \sqrt{2}\lambda$. This value is (almost) achieved when taking $p_n = \sqrt{2}/2 - |n|/\sqrt{2}N$ as it gives $\bar{E}^h \geq 1 - \sqrt{2}\lambda + \frac{3\lambda - \sqrt{2}}{2\sqrt{2}}h$ (but $3\lambda - \sqrt{2} < 0$). Let us denote

$$F(p, \lambda) = \frac{1}{2} + \frac{1}{\sqrt{2}} p_1 - \frac{\lambda}{\sqrt{2}h} \sum_{n=1}^{N-1} (p_{-n+1} - p_{-n} + p_n - p_{n+1})^2$$

$$\tilde{F}(\tilde{p}, \lambda) = \frac{1}{2} + \frac{1}{\sqrt{2}} \tilde{p}_1 - \frac{\lambda}{\sqrt{2}h} \sum_{n=1}^{N-1} (\tilde{p}_{-n+1} - \tilde{p}_{-n} + \tilde{p}_n - \tilde{p}_{n+1})^2 - \frac{\lambda}{\sqrt{2}h} (\tilde{p}_{-N+1} + \tilde{p}_N)^2$$

Note that the constraint on p in Dirichlet and Neumann problems is the same: $p_0 = \sqrt{2}/2$ and $p_n^2 + p_{-n}^2 \leq 1$. Now suppose p is the Dirichlet variable constructed in the previous sections, and form $\tilde{p}_n = p_n - \frac{|n|}{\sqrt{2}N}$. We want to compare $\tilde{F}(\tilde{p}, \lambda) - \bar{E}_N$ to $F(p, \lambda) - \bar{E}_D$. As $\bar{E}_N = 1 - \lambda\sqrt{2} = \bar{E}_D - \lambda\sqrt{2}$, we split $\lambda\sqrt{2}$ into $N \times \frac{\lambda}{\sqrt{2}h} \times \frac{2}{N^2}$ and allocate each $\frac{2}{N^2}$ to a term involving p^2 in the expression of \tilde{E} . We obtain:

$$(\tilde{p}_{-n+1} - \tilde{p}_{-n} + \tilde{p}_n - \tilde{p}_{n+1})^2 - \frac{2}{N^2} = (p_{-n+1} - p_{-n} + p_n - p_{n+1} + \frac{\sqrt{2}}{N})^2 - \frac{2}{N^2} = x_n^2 + \frac{2\sqrt{2}}{N} x_n$$

where we denoted $x_n = p_{-n+1} - p_{-n} + p_n - p_{n+1}$. When summing, we will recover the term in x_n^2 appearing in $E(p, \lambda)$, along with

$$\sum_{n=1}^{N-1} x_n = p_1 - p_N + p_0 - p_{-N+1} = (p_1 - \frac{\sqrt{2}}{2}) - (p_N + p_{N-1} - \sqrt{2})$$

Besides, one has

$$(\tilde{p}_{-N+1} + \tilde{p}_N)^2 - \frac{2}{N^2} = (p_{-N+1} + p_N - \sqrt{2})^2 + \frac{\sqrt{2}}{N}(p_{-N+1} + p_N - \sqrt{2}) - \frac{3}{N^2}$$

Then we obtain:

$$\begin{aligned} \tilde{F}(\tilde{p}, \lambda) - \bar{E}_N &= \frac{1}{2} + \frac{1}{\sqrt{2}}p_1 - \frac{1}{2N} - 1 - \frac{\lambda}{\sqrt{2}h} \sum_{n=1}^{N-1} x_n^2 \\ &\quad - \frac{\lambda}{\sqrt{2}h} \times \frac{2\sqrt{2}}{N} (p_1 - \frac{\sqrt{2}}{2}) - \frac{\lambda}{\sqrt{2}h} (p_{-N+1} + p_N - \sqrt{2})^2 \\ &\quad - \frac{\lambda}{\sqrt{2}h} \times \frac{\sqrt{2}}{N} (p_{-N+1} + p_N - \sqrt{2}) + \frac{\lambda}{\sqrt{2}h} \frac{3}{N^2} \\ &= F(p, \lambda) - \bar{E}_D - 2\lambda(p_1 - \frac{\sqrt{2}}{2}) + R \end{aligned} \quad (21)$$

where $R = \lambda(p_N + p_{-N+1} - \sqrt{2}) - \frac{\lambda}{\sqrt{2}h} (p_N + p_{-N+1} - \sqrt{2})^2 + \frac{3\sqrt{2}\lambda-1}{2N}$.

At this point, remember p was obtained from continuous functions σ and ρ through

$$\begin{cases} p_n = \frac{1}{\sqrt{2}}(\sigma_n + 1 + \rho_n) ; p_{-n} = \frac{1}{\sqrt{2}}(\sigma_n + 1 - \rho_n) \\ \text{with } \sigma_n = N^{-2/3}(\sigma(\tau n) - \tau) ; \rho_n = N^{-1/3}\rho(\tau n) \end{cases}$$

As ρ and σ are bounded, one sees that p_n converges to $\frac{\sqrt{2}}{2}$ uniformly as N goes to infinity (that is $\max_{-N \leq n \leq N} |p_n - \frac{\sqrt{2}}{2}| \rightarrow 0$ as $N \rightarrow \infty$). This first shows that \tilde{p} is admissible in the dual problem (meaning that $\tilde{p}_n^2 + \tilde{p}_{-n}^2 \leq 1$): indeed p is itself admissible and $p_n \geq \tilde{p}_n \geq -1 \geq -p_n$ for N sufficiently large. Second, remember that, at infinity, σ converges to $-k(a) < 0$ or 0, and ρ converges to 0. Writing

$$p_{-N+1} + p_N - \sqrt{2} = \frac{1}{\sqrt{2}}(\sigma_N + \sigma_{N-1} + \rho_N - \rho_{N-1})$$

one sees that $N^{2/3}R \rightarrow 0$ when $N \rightarrow \infty$. Then we apply a last trick to include $2\lambda(p_1 - \frac{\sqrt{2}}{2})$ from (21) into our energies: we remark that

$$F(p, \lambda) - \bar{E}_D - 2\lambda(p_1 - \frac{\sqrt{2}}{2}) = (1 - 2\sqrt{2}\lambda)(F(p, \frac{\lambda}{1 - 2\sqrt{2}\lambda}) - \bar{E}_D)$$

This finally shows that

$$N^{2/3}(\tilde{F}(\tilde{p}, \lambda) - \bar{E}_N) = N^{2/3}\left((1 - 2\sqrt{2}\lambda)(F(p, \frac{\lambda}{1 - 2\sqrt{2}\lambda}) - \bar{E}_D)\right) + N^{2/3}R$$

converges to a positive value when N tends to infinity; hence the $O(h^{2/3})$ rate is also true in the Neumann setting.

5 Raviart-Thomas total variation

5.1 Definitions

The idea behind the definition of the isotropic total variation is of course to catch the L^1 norm of the gradient of u based on a discretization of the expression $\text{TV}(u) = \int_{\Omega} |\nabla u|$. To do so, one chooses a finite differences operator D , defined on the mesh $\Omega = \cup C_{i,j}$, to approximate ∇ . However, the non isotropy of the grid itself prevents D from being isotropic, as it has to involve a notion of neighbor on this two-directional grid. On the contrary, the dual definition of TV offers the possibility to discretize a field rather than an operator. In the formulas

$$\text{TV}_N(u) = \sup \left\{ - \int_{\Omega} u \operatorname{div} \phi, \phi \in \mathcal{C}_c^1(\Omega, \mathbb{R}^2) \text{ s.t. } \|\phi\|_{\infty} \leq 1 \right\} \quad (22)$$

$$\text{TV}_D(u) = \sup \left\{ - \int_{\Omega} u \operatorname{div} \phi + \int_{\partial\Omega} b \langle \phi | \vec{n} \rangle, \phi \in \mathcal{C}^1(\Omega, \mathbb{R}^2) \text{ s.t. } \|\phi\|_{\infty} \leq 1 \right\} \quad (23)$$

we will keep the exact operator div but replace the spaces $\mathcal{C}_c^1(\Omega, \mathbb{R}^2)$ and $\mathcal{C}^1(\Omega, \mathbb{R}^2)$ of (compactly supported) \mathcal{C}^1 fields from Ω to \mathbb{R}^2 , by a space of discrete fields favouring no direction.

The most simple space available is the so-called "Raviart-Thomas" finite elements space [14], which first seems to have been used in this context in [10]. Raviart-Thomas fields are defined via their fluxes through the edges of the squares, we will denote $f_{i+\frac{1}{2},j}$ (resp. $f_{i,j+\frac{1}{2}}$) the flux through the edge between the squares $C_{i,j}$ and $C_{i+1,j}$ (resp. $C_{i,j}$ and $C_{i,j+1}$), and $(x_{i,j}, y_{i,j})$ the center of the square $C_{i,j}$. Then the Raviart-Thomas fields are the elements of

$$\begin{aligned} RT0 = & \left\{ \phi : \Omega \rightarrow \mathbb{R}^2 : \exists (f_{i+\frac{1}{2},j}, f_{i,j+\frac{1}{2}})_{i,j} \text{ s.t. } \forall 1 \leq i, j \leq N, \right. \\ & \left. \phi(x, y) = \left(\begin{array}{l} \frac{f_{i+\frac{1}{2},j} + f_{i-\frac{1}{2},j}}{2} + (f_{i+\frac{1}{2},j} - f_{i-\frac{1}{2},j}) \frac{x - x_{i,j}}{h} \\ \frac{f_{i,j+\frac{1}{2}} + f_{i,j-\frac{1}{2}}}{2} + (f_{i,j+\frac{1}{2}} - f_{i,j-\frac{1}{2}}) \frac{y - y_{i,j}}{h} \end{array} \right) \text{ in } C_{i,j} \right\} \end{aligned} \quad (24)$$

In the sequel, we will write $\phi = \phi_f \in RT0$ to precise that f denotes the fluxes of the Raviart-Thomas fields ϕ according to (24). In the Neumann setting, we use Raviart-Thomas fields vanishing on the boundary of Ω , which we denote $RT0_0$:

$$RT0_0 = \left\{ \phi_f \in RT0 \text{ s.t. } \forall 1 \leq i, j \leq N, f_{\frac{1}{2},j} = f_{N+\frac{1}{2},j} = f_{i,\frac{1}{2}} = f_{i,N+\frac{1}{2}} = 0 \right\}$$

Finally, in the Neumann setting, the Raviart-Thomas total variation we study is, for any $u^h \in P0$:

$$\text{TV}_{RT,N}^h(u^h) = \sup \left\{ - \int_{\Omega} u^h \operatorname{div} \phi, \phi = \phi_f \in RT0_0 \text{ s.t. } \|\phi\|_{\infty} \leq 1 \right\} \quad (25)$$

while in the Dirichlet setting, we directly use the source term g^h of the ROF problem in the integral on $\partial\Omega$:

$$\text{TV}_{RT,D}^h(u^h) = \sup \left\{ - \int_{\Omega} u^h \operatorname{div} \phi + \int_{\partial\Omega} b^h \langle \phi | \vec{n} \rangle, \phi = \phi_f \in RT0 \text{ s.t. } \|\phi\|_{\infty} \leq 1 \right\} \quad (26)$$

We stress the fact that no continuity jump appears in the calculus of $\operatorname{div} \phi_f$ so that, for instance in the Neumann setting, for $\phi_f \in RT0_0$:

$$\begin{aligned} - \int_{\Omega} u^h \operatorname{div} \phi_f &= - \sum_{i,j} h^2 u_{i,j}^h \frac{1}{h} (f_{i+\frac{1}{2},j} - f_{i-\frac{1}{2},j} + f_{i,j+\frac{1}{2}} - f_{i,j-\frac{1}{2}}) \\ &= h \sum_{i,j} f_{i+\frac{1}{2},j} (u_{i+1,j}^h - u_{i,j}^h) + h \sum_{i,j} f_{i,j+\frac{1}{2}} (u_{i,j+1}^h - u_{i,j}^h) \\ &= h \sum_{i,j} \left\langle \begin{pmatrix} f_{i+\frac{1}{2},j} \\ f_{i,j+\frac{1}{2}} \end{pmatrix} \middle| \begin{pmatrix} (u^h)_{i+1,j} - (u^h)_{i,j} \\ (u^h)_{i,j+1} - (u^h)_{i,j} \end{pmatrix} \right\rangle = h \langle \langle f | Du^h \rangle \rangle \end{aligned}$$

In particular, as noted by the authors of [13], the isotropic total variation can be recovered in the context of Raviart-Thomas field total variation as

$$\text{TV}_i^h(u^h) = \sup \left\{ - \int_{\Omega} u^h \operatorname{div} \phi, \phi = \phi_f \in RT0_0 \text{ s.t. } \forall i, j, \left| \begin{pmatrix} f_{i+\frac{1}{2},j} \\ f_{i,j+\frac{1}{2}} \end{pmatrix} \right| \leq 1 \right\}$$

In TV_{RT}^h , the constraint on ϕ_f is the same as on ϕ on the continuous TV, namely that $|\phi_f(x)| \leq 1$ for all $x \in \Omega$. Note that since the two components of ϕ_f are piecewise affine, the constraint of being less than 1 everywhere on Ω reduces to being less than 1 in the corners of the mesh, that is

$$\text{TV}_{RT,N}^h(u_h) = \sup \left\{ - \int_{\Omega} u_h \operatorname{div} \phi_f, \phi_f \in RT0_0 \text{ s.t.} \right. \\ \left. \forall 1 \leq i, j \leq N, \max_{1 \leq k \leq 4} |(L_k \phi_f)_{i,j}| \leq 1 \right\}$$

$$\text{TV}_{RT,D}^h(u_h) = \sup \left\{ - \int_{\Omega} u_h \operatorname{div} \phi_f + \int_{\partial\Omega} b^h \langle \phi_f | \vec{n} \rangle, \phi_f \in RT0 \text{ s.t.} \right. \\ \left. \forall 1 \leq i, j \leq N, \max_{1 \leq k \leq 4} |(L_k \phi_f)_{i,j}| \leq 1 \right\}$$

$$\text{with } \begin{cases} (L_1 \phi_f)_{i,j} = \phi_f(x_{i,j} + \frac{h}{2}^-, y_{i,j} + \frac{h}{2}^-) \\ (L_2 \phi_f)_{i,j} = \phi_f(x_{i,j} - \frac{h}{2}^-, y_{i,j} + \frac{h}{2}^-) \\ (L_3 \phi_f)_{i,j} = \phi_f(x_{i,j} - \frac{h}{2}^-, y_{i,j} - \frac{h}{2}^-) \\ (L_4 \phi_f)_{i,j} = \phi_f(x_{i,j} + \frac{h}{2}^-, y_{i,j} - \frac{h}{2}^-) \end{cases} \quad \begin{array}{|c|c|} \hline \bullet & \bullet \\ \hline 2 & 1 \\ \hline (x_{i,j}, y_{i,j}) & \\ \hline \times & \\ \hline 3 & 4 \\ \hline \bullet & \bullet \\ \hline \end{array}$$

where we used the notation a^- (a^+) to denote we take the left (right) limit. Other choices of constraints on ϕ_f proposed in [9, 11] lead to better numerical results, however we do not have convergence rates for them. Nevertheless, they also fit the framework of Raviart-Thomas total variations, see Section 6.

5.2 Convergence rate for TV_{RT}^h

In [7] the authors have studied Crouzeix-Raviart finite elements based total variation on a triangular mesh, which can be computed by approximating the dual fields with Raviart-Thomas fields with a norm constraint only in the center of each triangle. Given a source term $g \in L^\infty$, and under a regularity assumption on the dual field, they show there exists a constant c (depending on g and the value of the continuous ROF problem) such that $|\bar{E} - \bar{E}^h| \leq ch$ where we recall that \bar{E} and \bar{E}^h are respectively the optimal values of the continuous and discrete problems:

$$\bar{u} = \arg \min_{u \in BV(\Omega)} \frac{1}{2\lambda} \|u - g\|_{L^2}^2 + \text{TV}(u) =: E(u) \quad (27)$$

$$\bar{u}^h = \arg \min_{u^h \in P_0} \frac{1}{2\lambda} \|u^h - g^h\|_{L^2}^2 + \text{TV}_{RT}^h(u^h) =: E^h(u^h) \quad (28)$$

with appropriate variants for Dirichlet and Neumann boundary conditions (recall that when no subscript N or D is specified, the proposed results are valid for both settings). Thanks to the strong convexity of the energy these estimates are also controlling the squared L^2 error between \bar{u} and \bar{u}^h . This study easily transposes to our context. We emphasize in this section the main lemmas that result in the desired convergence rate. We refer the reader to [7] for more details and to [2] for a generalization.

The proof of this rate is two-fold: a first estimate comes from the primal problems, a second from the dual. The first one relies on the conformal aspect of our discrete total variation TV_{RT}^h (25), (26) with respect to the continuous TV (22), (23). As in [1], it follows from the TV-diminishing lemma:

Lemma 4. *For any $u \in BV \cap L^2(\Omega)$ admissible in the continuous ROF problem (1), one has $\text{TV}_{RT}^h(\Pi_{P_0} u) \leq \text{TV}(u)$.*

Proof. The main argument is that if $\phi \in RT_0$, then $\text{div } \phi$ is piecewise constant so that $u^h = \Pi_{P_0}(u)$ satisfies

$$\int_{\Omega} u^h \text{div } \phi = \int_{\Omega} u \text{div } \phi$$

The result follows by approximation of Raviart-Thomas fields by regular fields. \square

Using strong convexity of the primal objectives leads to the first estimate:

Proposition 6. *The solutions \bar{u}, \bar{u}^h of (27), (28) satisfy*

$$\frac{1}{2\lambda} \|\bar{u}^h - \Pi_{P_0} \bar{u}\|_{L^2}^2 \leq \bar{E} - \bar{E}^h - \frac{1}{2\lambda} (\|\bar{u} - g\|_{L^2}^2 - \|\Pi_{P_0}(\bar{u} - g)\|_{L^2}^2)$$

The second part relies on the evaluation of the dual problems of (27) and (28). In the continuous setting, switching the min operator from (27) with the supremum defining

the total variation leads to the following dual problems:

$$\bar{\phi}_N \in \arg \max_{\substack{\phi \in \mathcal{H}^0 \text{ s.t.} \\ \|\phi\|_\infty \leq 1}} - \int_{\Omega} g \operatorname{div} \phi - \frac{\lambda}{2} \|\operatorname{div} \phi\|_2^2 =: D_N(\phi) \quad (29)$$

$$\bar{\phi}_D \in \arg \max_{\substack{\phi \in \mathcal{H} \text{ s.t.} \\ \|\phi\|_\infty \leq 1}} - \int_{\Omega} g \operatorname{div} \phi - \frac{\lambda}{2} \|\operatorname{div} \phi\|_2^2 + \int_{\partial\Omega} b \langle \phi | \vec{n} \rangle =: D_D(\phi) \quad (30)$$

where $\mathcal{H} = \{\phi \in L^\infty(\Omega) \text{ s.t. } \operatorname{div} \phi \in L^2(\Omega)\}$ and \mathcal{H}^0 is the subset of \mathcal{H} made of fields vanishing at the boundary in the weak sense $\mathcal{H}^0 = \{\phi \in \mathcal{H} \text{ s.t. } \forall u \in H^1(\Omega), \int_{\Omega} \langle \nabla u | \phi \rangle = - \int_{\Omega} u \operatorname{div} \phi\}$. Observe, for instance in the Neumann setting, that for any $\phi \in \mathcal{H}^0$ such that $\|\phi\|_\infty \leq 1$, one has $D_N(\phi) \leq E(\bar{u}) = \operatorname{TV}(\bar{u}) + \frac{1}{2\lambda} \|\bar{u} - g\|_{L^2}^2$. The Euler-Lagrange equation for the ROF problem (see [5]) shows that \bar{u} is a minimizer of (27) if and only if there exists $\bar{\phi} \in \mathcal{H}$ with $\bar{u} - g = \lambda \operatorname{div} \bar{\phi}$, $\|\bar{\phi}\|_\infty \leq 1$ and $-\int_{\Omega} \bar{u} \operatorname{div} \bar{\phi} = \operatorname{TV}(\bar{u})$. Choosing $\phi = \bar{\phi}$ in the above inequality shows that strong duality between primal and dual problems holds. Finally, $\bar{D} := D(\bar{\phi}) = \bar{E}$ through the relation $\bar{u} = g + \lambda \operatorname{div} \bar{\phi}$. The same relations hold in the discrete case which is completely similar and where dual problems are:

$$\bar{\phi}_N^h \in \arg \max_{\substack{\phi^h \in RT_0 \\ \|\phi^h\|_\infty \leq 1}} - \int_{\Omega} g^h \operatorname{div} \phi^h - \frac{\lambda}{2} \|\operatorname{div} \phi^h\|_2^2 =: D_N^h(\phi^h)$$

$$\bar{\phi}_D^h \in \arg \max_{\substack{\phi^h \in RT_0 \\ \|\phi^h\|_\infty \leq 1}} - \int_{\Omega} g^h \operatorname{div} \phi^h - \frac{\lambda}{2} \|\operatorname{div} \phi^h\|_2^2 + \int_{\partial\Omega} b^h \langle \phi^h | \vec{n} \rangle =: D_D^h(\phi^h)$$

As previously, one has to be able to get a discrete field from a continuous one through a projection operator like Π_{P_0} . This will be achieved by the operator $\Pi_{RT_0} : \mathcal{H} \rightarrow RT_0$ which takes $z = (z_1, z_2) : \Omega \rightarrow \mathbb{R}^2$ to $\phi = \phi_f \in RT_0$ where the fluxes through the edges of the mesh f are defined as

$$f_{i+1/2,j} = \frac{1}{h} \int_{E_{i+1/2,j}} z_2 ; f_{i,j+1/2} = \frac{1}{h} \int_{E_{i,j+1/2}} z_1$$

where $E_{i+1/2,j} = \partial C_{i,j} \cap \partial C_{i+1,j}$ and $E_{i,j+1/2} = \partial C_{i,j} \cap \partial C_{i,j+1}$. This projection operator enjoys two properties that derive from simple integration formulas. The first one is classical:

Lemma 5. $\forall \phi \in \mathcal{H}$, $\operatorname{div}(\Pi_{RT_0}(\phi)) = \Pi_{P_0}(\operatorname{div} \phi)$.

The second one deals with the behavior of Π_{RT_0} with respect to the infinite norm: when $\phi^h = \Pi_{RT_0}(\phi)$, one sees that the value of $|\phi^h|^2$ at a corner of the square writes as an average of $|\phi|^2$ on an adjacent edge up to a first order derivative term. This leads to the following result:

Lemma 6. *If $\phi : \Omega \rightarrow \mathbb{R}^2$ is L -Lipschitz and if $\|\phi\|_\infty \leq 1$ then its projection $\phi^h = \Pi_{RT_0}(\phi)$ satisfies $\|\phi^h\|_\infty \leq 1 + Lh$.*

In our analysis, we will consequently need a Lipschitz hypothesis to hold on the optimal dual field $\bar{\phi}$. As noticed by [7], this hypothesis is reasonable in the sense that it is known to hold when g is the characteristic of a disk and $\Omega = \mathbb{R}^2$, as well as in the case $g = g_\nu$ (where one can even take $L = 0$ as $\bar{\phi} = \nu$ is a solution). It seems plausible that this hypothesis is satisfied as long as $g \in L^\infty$ (when working in a bounded convex domain of \mathbb{R}^2), however no such result is known for the time being. We now apply these two lemmas to get an admissible solution in the discrete dual problem from a continuous dual solution, and get the second estimate:

Proposition 7. *Suppose the dual continuous problem (29), (30) admits a L -Lipschitz solution, then one has:*

$$\begin{aligned}\bar{D}_N &\leq (1 + Lh)\bar{D}_N^h + \frac{1}{2\lambda}\|g - g^h\|_{L^2(\Omega)}^2 \\ \bar{D}_D &\leq (1 + Lh)\bar{D}_D^h + \frac{1}{2\lambda}\|g - g^h\|_{L^2(\Omega)}^2 + \|b - b^h\|_{L^1(\partial\Omega)}\end{aligned}$$

Now combining Propositions 6 and 7, we deduce, for instance in the Neumann setting, that there exist a constant $c > 0$, depending on the optimal energy \bar{E} such that, thanks to Jensen's inequality:

$$\begin{aligned}\frac{1}{2\lambda}\|\bar{u}^h - \Pi_{P_0}\bar{u}\|_{L^2}^2 &\leq ch + \frac{1}{2\lambda}\left(\|g - g^h\|_{L^2}^2 - (\|\bar{u} - g\|_{L^2}^2 + \|\Pi_{P_0}(\bar{u} - g)\|_{L^2}^2)\right) \\ &= ch + \frac{1}{2\lambda}\left(\|\Pi_{P_0}\bar{u}\|_{L^2}^2 - \|\bar{u}\|_{L^2}^2 + 2\int_{\Omega}\bar{u}g - 2\int_{\Omega}(\Pi_{P_0}\bar{u})g^h\right) \\ &\leq ch + \frac{1}{\lambda}\int_{\Omega}g(\bar{u} - \Pi_{P_0}\bar{u}) \\ &\leq ch + \frac{1}{\lambda}\|g\|_{L^\infty}\|\bar{u} - \Pi_{P_0}\bar{u}\|_{L^1} \\ &\leq c'h\end{aligned}$$

where in the final inequality we used the following lemma:

Lemma 7. *For any bounded domain $\omega \subset \mathbb{R}^2$, there exists a constant $c > 0$ such that*

$$\forall f \in BV(\omega), \|f - \Pi_{P_0}f\|_{L^1(\omega)} \leq chTV(f)$$

This lemma also applies (with now $\omega = \partial\Omega$) to get the same estimate in the Dirichlet setting as $b \in BV(\partial\Omega)$. We finally showed that, under the hypotheses $g \in L^\infty(\Omega)$ and $\bar{\phi}$ is Lipschitz, there exists a constant $c > 0$ such that $\|\bar{u}^h - \Pi_{P_0}\bar{u}\|_{L^2} \leq c\sqrt{h}$. However, to estimate the convergence of the energies $|\bar{E} - \bar{E}^h|$ it seems mandatory to control the term $\|g - g^h\|_{L^2}^2$ through lemma 7 and ask that $g \in BV(\Omega)$. In this situation, we finally get the announced convergence rate: for some $c > 0$ depending on g and on the continuous energy \bar{E} , provided $\bar{\phi}$ is Lipschitz,

$$\forall h > 0, |\bar{E} - \bar{E}^h| \leq ch$$

Note finally that the same rates would be obtained with the weaker TV diminishing lemma: $\text{TV}^h(\Pi_{P_0}(u)) \leq (1 + ch)\text{TV}(u)$ which could be true for other discrete total variations.

Remark. One could also chose to consider the discrete problem where g^h is replaced by g in the L^2 term, that is to minimize $\widetilde{E}^h(u^h) = \frac{1}{2\lambda}\|u^h - g\|_{L^2}^2 + \text{TV}_{RT}^h(u^h)$. Actually this leads to the same optimizer \bar{u}^h as $\widetilde{E}^h = E^h + \frac{1}{2\lambda}\|g - g^h\|_{L^2}^2$. However, denoting $\overline{\widetilde{E}^h}$ the optimal value of this energy, Proposition 7 then writes $\overline{E} - \overline{\widetilde{E}^h} \leq ch$. Meanwhile, after using the already mentioned calculation

$$\|\bar{u} - g\|_{L^2}^2 - \|\Pi_{P_0}(\bar{u} - g)\|_{L^2}^2 = \|\bar{u} - \Pi_{P_0}\bar{u}\|_{L^2}^2 + \|g - g^h\|_{L^2}^2 - 2 \int_{\Omega} g(\bar{u} - \Pi_{P_0}\bar{u})$$

one sees that Proposition 6 implies $\overline{E} - \overline{\widetilde{E}^h} \geq -\frac{1}{\lambda} \int_{\Omega} g(\bar{u} - \Pi_{P_0}\bar{u}) \geq -ch$ so that finally, even when $g \notin BV$ (but still assuming there exists a Lipschitz dual field $\bar{\phi}$), one has

$$|\overline{E} - \overline{\widetilde{E}^h}| \leq ch$$

6 Implementation and results

6.1 A united framework

As we have seen, the Raviart-Thomas fields offer a united framework to deal with different total variations. Indeed, TV_i^h , TV_{RT}^h as well as the total variation proposed in [9, 11] (that we will refer to as ‘‘Condat TV’’, referring to the implementation in [9]) can all be expressed in the form:

$$\begin{aligned} \text{TV}_N^L(u^h) &= \sup \left\{ - \int_{\Omega} u^h \operatorname{div} \phi, \phi \in RT0_0 \text{ s.t. } \| |L\phi| \|_{\infty} \leq 1 \right\} \\ \text{TV}_D^L(u^h) &= \sup \left\{ - \int_{\Omega} u^h \operatorname{div} \phi + \int_{\partial\Omega} b^h \langle \phi | \vec{n} \rangle, \phi \in RT0 \text{ s.t. } \| |L\phi| \|_{\infty} \leq 1 \right\} \end{aligned}$$

where $L : RT0 \rightarrow (\mathbb{R}^2)^{\mathcal{T}}$ is some linear operator giving the constraints that the dual field must satisfy, namely that $\forall i, j, |(L\phi)_{i,j}| \leq 1$.

In the case of the isotropic total variation, one has $L = L_1$, for Raviart-Thomas $L = (L_1, L_2, L_3, L_4)$, and for Condat $L = (L_{\bullet}, L_{\leftrightarrow}, L_{\leftrightarrow})$ where all these operators are

given $\forall 0 \leq i, j \leq N$ by:

$$\begin{aligned}
(L_1(\phi_f))_{i,j} &= \begin{pmatrix} f_{i+\frac{1}{2},j} \\ f_{i,j+\frac{1}{2}} \end{pmatrix} & (L_2(\phi_f))_{i,j} &= \begin{pmatrix} f_{i-\frac{1}{2},j} \\ f_{i,j+\frac{1}{2}} \end{pmatrix} \\
(L_3(\phi_f))_{i,j} &= \begin{pmatrix} f_{i-\frac{1}{2},j} \\ f_{i,j-\frac{1}{2}} \end{pmatrix} & (L_4(\phi_f))_{i,j} &= \begin{pmatrix} f_{i+\frac{1}{2},j} \\ f_{i,j-\frac{1}{2}} \end{pmatrix} \\
(L_\bullet \phi_f)_{i,j} &= \frac{1}{2} \begin{pmatrix} f_{i+\frac{1}{2},j} + f_{i-\frac{1}{2},j} \\ f_{i,j+\frac{1}{2}} + f_{i,j-\frac{1}{2}} \end{pmatrix} \\
(L_{\leftrightarrow} \phi_f)_{i,j} &= \begin{pmatrix} \frac{1}{4}(f_{i+\frac{1}{2},j} + f_{i-\frac{1}{2},j} + f_{i+\frac{1}{2},j+1} + f_{i-\frac{1}{2},j+1}) \\ f_{i,j+\frac{1}{2}} \end{pmatrix} \\
(L_{\updownarrow} \phi_f)_{i,j} &= \begin{pmatrix} f_{i+\frac{1}{2},j} \\ \frac{1}{4}(f_{i,j+\frac{1}{2}} + f_{i,j-\frac{1}{2}} + f_{i+1,j+\frac{1}{2}} + f_{i+1,j-\frac{1}{2}}) \end{pmatrix}
\end{aligned}$$

with $f_{k,l} = 0$ for couples (k, l) such that this quantity is not defined.

Note that the four variants of the isotropic total variation (obtained through the four combinations of directions selected to discretize the ∇ operator) correspond to enforcing the constraints $\| |L_k(\phi_f)| \|_\infty \leq 1$ for $1 \leq k \leq 4$ separately. On the contrary, the Raviart-Thomas total variation enforces the four of them simultaneously.

6.2 Resolution by a primal-dual algorithm

We write the (dual) ROF problem in the following way, for instance for Neumann boundary conditions:

$$\begin{aligned}
& \min_{u^h \in P_0} \frac{1}{2\lambda} \|u^h - g^h\|_{L^2}^2 + \sup \left\{ - \int_{\Omega} u^h \operatorname{div} \phi_f, \phi_f \in RT_0 \text{ s.t. } \| |L\phi_f| \|_\infty \leq 1 \right\} \\
&= \sup_{\phi_f \in RT_0} \min_{u^h \in P_0} \frac{1}{2\lambda} \|u^h - g^h\|_{L^2}^2 - \int_{\Omega} u^h \operatorname{div} \phi_f - F(L\phi_f) \\
&= - \min_{\phi_f \in RT_0} G(\phi_f) + F(L\phi_f)
\end{aligned}$$

where $G(\phi_f) = \frac{\lambda}{2} \|\operatorname{div} \phi_f\|_{L^2}^2 + \int_{\Omega} g^h \operatorname{div} \phi_f$ and $F : (\mathbb{R}^2)^{\mathcal{I}} \rightarrow \mathbb{R}$ is given by $F(z) = 0$ if $\| |z| \|_\infty \leq 1$, $+\infty$ otherwise. Note that the optimal primal solution will be obtained from the optimal $\bar{\phi}_f$ through $\bar{u}^h = g^h + \lambda \operatorname{div} \bar{\phi}_f$.

This allows one to use one of the primal-dual algorithm presented in [6] for which one needs to calculate the following proximal operators (we denote F^* the convex conjugate

of F and use the Moreau identity to calculate its prox, see [3]):

$$(Id + \tau \partial G)^{-1}(\phi_f) = \left(\frac{1}{\tau} Id + \lambda DD^* \right)^{-1} \left(\frac{1}{\tau} \phi_f + Dg^h \right)$$

$$(Id + \sigma \partial F^*)^{-1}(z) = \left\{ \begin{array}{l} 0 \text{ if } |z_i| \leq \sigma \\ z_i \left(1 - \frac{\sigma}{|z_i|}\right) \text{ otherwise} \end{array} \right\}_{i \in \mathcal{I}}$$

where $D = -\text{div}^*$ is the opposite of the dual operator of the divergence on the $RT0$ fields, which corresponds to a finite difference approximation of the gradient. Finally, we use the simplest version of the proposed algorithm and obtain the following:

Algorithm. From $\phi_f^0 \in RT0_0$, $z^0 \in (\mathbb{R}^2)^{\mathcal{I}}$, and $\sigma, \tau > 0$ such that $\sigma\tau \|L\|^2 \leq 1$, set $\bar{\phi}_f^0 = \phi_f^0$ and do $\forall n \geq 0$:

$$z^{n+1} = (Id + \sigma \partial F^*)^{-1}(z^n + \sigma L^* \bar{\phi}_f^n)$$

$$\phi_f^{n+1} = \left(\frac{1}{\tau} Id + \lambda DD^* \right)^{-1} \left(\frac{1}{\tau} \phi_f^n - L^* z^{n+1} + Dg^h \right)$$

$$\bar{\phi}_f^{n+1} = 2\phi_f^{n+1} - \phi_f^n$$

One checks that in the Dirichlet setting, the function G is replaced by $G(\phi_f) = \frac{\lambda}{2} \|\text{div} \phi_f\|_{L^2}^2 + \int_{\Omega} g^h \text{div} \phi_f - \int_{\partial\Omega} b^h \langle \phi_f | \vec{n} \rangle$ and that the same algorithm applies just replacing Dg^h with the appropriate correction to take into account the boundary term (namely in Neumann $Dg^h = 0$ on the boundary edges while in Dirichlet Dg^h has value Dg_b^h such that $\int_{\partial\Omega} \langle \phi_f | Dg_b^h \rangle = \int_{\partial\Omega} b^h \langle \phi_f | \vec{n} \rangle$).

6.3 Numerical results

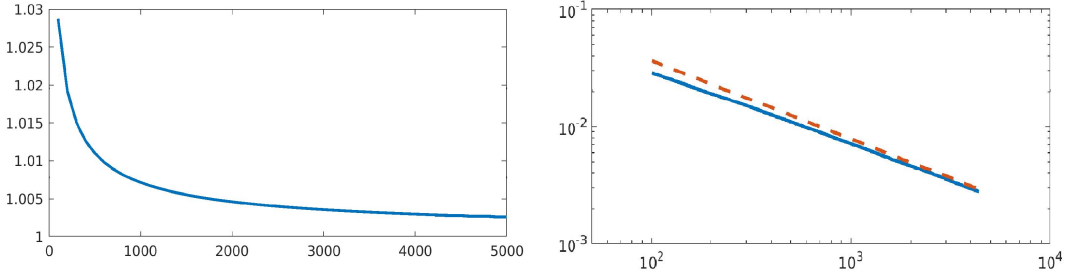


Figure 5: \bar{E}^h (plain) and $h^{2/3}$ (dashed) as functions of $N = \frac{1}{h}$ (Cartesian and log-log).

Numerical optimization of the 1D problem 7 reveals that the $O(h^{2/3})$ rate is almost observed in practice. In Figure 5 we plotted the value of the energy \bar{E}^h in the Dirichlet setting for N ranging in $[100, 5000]$ with a stepsize of 100. The corresponding log-log graph exhibits an empirical convergence rate of h^θ with $\theta = 0.6240$.

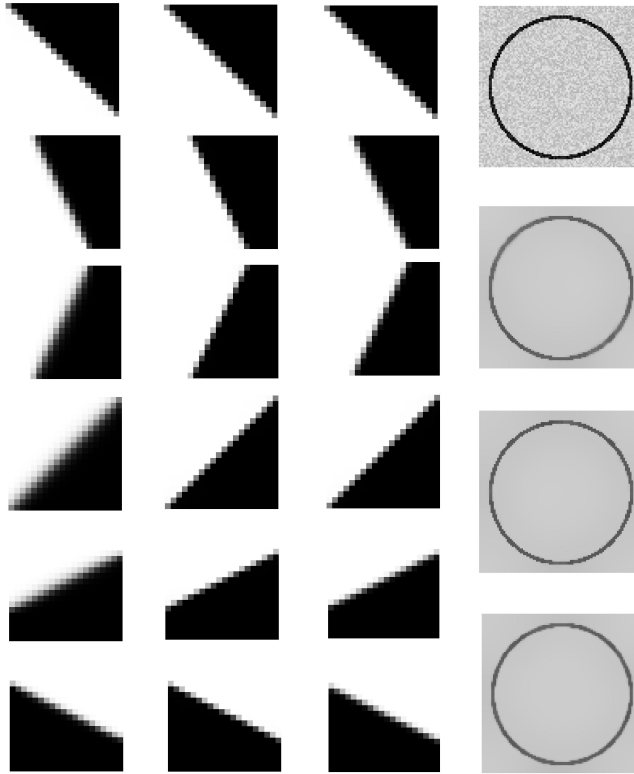


Figure 6: Denoising lines and a circle with isotropic (left col. and 2nd circle), RT (middle col. and 3rd circle) and Condat (right col. and 4th circle) total variations.

We present in Figure 6 the results for the denoising of a line, that is $g = g_\nu$ in the Dirichlet setting for different orientations ν and for the three total variations we considered: isotropic, Raviart-Thomas and “Condat”. We give also the results for the denoising of the circle we showed in the introduction, this time in the Neumann setting.

We see that the Raviart-Thomas TV performs as well as the TV of [9, 11]. However, it is important to notice that this good behavior relies heavily on the presence of the L^2 term $\|u - g\|_2^2$ in the problem we considered. Indeed, when tackling the inpainting problem, that is the completion of a missing image (here, a plain discontinuity) from its boundary datum:

$$\begin{aligned} \arg \min_{\substack{u^h \in P0 \\ \text{s.t. } u^h|_B = g|_B}} \text{TV}_D^h(u^h) \end{aligned} \quad (31)$$

where B denotes the $4N - 4$ border pixels of our image, the Raviart-Thomas TV does *worse* than the isotropic TV, while the “Condat” TV still produces sharp discontinuities, see Figure 7.

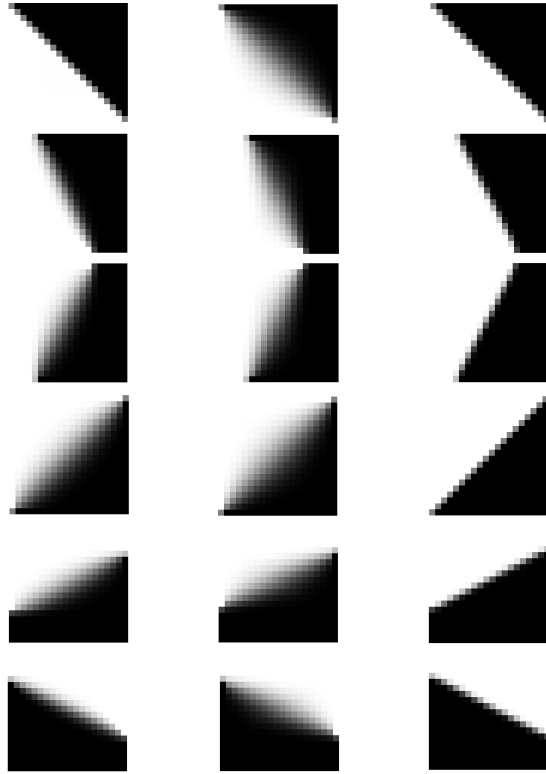


Figure 7: Inpainting with isotropic (left), RT (mid.) and Condat (right) total variations.

7 Conclusion and perspectives

In this article we developed a study of the convergence rate of the discrete towards the continuous energies of the ROF model for two discretizations of the total variation. These two discrete TV, as well as the one introduced in [9, 11] can be united under the framework of constrained Raviart-Thomas fields. Future works include estimations on convergence of the minimizers \bar{u}^h towards \bar{u} , investigations on convergence rates for the inpainting problem (31), for “Condat” TV and for other directions in the isotropic TV.

References

- [1] Sören Bartels, Ricardo H. Nochetto, and Abner J. Salgado. A total variation diminishing interpolation operator and applications. *Math. Comp.*, 84(296):2569–2587, 2015.
- [2] Sören Bartels. Nonconforming discretizations of convex minimization problems and precise relations to mixed methods, 2020. Preprint arXiv:2002.02359.
- [3] H. H. Bauschke and P. L. Combettes. *Convex analysis and monotone operator theory in Hilbert spaces*. CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC. Springer, New York, 2011. With a foreword by Hédÿ Attouch.

- [4] Antonin Chambolle, Vicent Caselles, Daniel Cremers, Matteo Novaga, and Thomas Pock. An introduction to total variation for image analysis. In *Theoretical foundations and numerical methods for sparse recovery*, volume 9 of *Radon Ser. Comput. Appl. Math.*, pages 263–340. Walter de Gruyter, Berlin, 2010.
- [5] Antonin Chambolle, Vicent Caselles, Daniel Cremers, Matteo Novaga, and Thomas Pock. An introduction to total variation for image analysis. *Theoretical foundations and numerical methods for sparse recovery*, 9(263-340):227, 2010.
- [6] 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.
- [7] Antonin Chambolle and Thomas Pock. Crouzeix–raviart approximation of the total variation on simplicial meshes. *J. Math. Imaging Vision*, 2020. (to appear).
- [8] Antonin Chambolle, Pauline Tan, and Samuel Vaiter. Accelerated alternating descent methods for Dykstra-like problems. *J. Math. Imaging Vision*, 59(3):481–497, 2017.
- [9] Laurent Condat. Discrete total variation: new definition and minimization. *SIAM J. Imaging Sci.*, 10(3):1258–1290, 2017.
- [10] P. Destuynder, M. Jaoua, and H. Sellami. A dual algorithm for denoising and preserving edges in image processing. *J. Inverse Ill-Posed Probl.*, 15(2):149–165, 2007.
- [11] 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.
- [12] 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.
- [13] Chang-Ock Lee, Eun-Hee Park, and Jongho Park. A finite element approach for the dual Rudin-Osher-Fatemi model and its nonoverlapping domain decomposition methods. *SIAM J. Sci. Comput.*, 41(2):B205–B228, 2019.
- [14] P. A. Raviart and J. M. Thomas. A mixed finite element method for 2-nd order elliptic problems. In Ilio Galligani and Enrico Magenes, editors, *Mathematical Aspects of Finite Element Methods*, pages 292–315, Berlin, Heidelberg, 1977. Springer Berlin Heidelberg.
- [15] 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)].
- [16] 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.