Discrete Total Variation with Finite Elements and Applications to Imaging

Marc Herrmann, Roland Herzog, Stephan Schmidt, José Vidal, Gerd Wachsmuth

Discrete Total Variation with Finite Elements and Applications to Imaging[*]

Marc Herrmann[†], Roland Herzog[‡], Stephan Schmidt[†], José Vidal[‡], and Gerd Wachsmuth[§]

Abstract. The total variation (TV)-seminorm is considered for piecewise polynomial, globally discontinuous (DG) and continuous (CG) finite element functions on simplicial meshes. A novel, discrete variant (DTV) based on a nodal quadrature formula is defined. DTV has favorable properties, compared to the original TV-seminorm for finite element functions. These include a convenient dual representation in terms of the supremum over the space of Raviart–Thomas finite element functions, subject to a set of simple constraints. It can therefore be shown that a variety of algorithms for classical image reconstruction problems, including TV-L^2 and TV-L^1, can be implemented in low and higher-order finite element spaces with the same efficiency as their counterparts originally developed for images on Cartesian grids.

Key words. discrete total variation, dual problem, image reconstruction, numerical algorithms

AMS subject classifications. 94A08, 68U10, 49M29, 65K05

1. Introduction. The total-variation (TV)-seminorm $|\cdot|_{TV}$ is ubiquitous as a regularizing functional in image analysis and related applications; see for instance [43, 25, 15, 13]. When $\Omega \subset \mathbb{R}^2$ is a bounded domain, this seminorm is defined as

$$(1.1) \qquad |u|_{TV(\Omega)} := \sup \left\{ \int_\Omega u \operatorname{div} \boldsymbol{p} \, \mathrm{d}x : \boldsymbol{p} \in C_c^\infty(\Omega; \mathbb{R}^2), \ |\boldsymbol{p}|_{s^*} \leq 1 \right\},$$

where $s \in [1, \infty]$, $s^* = \frac{s}{s-1}$ denotes the conjugate of s and $|\cdot|_{s^*}$ is the usual s^*-norm of vectors in $\mathbb{R}^2$. Frequent choices include $s = 2$ (the isotropic case) and $s = 1$, see Figure 1.1.

It has been observed in [19] that "the rigorous definition of the TV for discrete images has received little attention." In this paper we propose and analyze a discrete analog of (1.1) for

[†]Julius-Maximilians-Universität Würzburg, Faculty of Mathematics and Computer Science, Lehrstuhl für Mathematik VI, Emil-Fischer-Straße 40, 97074 Würzburg, Germany (MH: marc.herrmann@mathematik.uni-wuerzburg.de; STS: stephan.schmidt@mathematik.uni-wuerzburg.de, https://www.mathematik.uni-wuerzburg.de/~schmidt).

[‡]Technische Universität Chemnitz, Faculty of Mathematics, Professorship Numerical Mathematics (Partial Differential Equations), 09107 Chemnitz, Germany (RH: roland.herzog@mathematik.tu-chemnitz.de, https://www.tu-chemnitz.de/herzog, JV: jose.vidal-nunez@mathematik.tu-chemnitz.de, https://www.tu-chemnitz.de/mathematik/part_dgl/people/vidal).

[§]Brandenburgische Technische Universität Cottbus-Senftenberg, Institute of Mathematics, Chair of Optimal Control, Platz der Deutschen Einheit 1, 03046 Cottbus, Germany (gerd.wachsmuth@b-tu.de, https://www.b-tu.de/fg-optimale-steuerung).
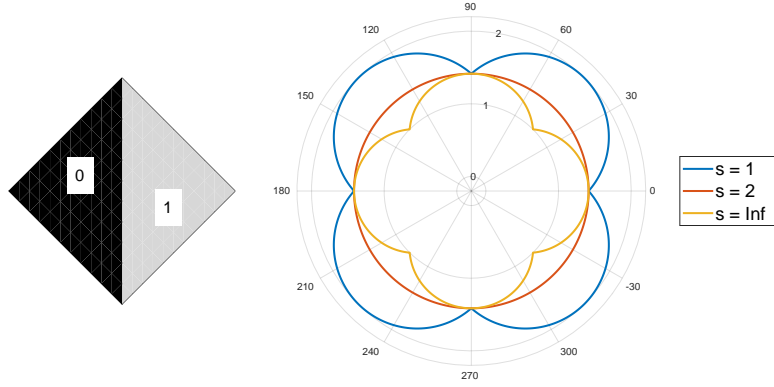
Figure 1.1. *A $\mathcal{DG}_0(\Omega)$ function u with values 0 and 1 on two triangles forming the unit square Ω (left), and the value of the associated TV-seminorm $|u|_{TV(\Omega)} = |u|_{DTV(\Omega)}$ as a function of the rotation angle of the mesh.*

functions u belonging to a space $\mathcal{DG}_r(\Omega)$ or $\mathcal{CG}_r(\Omega)$ of globally discontinuous or continuous finite element functions of polynomial degree[1] $0 \leq r \leq 4$ on a geometrically conforming, simplicial triangulation of Ω, consisting of triangles T and interior edges E.

In this case, it is not hard to see that the TV-seminorm (1.1) can be evaluated as

$$(1.2) \qquad |u|_{TV(\Omega)} = \sum_T \int_T |\nabla u|_s \, dx + \sum_E \int_E |[\![u]\!]|_s \, dS,$$

where $[\![u]\!]$ denotes the vector-valued jump of a function in normal direction across an interior edge of the triangulation.

It is intuitively clear that when u is confined to a finite element space such as $\mathcal{DG}_r(\Omega)$ or $\mathcal{CG}_r(\Omega)$, then it ought to be sufficient to consider the supremum in (1.1) over all vector fields $\boldsymbol{p}$ from an appropriate finite dimensional space as well. Indeed, we show that this is the case, provided that the TV-seminorm (1.2) is replaced by its discrete analog

$$(1.3) \qquad |u|_{DTV(\Omega)} := \sum_T \int_T \mathcal{I}_T\{|\nabla u|_s\} \, dx + \sum_E \int_E \mathcal{I}_E\{|[\![u]\!]|_s\} \, dS,$$

which we term the *discrete TV-seminorm*. Here $\mathcal{I}_T$ and $\mathcal{I}_E$ are local interpolation operators into the polynomial spaces $\mathcal{P}_{r-1}(T)$ and $\mathcal{P}_r(E)$, respectively. Therefore, (1.3) amounts to the application of a nodal quadrature formula for the integrals appearing in (1.2). We emphasize that both (1.2) and (1.3) are isotropic when $s = 2$, i.e., invariant w.r.t. rotations of the coordinate system. In the lowest-order case ($r = 0$) of piecewise constant functions, the first sum in (1.3) is zero and only edge contributions appear. Moreover, in this case (1.2) and (1.3) coincide since $[\![u]\!]$ is constant on edges. In general, we will show that the difference between (1.2) and (1.3) is of the order of the mesh size, see Proposition 3.4.

---

[1]It will become clear in section 3 why the discussion is restricted to polynomial degrees at most 4. Although this should be sufficient for most practical purposes, we briefly discuss extensions in section 10.

Using (1.3) in place of (1.2) in optimization problems in imaging offers a number of significant advantages. Specifically, we will show in Theorem 3.2 that (1.3) has a discrete dual representation

$$(1.4) \quad |u|_{DTV(\Omega)} = \max \left\{ \int_\Omega u \operatorname{div} \boldsymbol{p} \, dx : \boldsymbol{p} \in \mathcal{RT}_{r+1}^0(\Omega) \text{ s.t. a number of simple constraints} \right\}$$

for $u \in \mathcal{DG}_r(\Omega)$, where $\mathcal{RT}_{r+1}(\Omega)$ denotes the space of Raviart–Thomas finite element functions of order $r+1$, and $\mathcal{RT}_{r+1}^0(\Omega)$ is the subspace defined by $\boldsymbol{p} \cdot \boldsymbol{n} = 0$ (where $\boldsymbol{n}$ is the outer normal of unit Euclidean length) on the boundary of Ω. In the lowest-order case $r = 0$ in particular, one obtains
(1.5)

$$|u|_{DTV(\Omega)} = \max \left\{ \int_\Omega u \operatorname{div} \boldsymbol{p} \, dx : \boldsymbol{p} \in \mathcal{RT}_1^0(\Omega), \int_E |\boldsymbol{p} \cdot \boldsymbol{n}_E| \, dS \le |E| \, |\boldsymbol{n}_E|_s \text{ on interior edges} \right\}.$$

Here $\boldsymbol{n}_E$ denotes a normal vector of arbitrary orientation and unit Euclidean length, i.e., $|\boldsymbol{n}_E|_2 = 1$, on an interior edge E, and $|E|$ denotes the (Euclidean) edge length. Since the expressions $\int_E |\boldsymbol{p} \cdot \boldsymbol{n}_E| \, dS$ are exactly the degrees of freedom typically used to define the basis in $\mathcal{RT}_1(\Omega)$, the constraints in (1.5) are in fact simple *bound constraints on the coefficient vector* of $\boldsymbol{p}$. For comparison, the pointwise restrictions $|\boldsymbol{p}|_{s^*} \le 1$ appearing in (1.1) are nonlinear unless $s^* \in \{1, \infty\}$. For the case of higher-order finite elements, i.e., $1 \le r \le 4$, further constraints in (1.4) impose an upper bound on the $|\cdot|_{s^*}$-norm of *pairs of coefficients* of $\boldsymbol{p}$, see Theorem 3.2. Consequently, these constraints are likewise linear in the important special case $s = 1$. In any case, each coefficient of $\boldsymbol{p}$ is constrained only once.

As a consequence of (1.4), we establish that optimization problems utilizing the discrete TV-seminorm (1.3) as a regularizer possess a discrete dual problem with very simple constraints. This applies, in particular, to the famous TV-L^2 and TV-L^1 models; see [43] and [38, 25, 15], respectively. The structure of the primal and dual problems is in turn essential for the efficient implementation of appropriate solution algorithms. As one of the main contributions of this paper, we are able to show that a variety of popular algorithms for TV-L^2 and TV-L^1, originally developed in the context of finite difference discretizations on Cartesian grids, apply with little or no changes to discretizations with low or higher-order finite elements. Specifically, we consider the split Bregman algorithm [29], the primal-dual method of [14], Chambolle's projection method [12], a primal-dual active set method similar to [32] for TV-L^2, as well as the primal-dual method and the ADMM of [47] for TV-L^1. A 'Huberized' version of (1.3) can also be considered with minor modifications to the algorithms.

There are multiple motivations to study finite element discretizations of the TV-seminorm. First, finite element discretizations lend themselves in imaging applications when the image data is not represented on a Cartesian grid. This happens, for instance, due to the use of acquisition equipment with, e.g., hexagonal sensor layouts [34, 18], as well as for images defined on manifolds, e.g., in geodesy [33], or medical imaging applications [37]. Second, (1.1) is popular as a regularizer in inverse coefficient problems for partial differential equations; see for instance [16, 4, 17]. In this situation, a discretization by finite elements of both the state and the unknown coefficient is often the natural choice, in particular on non-trivial geometries.

Third, finite element discretizations generalize easily to higher-order simply by increasing the polyomial degree. It is well known that higher-order discretizations can outperform mesh refinement approaches when the function to be approximated is sufficiently smooth. Finally, we anticipate that our approach can be extended to total *generalized* variation (TGV) introduced in [9] as well, although this is not the subject of the present paper.

The vast majority of all publications to date dealing with the TV-seminorm use a (lowest order) finite difference approximation of (1.1) on Cartesian grids, where the divergence is approximated by one-sided differences. We are aware of only a few contributions including [26, 23, 50, 5, 6, 1, 8, 17] using lowest-order ($r = 1$) *continuous* finite elements, i.e., $u \in \mathcal{CG}_1(\Omega)$. In this case the edge jump contributions in (1.2) and (1.3) vanish, and since $\nabla u \in \mathcal{DG}_0(\Omega)$ holds, formulas (1.2) and (1.3) coincide. Moreover, the case $u \in \mathcal{DG}_0(\Omega)$ on uniform, rectangular grids is dicussed in [46]. To the best of our knowledge, the definition of the discrete TV-seminorm (1.3) as well as role of Raviart–Thomas finite element space to establish the dual representation (1.4) are novel contributions of the present work.

This paper is structured as follows. We collect some background material on finite elements in section 2. In section 3 we establish the dual representation (1.3) of the discrete TV-seminorm (1.4). We also derive an estimate of the error between (1.3) and (1.2). We present discrete TV-L^2 and TV-L^1 models along with their duals in section 4. In section 5 we show that a variety of well known algorithms for TV-L^2 image denoising can be applied in our (possibly higher-order) finite element setting with little or no changes compared to their classical counterparts in the finite difference domain. Further implementation details in the finite element framework FEniCS are given in section 6 and numerical results for TV-L^2 are presented in section 7. In section 8 we briefly also consider two methods for the TV-L^1 case. In section 9 we comment on extensions such as Huber regularized variants of TV-L^2 and TV-L^1, as well as on the simplifications that apply when images belong to globally *continuous* finite element spaces $\mathcal{CG}_r(\Omega)$. We conclude with an outlook in section 10.

Notation. Let $\Omega \subset \mathbb{R}^2$ be a bounded domain with polygonal boundary. We denote by $L^2(\Omega)$ and $H^1(\Omega)$ the usual Lebesgue and Sobolev spaces. $H^1_0(\Omega)$ is the subspace of $H^1(\Omega)$ of functions having zero trace on the boundary $\partial\Omega$. The vector valued counterparts of these spaces as well as all vector valued functions will be written in bold-face notation. Moreover, we define

$$\boldsymbol{H}(\operatorname{div};\Omega) \coloneqq \left\{ \boldsymbol{p} \in \boldsymbol{L}^2(\Omega) : \operatorname{div} \boldsymbol{p} \in L^2(\Omega) \right\}$$

and $\boldsymbol{H}_0(\operatorname{div};\Omega)$ is the subspace of functions having zero normal trace on the boundary, i.e., $\boldsymbol{p} \cdot \boldsymbol{n} = 0$.

2. Finite Element Spaces. Suppose that Ω is triangulated by a geometrically conforming mesh (no hanging nodes) consisting of non-degenerate triangular cells T and interior edges E. Recall that on each interior edge, $\boldsymbol{n}_E$ denotes the unit normal vector (of arbitrary but fixed orientation). Throughout, $r \geq 0$ denotes the degree of certain polynomials.

Lagrangian Finite Elements. Let $\mathcal{P}_r(T)$ denote the space of scalar, bivariate polynomials on T with total maximal degree r. The dimension of $\mathcal{P}_r(T)$ is $(r+1)(r+2)/2$. Let $\{\Phi_{T,k}\}$ denote the standard nodal basis of $\mathcal{P}_r(T)$ with associated Lagrange nodes $\{X_{T,k}\}$, $k = 1, \ldots, (r+1)(r+2)/2$. In other words, each $\Phi_{T,k}$ is a function in $\mathcal{P}_r(T)$ satisfying $\Phi_{T,k}(X_{T,k'}) = \delta_{kk'}$, see Figure A.1 in Appendix A. We denote by

$$(2.1) \qquad \mathcal{DG}_r(\Omega) \coloneqq \left\{ u \in L^2(\Omega) : u|_T \in \mathcal{P}_r(T) \right\}, \quad r \geq 0,$$

$$(2.2) \qquad \mathcal{CG}_r(\Omega) \coloneqq \left\{ u \in C(\Omega) : u|_T \in \mathcal{P}_r(T) \right\}, \quad r \geq 1,$$

the standard finite element spaces of globally discontinuous (L^2-conforming) or continuous (H^1-conforming) piecewise polynomials of degree r. A function $u \in \mathcal{DG}_r(\Omega)$ or $\mathcal{CG}_r(\Omega)$, restricted to T, is represented by its coefficient vector w.r.t. the basis $\{\Phi_{T,k}\}$, which is simply given by point evaluations. We use the notation

$$u_{T,k} = \{u|_T(X_{T,k})\}$$

to denote the elements of the coefficient vector of a function $u \in \mathcal{DG}_r(\Omega)$ or $\mathcal{CG}_r(\Omega)$.

Frequently we will also work with the space $\mathcal{P}_{r-1}(T)$, whose standard nodal basis and Lagrange nodes we denote by $\{\varphi_{T,i}\}$ and $\{x_{T,i}\}$, $i = 1, \ldots, r(r+1)/2$. The interpolation operator into this space (used in the definition (1.3) of $|u|_{DTV(\Omega)}$) is defined by $\mathcal{I}_T\{v\} \coloneqq \sum_{i=1}^{r(r+1)/2} v(x_{T,i})\, \varphi_{T,i}$. Similarly, $\mathcal{P}_r(E)$ denotes the space of univariate scalar polynomials on E of maximal degree r, which has dimension $r+1$. Let $\{\varphi_{E,j}\}$ denote the standard nodal basis of $\mathcal{P}_r(E)$ with associated Lagrange nodes $\{x_{E,j}\}$, $j = 1, \ldots, r+1$, see Figure A.2 in Appendix A. The associated interpolation operator becomes $\mathcal{I}_E\{v\} \coloneqq \sum_{j=1}^{r+1} v(x_{E,j})\, \varphi_{E,j}$.

Finally, we address the definition of the jump of a $\mathcal{DG}_r(\Omega)$ function across an interior edge E connecting two cells T_1 and T_2 with their respective outer normals $\boldsymbol{n}_1$ and $\boldsymbol{n}_2 = -\boldsymbol{n}_1$ of unit length. We recall that the edge normal $\boldsymbol{n}_E$ coincides either with $\boldsymbol{n}_1$ or $\boldsymbol{n}_2$ and we distinguish between the

$$(2.3a) \qquad \text{vector-valued jump} \quad [\![u]\!] = u|_{T_1}\boldsymbol{n}_1 + u|_{T_2}\boldsymbol{n}_2$$

$$(2.3b) \qquad \text{and scalar jump} \quad [u] = [\![u]\!] \cdot \boldsymbol{n}_E.$$

Notice that the sign of $[u]$ depends on the orientation of $\boldsymbol{n}_E$, while $[\![u]\!]$ does not. For instance when $\boldsymbol{n}_E = \boldsymbol{n}_1$, then $[u] \coloneqq u|_{T_1} - u|_{T_2}$ holds. Moreover, we point out that $[\![u]\!] = [u]\,\boldsymbol{n}_E$ holds.

Raviart–Thomas Finite Elements. We denote by

$$(2.4) \qquad \mathcal{RT}_{r+1}(\Omega) \coloneqq \left\{ \boldsymbol{p} \in \boldsymbol{H}(\mathrm{div};\Omega) : \boldsymbol{p}|_T \in \mathcal{P}_r(T)^2 + \boldsymbol{x}\,\mathcal{P}_r(T) \right\}, \quad r \geq 0$$

the ($\boldsymbol{H}(\mathrm{div};\Omega)$-conforming) Raviart–Thomas finite element space of order $r+1$.[2] Moreover, $\mathcal{RT}_{r+1}^0(\Omega)$ is the subspace of functions satisfying $\boldsymbol{p} \cdot \boldsymbol{n} = 0$ along the boundary of Ω. The

---

[2] Notice that while denote the lowest-order $\mathcal{RT}$ space by $\mathcal{RT}_1$, some authors use $\mathcal{RT}_0$ for this purpose.

dimension of the polynomial space on each cell is $(r+1)(r+3)$. Notice that several choices of local bases for $\mathcal{RT}_{r+1}(T)$ are described in the literature, based on either point evaluations or integral moments as degrees of freedom (dofs). Clearly, a change of the basis does not alter the finite element space but only the representation of its members, which can be identified with their coefficient vectors w.r.t. a particular basis. For the purpose of this paper, it is convenient to work with the following global degrees of freedom of integral type for $\boldsymbol{p} \in \mathcal{RT}_{r+1}(\Omega)$; see [36, Ch. 3.4.1]:

$$(2.5a) \qquad \boldsymbol{\sigma}_{T,i}(\boldsymbol{p}) := \int_T \varphi_{T,i}\,\boldsymbol{p}\,\mathrm{d}x, \quad i = 1, \ldots, r\,(r+1)/2,$$

$$(2.5b) \qquad \sigma_{E,j}(\boldsymbol{p}) := \int_E \varphi_{E,j}\,(\boldsymbol{p} \cdot \boldsymbol{n}_E)\,\mathrm{d}S, \quad j = 1, \ldots, r+1.$$

We will refer to (2.5a) as triangle-based, or interior, dofs and to (2.5b) as edge-based dofs. Notice that while the edge-based dofs are scalar, the triangle-based dofs have values in $\mathbb{R}^2$ for notational convenience. The global basis functions for the space $\mathcal{RT}_{r+1}(\Omega)$ are denoted by $\boldsymbol{\psi}_i^T$ and $\boldsymbol{\psi}_j^E$, respectively. Notice that $\boldsymbol{\psi}_i^T$ is $\mathbb{R}^{2\times2}$-valued. As is the case for all finite element spaces, any dof applied to any of the basis functions evaluates to zero except

$$(2.6) \qquad \boldsymbol{\sigma}_{T,i}(\boldsymbol{\psi}_{i'}^T) = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)\delta_{ii'} \quad \text{and} \quad \sigma_{E,j}(\boldsymbol{\psi}_{j'}^E) = \delta_{jj'}.$$

Some basis functions of type $\boldsymbol{\psi}_j^E$ are shown in Figure A.3 in Appendix A. Let us emphasize that for any function $\boldsymbol{p} \in \mathcal{RT}_{r+1}^0(\Omega)$, the dof values (2.5) are precisely the coefficients of $\boldsymbol{p}$ w.r.t. the basis, i.e.,

$$(2.7) \qquad \boldsymbol{p} = \sum_T \sum_{i=1}^{r\,(r+1)/2} \boldsymbol{\sigma}_{T,i}(\boldsymbol{p})\,\boldsymbol{\psi}_i^T + \sum_E \sum_{j=1}^{r+1} \sigma_{E,j}(\boldsymbol{p})\,\boldsymbol{\psi}_j^E.$$

Index Conventions. In order to reduce the notational overhead, we are going to associate specific ranges for any occurence of the indices i, j and k in the sequel:

$i \in \{1, \ldots, r\,(r+1)/2\}$ as in the basis functions $\varphi_{T,i}$ of $\mathcal{P}_{r-1}(T)$ and dofs $\boldsymbol{\sigma}_{T,i}$ in $\mathcal{RT}_{r+1}(\Omega)$,

$j \in \{1, \ldots, r+1\}$ as in the basis functions $\varphi_{E,j}$ of $\mathcal{P}_r(E)$ and dofs of $\sigma_{E,j}$ in $\mathcal{RT}_{r+1}(\Omega)$,

$k \in \{1, \ldots, (r+1)(r+2)/2\}$ as in the basis functions $\Phi_{T,k}$ of $\mathcal{P}_r(T)$.

For instance, (2.7) will simply be written as $\boldsymbol{p} = \sum_{T,i} \boldsymbol{\sigma}_{T,i}(\boldsymbol{p})\,\boldsymbol{\psi}_i^T + \sum_{E,j} \sigma_{E,j}(\boldsymbol{p})\,\boldsymbol{\psi}_j^E$ in the sequel. For convenience, we summarize the notation for the degrees of freedom and basis functions needed throughout the paper in Table 2.1.

3. Properties of the Discrete Total Variation. In this section we investigate the properties of the discrete total variation seminorm

$$|u|_{DTV(\Omega)} := \sum_T \int_T \mathcal{I}_T\big\{|\nabla u|_s\big\}\,\mathrm{d}x + \sum_E \int_E \mathcal{I}_E\big\{\big|[\![u]\!]\big|_s\big\}\,\mathrm{d}S$$

FE space	local dimension	dofs	basis functions	global dimension
$\mathcal{CG}_r(\Omega)$ $(r \geq 1)$	$(r+1)(r+2)/2$	eval. in $X_{T,k}$	$\{\Phi_{T,k}\}$	$N_T\,(r-2)^+(r-1)/2$ $+\,N_E\,(r-1)^+ + N_V$
$\mathcal{DG}_r(\Omega)$	$(r+1)(r+2)/2$	eval. in $X_{T,k}$	$\{\Phi_{T,k}\}$	$N_T\,(r+1)(r+2)/2$
$\mathcal{DG}_{r-1}(\Omega)$	$r\,(r+1)/2$	eval. in $x_{T,i}$	$\{\varphi_{T,i}\}$	$N_T\,r\,(r+1)/2$
$\mathcal{DG}_r(\cup E)$	$r+1$	eval. in $x_{E,j}$	$\{\varphi_{E,j}\}$	$N_E\,(r+1)$
$\mathcal{RT}_{r+1}^0(\Omega)$	$(r+1)(r+3)$	$\boldsymbol{\sigma}_{T,i}$, see (2.5a) $\sigma_{E,j}$, see (2.5b)	$\{\boldsymbol{\psi}_i^T\}$ $\{\boldsymbol{\psi}_j^E\}$	$N_T\,r\,(r+1)$ $+\,N_E\,(r+1)$

Table 2.1
Finite element spaces, their degrees of freedom and corresponding bases. Here N_V, N_T and N_E denote the number of vertices, triangles, and interior edges in the triangular mesh. A term like $(r-a)^+$ should be understood as $\max\{r-a,0\}$.

for functions $u \in \mathcal{DG}_r(\Omega)$. Recall that $\mathcal{I}_T$ and $\mathcal{I}_E$ are local interpolation operators into the polynomial spaces $\mathcal{P}_{r-1}(T)$ and $\mathcal{P}_r(E)$, respectively. In terms of the Lagrangian bases $\{\varphi_{T,i}\}$ and $\{\varphi_{E,j}\}$ of these spaces, we have

$$(3.1a) \qquad \int_T \mathcal{I}_T\{|\nabla u|_s\}\,\mathrm{d}x = \sum_{i=1}^{r\,(r+1)/2} \big|\nabla u(x_{T,i})\big|_s\, c_{T,i},$$

$$(3.1b) \qquad \int_E \mathcal{I}_E\{|\llbracket u \rrbracket|_s\}\,\mathrm{d}S = \sum_{j=1}^{r+1} \big|\llbracket u \rrbracket(x_{E,j})\big|\,|\boldsymbol{n}_E|_s\, c_{E,j},$$

where the weights are given by

$$(3.2) \qquad c_{T,i} := \int_T \varphi_{T,i}\,\mathrm{d}x \quad \text{and} \quad c_{E,j} := \int_E \varphi_{E,j}\,\mathrm{d}S.$$

Figure 3.1 provides an illustration of the difference between the contributions $\int_E |\llbracket u \rrbracket|_s\,\mathrm{d}S$ and $\int_E \mathcal{I}_E\{|\llbracket u \rrbracket|_s\}\,\mathrm{d}S$ to $|u|_{TV(\Omega)}$ and $|u|_{DTV(\Omega)}$.

In virtue of the fact that $\nabla u|_T \in \mathcal{P}_{r-1}(T)^2$ and $\llbracket u \rrbracket \in \mathcal{P}_r(E)$, it is clear that $|\cdot|_{DTV(\Omega)}$ is indeed a seminorm on $\mathcal{DG}_r(\Omega)$, provided that all weights $c_{T,i}$ and $c_{E,j}$ are non-negative. The following lemma shows that this is the case for polynomial degrees $0 \leq r \leq 4$.

Lemma 3.1 (Lagrange basis functions with positive integrals).

(a) *Let $T \subset \mathbb{R}^2$ be a triangle and $1 \leq r \leq 4$. Then $c_{T,i} \geq 0$ holds for all $i = 1,\ldots,r\,(r+1)/2$. When $r \neq 3$, then all $c_{T,i} > 0$.*

(b) *Let $E \subset \mathbb{R}^2$ be an edge and $0 \leq r \leq 7$. Then $c_{E,j} > 0$ holds for all $j = 1,\ldots,r+1$.*
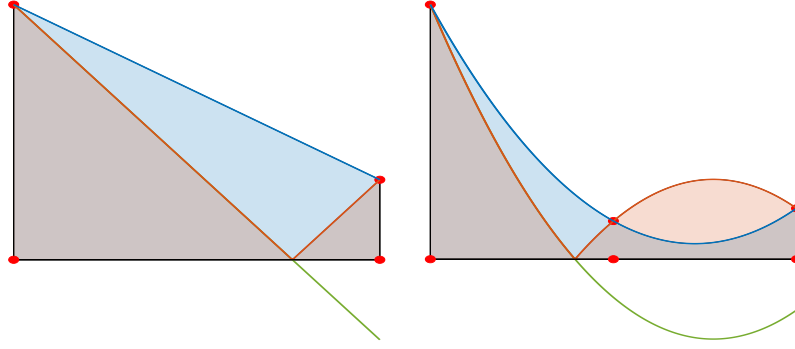
Figure 3.1. *Illustration of typical edge-jump contributions to $|u|_{TV(\Omega)}$ and to $|u|_{DTV(\Omega)}$. The green and red curves show $[\![u]\!]$ and $|[\![u]\!]|$, respectively, and the blue curve shows $\mathcal{I}_E\{|[\![u]\!]|\}$ for polynomial degrees $r = 1$ (left) and $r = 2$ (right). The picture also confirms $|u|_{TV(\Omega)} \leq |u|_{DTV(\Omega)}$ when $r = 1$, see Corollary 3.5, while $|u|_{TV(\Omega)}$ may be larger or smaller than $|u|_{DTV(\Omega)}$ when $r \in \{2, 3, 4\}$.*

Proof. Given that the Lagrange points form a unifom lattice on either T or E, the values of $c_{T,i}$ and $c_{E,j}$ are precisely the integration weights of the closed Newton–Cotes formulas. For triangles, these weights are tabulated, e.g., in [45, Tab. I] for orders $0 \leq r \leq 8$, and they confirm (a). For edges (intervals), we refer the reader to, e.g., [21, Ch. 2.5] or [20, Ch. 5.1.5], which confirms (b). ∎

We can now prove the precise form of the dual representation (1.4) of the discrete TV-seminorm (1.3).

Theorem 3.2 (Dual Representation of $|u|_{DTV(\Omega)}$). *Suppose $0 \leq r \leq 4$. Then for any $u \in \mathcal{DG}_r(\Omega)$, the discrete TV-seminorm (1.3) satisfies*

$$|u|_{DTV(\Omega)} = \sup \left\{ \int_\Omega u \operatorname{div} \boldsymbol{p} \, \mathrm{d}x : \boldsymbol{p} \in \mathcal{RT}_{r+1}^0(\Omega), \right.$$

$$|\boldsymbol{\sigma}_{T,i}(\boldsymbol{p})|_{s^*} \leq c_{T,i} \text{ for all triangles } T \text{ and all } i = 1, \ldots, r\,(r+1)/2,$$

(3.3) $$\left. |\sigma_{E,j}(\boldsymbol{p})| \leq |\boldsymbol{n}_E|_s \, c_{E,j} \text{ for all interior edges } E \text{ and all } j = 1, \ldots, r+1 \right\}.$$

Proof. We begin with the observation that integration by parts yields

(3.4) $$-\int_\Omega u \operatorname{div} \boldsymbol{p} \, \mathrm{d}x = -\sum_T \int_T u \operatorname{div} \boldsymbol{p} \, \mathrm{d}x = \sum_T \int_T \nabla u \cdot \boldsymbol{p} \, \mathrm{d}x + \sum_E \int_E [\![u]\!] \, (\boldsymbol{p} \cdot \boldsymbol{n}_E) \, \mathrm{d}S$$

for any $u \in \mathcal{DG}_r(\Omega)$ and $\boldsymbol{p} \in \mathcal{RT}_{r+1}^0(\Omega)$, i.e., $\boldsymbol{p} \cdot \boldsymbol{n} = 0$ on the boundary $\partial\Omega$.

Let us consider one of the edge integrals first. Notice that $[\![u]\!] \in \mathcal{P}_r(E)$ holds and thus $[\![u]\!] = \sum_j v_j \, \varphi_{E,j}$ with coeffients $v_j = [\![u]\!](x_{E,j})$. By the duality property (2.6) of the basis of $\mathcal{RT}_{r+1}(\Omega)$, we obtain

$$\int_E [\![u]\!] \, (\boldsymbol{p} \cdot \boldsymbol{n}_E) \, \mathrm{d}S = \sum_j v_j \int_E \varphi_{E,j} \, (\boldsymbol{p} \cdot \boldsymbol{n}_E) \, \mathrm{d}S = \sum_j v_j \, \sigma_{E,j}(\boldsymbol{p}).$$

The maximum of this expression w.r.t. $\boldsymbol{p}$ verifying the constraints in (3.3) is attained when $\sigma_{E,j}(\boldsymbol{p}) = \operatorname{sgn}(v_j)\,|\boldsymbol{n}_E|_s\,c_{E,j}$ holds. Here we are using the fact that $c_{E,j} > 0$ holds; see Lemma 3.1. Choosing $\boldsymbol{p}$ as the maximizer yields

$$\int_E [\![u]\!]\,(\boldsymbol{p}\cdot\boldsymbol{n}_E)\,\mathrm{d}S = \sum_j |v_j|\,|\boldsymbol{n}_E|_s\,c_{E,j} = \sum_j \int_E |v_j|\,\varphi_{E,j}\,|\boldsymbol{n}_E|_s\,\mathrm{d}S = \int_E \mathcal{I}_E\big\{\big|[\![u]\!]\big|_s\big\}\,\mathrm{d}S,$$

where we used $|v_j| = \big|[\![u]\!](x_{E,j})\big| = \big|[\![u]\!]\big|(x_{E,j})$ and thus $|v_j|\,|\boldsymbol{n}_E|_s = \big|[\![u]\!]\big|_s(x_{E,j})$ in the last step.

Next we consider an integral over a triangle, which is relevant only when $r \geq 1$. Since $u \in \mathcal{P}_r(T)$ holds, we have $\nabla u \in \mathcal{P}_{r-1}(T)^2$ and thus $\nabla u = \sum_i \varphi_{T,i}\,\boldsymbol{w}_i$ with vector-valued coefficients $\boldsymbol{w}_i = \nabla u(x_{T,i})$. Using again the duality property (2.6) of the basis of $\mathcal{RT}_{r+1}(\Omega)$, we obtain

$$\int_T \nabla u\cdot\boldsymbol{p}\,\mathrm{d}x = \sum_i \boldsymbol{w}_i\cdot\int_T \varphi_{T,i}\,\boldsymbol{p}\,\mathrm{d}x = \sum_i \boldsymbol{w}_i\cdot\boldsymbol{\sigma}_{T,i}(\boldsymbol{p}).$$

By virtue of Hölder's inequality, the maximum of this expression w.r.t. $\boldsymbol{p}$ verifying the constraints in (3.3) can be characterized explicitly. When $\boldsymbol{w}_i \neq \boldsymbol{0}$ and $1 \leq s < \infty$, then the maximum is attained when

$$\boldsymbol{\sigma}_{T,i}(\boldsymbol{p}) = \begin{pmatrix} (\operatorname{sgn}\boldsymbol{w}_{i,1})\,|\boldsymbol{w}_{i,1}|^{s-1} \\ (\operatorname{sgn}\boldsymbol{w}_{i,2})\,|\boldsymbol{w}_{i,2}|^{s-1} \end{pmatrix}\frac{c_{T,i}}{|\boldsymbol{w}_i|_s^{s-1}}.$$

Similarly, in case $\boldsymbol{w}_i \neq \boldsymbol{0}$ and $s = \infty$, we choose

$$\boldsymbol{\sigma}_{T,i}(\boldsymbol{p}) = \begin{cases} c_{T,i}\,(\operatorname{sgn}\boldsymbol{w}_{i,\ell}) & \text{for exactly one component } \ell \in \{1,2\} \text{ realizing } |\boldsymbol{w}_{i,\ell}| = |\boldsymbol{w}_i|_\infty, \\ 0 & \text{otherwise.} \end{cases}$$

When $\boldsymbol{w}_i = \boldsymbol{0}$ holds, $\boldsymbol{\sigma}_{T,i}(\boldsymbol{p})$ can be chosen arbitrarily but subject to $|\boldsymbol{\sigma}_{T,i}(\boldsymbol{p})|_{s^*} \leq c_{T,i}$. In any case, we arrive at the optimal value $\boldsymbol{w}_i\cdot\boldsymbol{\sigma}_{T,i}(\boldsymbol{p}) = c_{T,i}\,|\boldsymbol{w}_i|_s$. As before, we are using here the fact that $c_{T,i} \geq 0$ holds; see again Lemma 3.1. For an optimal $\boldsymbol{p}$, we thus have

$$\int_T \nabla u\cdot\boldsymbol{p}\,\mathrm{d}x = \sum_i |\boldsymbol{w}_i|_s\,c_{T,i} = \sum_i \int_T |\boldsymbol{w}_i|_s\,\varphi_{T,i}\,\mathrm{d}x = \int_T \mathcal{I}_T\big\{|\nabla u|_s\big\}\,\mathrm{d}x,$$

where we used $|\boldsymbol{w}_i|_s = |\nabla u(x_{T,i})|_s = |\nabla u|_s(x_{T,i})$ in the last step.

Finally, we point out that each summand in (3.4) depends on $\boldsymbol{p}$ only through the dof values $\boldsymbol{\sigma}_{T,i}(\boldsymbol{p})$ or $\sigma_{E,j}(\boldsymbol{p})$ associated with one particular triangle or edge. Consequently, the maximum of (3.4) is attained if and only if each summand attains its maximum subject to the constraints on the dof values set forth in (3.3). Since $-\boldsymbol{p}$ verifies the same constraints as $\boldsymbol{p}$, the maxima over $\pm\int_\Omega u\operatorname{div}\boldsymbol{p}\,\mathrm{d}x$ coincide and (3.3) is proved. ∎

Remark 3.3 (The lowest-order case $r = 0$). *In the lowest-order case $r = 0$, the only basis function on any interior edge E is $\varphi_{E,1} \equiv 1$ so that $c_{E,1} = |E|$ holds. Consequently, (3.3) reduces to (1.5).*

It may appear peculiar that the constraints for the edge dofs in (3.3) are scalar and linear, while the constraints for the pairwise triangle dofs $\boldsymbol{\sigma}_{T,i}(\boldsymbol{p}) \in \mathbb{R}^2$ are generally nonlinear. Notice, however, that it becomes evident in the proof of Theorem 3.2 that the edge dofs are utilized to measure the contributions in $|u|_{DTV(\Omega)}$ associated with the edge jumps of u, while the triangle dofs account for the contributions attributed to the gradient ∇u. Since the edge jumps are maximal in the direction normal to the edge, scalar dofs suffice in order to determine the unknown jump height. On the other hand, both the norm and direction of the gradient are unknown and must be recovered from integration against suitable functions $\boldsymbol{p}$. To this end, a variation of $\boldsymbol{\sigma}_{T,i}(\boldsymbol{p})$ within a two-dimensional ball (w.r.t. the $|\cdot|_{s^*}$-norm) is required, leading to constraints $|\boldsymbol{\sigma}_{T,i}(\boldsymbol{p})|_{s^*} \le c_{T,i}$ on pairs of coefficients of $\boldsymbol{p}$. Notice that those constraints appear for polynomial degrees $r \ge 1$ and they are nonlinear unless $s^* \in \{1, \infty\}$, which correspond to variants of the TV-seminorm with maximal anisotropy; compare Figure 1.1.

We conclude this section by comparing the TV-seminorm (1.2) with our discrete variant (1.3) for $\mathcal{DG}_r(\Omega)$ functions. For the purpose of the following result, let us denote by $\llbracket u \rrbracket'$ the tangential derivative (in arbitrary direction of traversal) of the scalar jump of u along an edge E. The symbol

$$|u|_{W^{2,\infty}(T)} = \max \left\{ \max_{x \in T} \{|u_{x_1 x_1}(x)|\}, \ \max_{x \in T} \{|u_{x_1 x_2}(x)|\}, \ \max_{x \in T} \{|u_{x_2 x_2}(x)|\} \right\}$$

is the $W^{2,\infty}$-seminorm of u on T. Moreover, we recall that the aspect ratio $\gamma_T = h_T/\varrho_T$ of a triangle T is the ratio between its diameter (longest edge) h_T and the diameter ϱ_T of the maximal inscribed circle; see for instance [24, Definition 1.107].

Proposition 3.4. *There is a constant $C > 0$ such that*

$$(3.5) \qquad \left| |u|_{TV(\Omega)} - |u|_{DTV(\Omega)} \right| \le C\, h \left(\max_T |u|_{W^{2,\infty}(T)} + \sum_E \left\| \llbracket u \rrbracket' \right\|_{L^1(E)} \right)$$

holds for all $u \in \mathcal{DG}_r(\Omega)$, $0 \le r \le 4$, where $h := \max_T h_T$ is the mesh size. The constant C depends only on r, s, the maximal aspect ratio $\max_T \gamma_T$ and the area $|\Omega|$.

Proof. We use (3.1) to interpret the discrete TV-seminorm as a quadrature rule applied to the TV-seminorm (1.2). An application of [24, Lem. 8.4] (using $d = 2$, $p = \infty$, $k_q = 2$, and $s = 1$ therein) yields the estimate

$$\left| \int_T |\nabla u|_s \, \mathrm{d}x - \sum_i |\nabla u(x_{T,i})|_s \, c_{T,i} \right| \le C\, h_T^3 \, \big||\nabla u|_s\big|_{W^{1,\infty}(T)}.$$

(During the proof, C denotes a generic constant which may change from instance to instance.) Summing over T and using $\sum_T h_T^2 \le C$ (depending on $|\Omega|$ and the maximal aspect ratio $\max_T \gamma_T$), we find

$$\sum_T \left| \int_T \left(|\nabla u|_s - \mathcal{I}_T \{|\nabla u|_s\} \right) \mathrm{d}x \right| = \sum_T \left| \int_T |\nabla u|_s \, \mathrm{d}x - \sum_i |\nabla u(x_{T,i})|_s \, c_{T,i} \right|$$

$$\le C\, h \, \max_T \big||\nabla u|_s\big|_{W^{1,\infty}(T)}.$$

Since $\boldsymbol{v} \mapsto |\boldsymbol{v}|_s$ is globally Lipschitz continuous, we find that

$$\max_T \big||\nabla u|_s\big|_{W^{1,\infty}(T)} \leq C \max_T |u|_{W^{2,\infty}(T)}.$$

Similarly, for each edge E, we can apply [24, Lem. 8.4] (using $d = 1$, $p = 1$, $k_q = 2$, and $s = 1$ therein); note that the proof carries over to this limit case with $p = 1$ and $d = s$. This yields the estimate

$$\left|\int_E \big|[\![u]\!]\big|\, \mathrm{d}S - \sum_j \big|[\![u]\!](x_{E,j})\big|\, c_{E,j}\right| \leq C\, h\, \big\|\,|[\![u]\!]|'\big\|_{L^1(E)}.$$

Here, $|[\![u]\!]|'$ is the tangential derivative of the absolute value of the jump of u on E. Notice that $\big\|\,|[\![u]\!]|'\big\|_{L^1(E)} = \big\|[\![u]\!]'\big\|_{L^1(E)}$ holds. Summing over E yields

$$\sum_E \left|\int_E \big|[\![u]\!]\big| - \mathcal{I}_E\{|[\![u]\!]|\}\, \mathrm{d}S\right| = \sum_E \left|\int_E \big|[\![u]\!]\big|\, \mathrm{d}S - \sum_j \big|[\![u]\!](x_{E,j})\big|\, c_{E,j}\right|$$

$$\leq C\, h \sum_E \big\|[\![u]\!]'\big\|_{L^1(E)}.$$

By using $\big|[\![u]\!]\big|_s = |[\![u]\!]|\,|\boldsymbol{n}_E|_s$ on each edge, and combining the above estimates, we obtain the announced error bound. ∎

Corollary 3.5 (Low Order Polynomial Degrees).

(a) When $r = 0$, we have $|u|_{TV(\Omega)} = |u|_{DTV(\Omega)}$ for all $u \in \mathcal{DG}_r(\Omega)$.

(b) When $r = 1$, then $|u|_{TV(\Omega)} \leq |u|_{DTV(\Omega)}$ for all $u \in \mathcal{DG}_r(\Omega)$.

Proof. In case $r = 0$, the right-hand side of the estimate in Proposition 3.4 vanishes. In case $r = 1$, ∇u is piecewise constant and the corresponding terms in (1.2) and (1.3) coincide. Moreover, for affine functions $v : E \to \mathbb{R}$ it is easy to check that

$$\int_E |v|\, \mathrm{d}S \leq \frac{1}{2}\left(\big|v(x_{E,1})\big| + \big|v(x_{E,2})\big|\right) \int_E 1\, \mathrm{d}S,$$

where $x_{E,1}$ and $x_{E,2}$ are the two end points of E. This yields the claim in case $r = 1$. ∎

We also mention that the boundary perimeter formula

$$\mathrm{Per}(E) := |\chi_E|_{TV(\Omega)} = |\chi_E|_{DTV(\Omega)} = \mathrm{length}(E)$$

holds when E is a union of triangles and thus the characteristic function χ_E belongs to $\mathcal{DG}_0(\Omega)$.

4. Discrete Dual Problems. In this section we revisit the classical image denoising problems.

(TV-L2)	Minimize $\dfrac{1}{2}\|u - f\|_{L^2(\Omega)}^2 + \beta\, |u|_{TV(\Omega)},$
(TV-L1)	Minimize $\|u - f\|_{L^1(\Omega)} + \beta\, |u|_{TV(\Omega)},$

see [43, 38, 25, 15, 13]. We introduce their discrete counterparts and establish their Fenchel duals.

4.1. The TV-L^2 Problem.
The discrete counterpart of (TV-L2) we consider is

(DTV-L2) $$\text{Minimize} \quad \frac{1}{2}\|u - f\|_{L^2(\Omega)}^2 + \beta\,|u|_{DTV(\Omega)}.$$

The reconstructed image u is sought in $\mathcal{DG}_r(\Omega)$ for some $0 \leq r \leq 4$. We can assume that the given data f belongs to $\mathcal{DG}_r(\Omega)$ as well, possibly after applying interpolation or quasi-interpolation. Notice that we use the discrete TV-seminorm as regularizer.

The majority of algorithms considered in the literature utilize either the primal or the dual formulations of the problems at hand. The *continuous* (pre-)dual problem for (TV-L2) is well known, see for instance [32]:

(TV-L2-D) $$\text{Minimize} \quad \frac{1}{2}\|\text{div}\,\boldsymbol{p} + f\|_{L^2(\Omega)}^2 \text{ s.t. } |\boldsymbol{p}|_{s^*} \leq \beta,$$

with $\boldsymbol{p} \in \boldsymbol{H}_0(\text{div}; \Omega)$.

For future reference, we denote the constraint set appearing in (3.3) by

$$\boldsymbol{P} := \Big\{\boldsymbol{p} \in \mathcal{RT}_{r+1}^0(\Omega) : |\boldsymbol{\sigma}_{T,i}(\boldsymbol{p})|_{s^*} \leq c_{T,i} \text{ for all triangles } T \text{ and all } i,$$

(4.1) $$|\sigma_{E,j}(\boldsymbol{p})| \leq |\boldsymbol{n}_E|_s\, c_{E,j} \text{ for all interior edges } E \text{ and all } j\Big\}.$$

Our first result in this section shows that the dual of the discrete problem (DTV-L2) has a very similar structure as (TV-L2-D), but with the pointwise constraints replaced by coefficient-wise constraints.

Theorem 4.1 (Discrete dual problem for (DTV-L2)). *Let $0 \leq r \leq 4$. Then the dual problem of* (DTV-L2) *is*

(DTV-L2-D) $$\textit{Minimize} \quad \frac{1}{2}\|\text{div}\,\boldsymbol{p} + f\|_{L^2(\Omega)}^2 \text{ s.t. } \boldsymbol{p} \in \beta\boldsymbol{P}.$$

Proof. We cast (DTV-L2) in the common form $F(u) + \beta\,G(\Lambda u)$. Let us define $U := \mathcal{DG}_r(\Omega)$ and $F(u) := \frac{1}{2}\|u - f\|_{L^2(\Omega)}^2$. The operator Λ represents the gradient of u, which consists of the triangle-wise contributions plus measure-valued contributions due to (normal) edge jumps. We therefore define

(4.2a) $$\Lambda : U \to Y := \prod_T \mathcal{P}_{r-1}(T)^2 \times \prod_E \mathcal{P}_r(E).$$

The components of Λu will be addressed by $(\Lambda u)_T$ and $(\Lambda u)_E$ respectively, and they are defined by

(4.2b) $$(\Lambda u)_T := \nabla u|_T \quad \text{and} \quad (\Lambda u)_E := [\![u]\!]_E.$$

Finally, the function $G : Y \to \mathbb{R}$ is defined by

$$(4.3) \qquad G(\boldsymbol{d}) := \sum_T \int_T \mathcal{I}_T \{ |\boldsymbol{d}_T|_s \} \, \mathrm{d}x + \sum_E |\boldsymbol{n}_E|_s \int_E \mathcal{I}_E \{ |d_E| \} \, \mathrm{d}S,$$

A crucial observation now is that the dual space Y^* of Y can be identified with $\mathcal{RT}_{r+1}^0(\Omega)$ when the duality product is defined as

$$(4.4) \qquad \langle \boldsymbol{p}, \boldsymbol{d} \rangle := \sum_T \int_T \boldsymbol{p} \cdot \boldsymbol{d}_T \, \mathrm{d}x + \sum_E \int_E (\boldsymbol{p} \cdot \boldsymbol{n}_E) \, d_E \, \mathrm{d}S.$$

In fact, $\mathcal{RT}_{r+1}^0(\Omega)$ has the same dimension as Y and, for any $\boldsymbol{p} \in \mathcal{RT}_{r+1}^0(\Omega)$, (4.4) clearly defines a linear functional on Y. Moreover, the mapping $\boldsymbol{p} \mapsto \langle \boldsymbol{p}, \cdot \rangle$ is injective since $\langle \boldsymbol{p}, \boldsymbol{d} \rangle = 0$ for all $\boldsymbol{d} \in Y$ implies $\boldsymbol{p} = \boldsymbol{0}$; see (2.5). With this representation of Y^* available, we can evaluate $\Lambda^* : \mathcal{RT}_{r+1}^0(\Omega) \to U$, where we identify U with its dual space using the Riesz isomorphism induced by the $L^2(\Omega)$ inner product. Consequently, Λ^* is defined by the condition $\langle \boldsymbol{p}, \Lambda u \rangle = (u, \Lambda^* \boldsymbol{p})_{L^2(\Omega)}$ for all $\boldsymbol{p} \in \mathcal{RT}_{r+1}^0(\Omega)$ and all $u \in \mathcal{DG}_r(\Omega)$. The left hand side is

$$(4.5) \qquad \begin{aligned} \langle \boldsymbol{p}, \Lambda u \rangle &= \sum_T \int_T \boldsymbol{p} \cdot \nabla u \, \mathrm{d}x + \sum_E \int_E (\boldsymbol{p} \cdot \boldsymbol{n}_E) \, [\![u]\!] \, \mathrm{d}S \\ &= \sum_T - \int_T (\operatorname{div} \boldsymbol{p}) \, u \, \mathrm{d}x + \sum_T \int_{\partial T} (\boldsymbol{p} \cdot \boldsymbol{n}_T) \, u \, \mathrm{d}S + \sum_E \int_E (\boldsymbol{p} \cdot \boldsymbol{n}_E) \, [\![u]\!] \, \mathrm{d}S \\ &= - \int_\Omega (\operatorname{div} \boldsymbol{p}) \, u \, \mathrm{d}x, \end{aligned}$$

hence $\Lambda^* = -\operatorname{div}$ holds. Here $\boldsymbol{n}_T$ denotes the outward unit normal along the triangle boundary ∂T.

The dual problem can be cast as

$$(4.6) \qquad \text{Minimize} \quad F^*(-\Lambda^* \boldsymbol{p}) + \beta \, G^*(\boldsymbol{p}/\beta).$$

It is well known that the convex conjugate of $F(u) = \frac{1}{2} \| u - f \|_{L^2(\Omega)}^2$ is $F^*(u) = \frac{1}{2} \| u + f \|_{L^2(\Omega)}^2 - \frac{1}{2} \| f \|_{L^2(\Omega)}^2$. It remains to evaluate

$$\begin{aligned} G^*(\boldsymbol{p}) &= \sup_{\boldsymbol{d} \in Y} \langle \boldsymbol{p}, \boldsymbol{d} \rangle - G(\boldsymbol{d}) \\ &= \sup_{\boldsymbol{d} \in Y} \sum_T \int_T \left[\boldsymbol{p} \cdot \boldsymbol{d}_T - \mathcal{I}_T \{ |\boldsymbol{d}_T|_s \} \right] \, \mathrm{d}x + \sum_E \int_E \left[(\boldsymbol{p} \cdot \boldsymbol{n}_E) \, d_E - \mathcal{I}_E \{ |d_E| \} |\boldsymbol{n}_E|_s \right] \, \mathrm{d}S. \end{aligned}$$

Let us consider the contribution from $d_E = \alpha \, \varphi_{E,j}$ for some $\alpha \in \mathbb{R}$ on a single interior edge E, and $\boldsymbol{d} \equiv 0$ otherwise. By (2.5b) and (3.2), this contribution is $\alpha \, \sigma_{E,j}(\boldsymbol{p}) - |\alpha| \, |\boldsymbol{n}_E|_s \, c_{E,j}$, which is bounded above if and only if $|\sigma_{E,j}(\boldsymbol{p})| \le |\boldsymbol{n}_E|_s \, c_{E,j}$. In this case, the maximum is zero. Similarly, it can be shown that the contribution from $\boldsymbol{d}_T = \left(\begin{smallmatrix} \alpha_1 \\ \alpha_2 \end{smallmatrix} \right) \varphi_{T,i}$ remains bounded above if and only if $|\boldsymbol{\sigma}_{T,i}(\boldsymbol{p})|_{s^*} \le c_{T,i}$, in which case the maximum is zero as well. This shows that $G^* = I_{\boldsymbol{P}}$ is the indicator function of the constraint set $\boldsymbol{P}$ defined in (4.1), which concludes the proof. ∎

Notice that the discrete dual problem (DTV-L2-D) features the same, very simple set of constraints which already appeared in (3.3). As is the case for (TV-L2-D), the solution of the discrete dual problem (DTV-L2-D) is not necessarily unique. However its divergence is unique due to the strong convexity of the objective in terms of $\operatorname{div} \boldsymbol{p}$.

Theorem 4.2 (Recovery of the Primal Solution in (DTV-L2)). *Suppose that $\boldsymbol{p} \in \mathcal{RT}_{r+1}^0(\Omega)$ is a solution of* (DTV-L2-D). *Then the unique solution of* (DTV-L2) *is given by*

$$(4.7) \qquad\qquad u = \operatorname{div} \boldsymbol{p} + f \in \mathcal{DG}_r(\Omega).$$

Proof. From (4.6), the pair of optimality conditions is

$$-\Lambda^* \boldsymbol{p} \in \partial F(u), \quad \Lambda u \in \beta \, \partial G^*(\boldsymbol{p}/\beta).$$

The first condition turns out to be equivalent to $F(u) + F^*(-\Lambda^* \boldsymbol{p}) - (u, -\Lambda^* \boldsymbol{p})_{L^2(\Omega)} = 0$, which can be rewritten as

$$\|u - f\|_{L^2(\Omega)}^2 + \|\operatorname{div} \boldsymbol{p} + f\|_{L^2(\Omega)}^2 - \|f\|_{L^2(\Omega)}^2 - 2\,(u,\, \operatorname{div} \boldsymbol{p})_{L^2(\Omega)} = 0.$$

Developing each summand in terms of the inner product $(\cdot, \cdot)_{L^2(\Omega)}$ and rearranging appropriately, we obtain

$$(u - f - \operatorname{div} \boldsymbol{p},\, u)_{L^2(\Omega)} + (-u + f + \operatorname{div} \boldsymbol{p},\, f)_{L^2(\Omega)} + (\operatorname{div} \boldsymbol{p} + f - u,\, \operatorname{div} \boldsymbol{p})_{L^2(\Omega)} = 0,$$

which amounts to $\|u - f - \operatorname{div} \boldsymbol{p}\|_{L^2(\Omega)}^2 = 0$, and (4.7) is proved. ∎

4.2. The TV-L^1 Problem. The continuous (pre-)dual problem associated with

$$(\text{TV-L1}) \qquad\qquad \text{Minimize} \quad \|u - f\|_{L^1(\Omega)} + \beta \, |u|_{TV(\Omega)}$$

can be shown along the lines of [32, Thm. 2.2] to be

$$(\text{TV-L1-D}) \qquad\qquad \text{Minimize} \quad \int_\Omega (\operatorname{div} \boldsymbol{p})\, f \, \mathrm{d}x \text{ s.t. } |\operatorname{div} \boldsymbol{p}| \le 1 \text{ and } |\boldsymbol{p}|_{s^*} \le \beta$$

with $\boldsymbol{p} \in \boldsymbol{H}_0(\operatorname{div}; \Omega)$.

The definition of an appropriate discrete counterpart of (TV-L1) deserves some attention. Simply replacing $|u|_{TV(\Omega)}$ by $|u|_{DTV(\Omega)}$ would yield a discrete dual problem with an infinite number of pointwise constraints $|\operatorname{div} \boldsymbol{p}| \le 1$ as in (TV-L1-D), which would render the problem intractable. We therefore advocate to consider

$$(\text{DTV-L1}) \qquad\qquad \text{Minimize} \quad \sum_T \int_T \mathcal{J}_T\{|u - f|\} \, \mathrm{d}x + \beta \, |u|_{DTV(\Omega)}$$

as an appropriate discrete version of (TV-L1) with $u \in \mathcal{DG}_r(\Omega)$. Here $\mathcal{J}_T$ denotes the interpolation operator into $\mathcal{P}_r(T)$, i.e.,

$$\mathcal{J}_T\{|u - f|\} = \sum_k |u - f|(X_{T,k})\, \Phi_{T,k}.$$

This choice of applying an interpolatory quadrature formula to the data fidelity (loss) term as well is a decisive advantage, yielding a favorable dual problem.

Theorem 4.3 (Discrete dual problem for (DTV-L1)). *Let $0 \leq r \leq 4$. Then the dual problem of (DTV-L1) is*

(DTV-L1-D) $\qquad$ *Minimize* $\quad \int_\Omega (\operatorname{div} \boldsymbol{p}) f \, \mathrm{d}x \ \ s.t. \ \left| \int_T (\operatorname{div} \boldsymbol{p}) \Phi_{T,k} \, \mathrm{d}x \right| \leq C_{T,k} \ \ and \ \boldsymbol{p} \in \beta \boldsymbol{P}.$

Proof. We proceed similarly as in the proof of Theorem 4.1. The functions G, G^* and Λ remain unchanged, and we replace F by

$$(4.8) \qquad F(u) = \sum_T \int_T \mathcal{J}_T \{|u - f|\} \, \mathrm{d}x = \sum_{T,k} |u - f|(X_{T,k}) \, C_{T,k},$$

where $C_{T,k} := \int_T \Phi_{T,k} \, \mathrm{d}x$ is non-negative due to Lemma 3.1. We identify again $U = \mathcal{DG}_r(\Omega)$ with its dual but this time not via the regular $L^2(\Omega)$ inner product but via its lumped approximation, i.e.,

$$(4.9) \qquad (u, v)_{\text{lumped}} := \sum_{T,k} u(X_{T,k}) \, v(X_{T,k}) \, C_{T,k}$$

for $u, v \in \mathcal{DG}_r(\Omega)$. Notice that this choice first of all affects the representation of $\Lambda^* : \mathcal{RT}_{r+1}^0(\Omega) \to U$. Indeed, using (4.5) it follows that $v = \Lambda^* \boldsymbol{p}$ is now defined by

$$(4.10) \qquad (u, v)_{\text{lumped}} = -\int_\Omega (\operatorname{div} \boldsymbol{p}) \, u \, \mathrm{d}x \quad \text{for all } u \in \mathcal{DG}_r(T).$$

For the particular choice $u = \Phi_{T,k}$, this yields

$$(4.11) \qquad v(X_{T,k}) = (\Lambda^* \boldsymbol{p})(X_{T,k}) = -\frac{1}{C_{T,k}} \int_T (\operatorname{div} \boldsymbol{p}) \, \Phi_{T,k} \, \mathrm{d}x$$

when $C_{T,k} > 0$. As a side remark, we mention that (4.11) means that $\Lambda^* \boldsymbol{p}$ is given locally by Carstensen's quasi-interpolant of $-\operatorname{div} \boldsymbol{p}$ into $\mathcal{P}_r(T)$; see [10]. When $C_{T,k} = 0$, then (4.10) can only be satisfied when $\int_T (\operatorname{div} \boldsymbol{p}) \, \Phi_{T,k} \, \mathrm{d}x = 0$ holds, in which case $v(X_{T,k})$ is arbitrary.

Next, since F is a weighted ℓ_1-norm, its convex conjugate can be easily seen to be

$$F^*(u) = \begin{cases} \displaystyle\sum_{T,k} u(X_{T,k}) \, f(X_{T,k}) \, C_{T,k} & \text{if } |u(X_{T,k})|_\infty \leq 1 \text{ for all } T, \, k \text{ such that } C_{T,k} > 0, \\ \infty & \text{otherwise.} \end{cases}$$

Consequently, by (4.11),

$$F^*(-\Lambda^* \boldsymbol{p}) = \sum_{T,k} \int_T (\operatorname{div} \boldsymbol{p}) \, \Phi_{T,k} \, \mathrm{d}x \, f(X_{T,k}) = \sum_T \int_T (\operatorname{div} \boldsymbol{p}) f \, \mathrm{d}x = \int_\Omega (\operatorname{div} \boldsymbol{p}) f \, \mathrm{d}x$$

holds when $\left| \int_T (\operatorname{div} \boldsymbol{p}) \Phi_{T,k} \, \mathrm{d}x \right| \leq C_{T,k}$ is satisfied, and $F^*(-\Lambda^* \boldsymbol{p}) = \infty$ otherwise. Plugging this into (4.6) concludes the proof. $\blacksquare$

Remark 4.4 (Discrete dual problem (DTV-L1-D)).

(a) *The replacement of $\|\cdot\|_{L^1(\Omega)}$ in the objective as well as of the $L^2(\Omega)$ inner product in U by lumped versions obtained by interpolatory quadrature has been successful in other contexts before; see for instance [11]. Here, it is essential in converting the otherwise infinitely many* pointwise *constraints $|\operatorname{div}\boldsymbol{p}| \leq 1$ into just finitely many constraints on $\operatorname{div}\boldsymbol{p}$.*

(b) *Notice that when $s^* \in \{1, \infty\}$ holds, then (DTV-L1-D) is a linear program.*

(c) *One may ask what would have happened if we had applied the same quadrature formula to the $L^2(\Omega)$ inner product already in (DTV-L2). It can be seen by straightforward calculations that the objective in (DTV-L2-D) would have been replaced by*

$$\frac{1}{2} \sum_{T,k} \left(\frac{1}{C_{T,k}} \int_T (\operatorname{div}\boldsymbol{p})\, \Phi_{T,k} \,\mathrm{d}x + f(X_{T,k}) \right)^2 C_{T,k}$$

with summands involving $C_{T,k} = 0$ omitted. There is, however, no structural advantage compared to (DTV-L2-D).

5. Algorithms for (DTV-L2). Our goal in this section is to show that a variety of standard algorithms developed for images on Cartesian grids, with finite difference approximations of gradient and divergence operations, are implementable with the same efficiency in our framework of higher-order finite elements on triangular meshes. We focus in this section on (DTV-L2) and come back to (DTV-L1) in section 8. Specifically, we consider in the following the split Bregman iteration [29], the primal-dual method of [14], Chambolle's projection method [12], and a primal-dual active set method similar to [32]. Since all algorithms are well known, we only focus on the main steps in each case. Let us recall that we are seeking a solution $u \in \mathcal{DG}_r$. For simplicity, we exclude the case $r = 3$ in this section, i.e., we restrict the discussion to the polynomial degrees $r \in \{0, 1, 2, 4\}$ so that all weights $c_{T,i}$ and $c_{E,j}$ are strictly positive. The case $r = 3$ can be included provided that zero weights are properly treated and we come back to this in subsection 9.2.

5.1. Split Bregman Method. The split Bregman method (also known as alternating direction method of multipliers (ADMM)) considers the primal problem (DTV-L2). It introduces an additional variable $\boldsymbol{d}$ and enforces the constraint $\boldsymbol{d} = \Lambda u = \nabla u$ by an augmented Lagrangian approach. As detailed in (4.2), $\boldsymbol{d}$ has contributions $\nabla u|_T$ per triangle, as well as contributions $[\![u]\!]_E$ per interior edge. We can thus express $\boldsymbol{d}$ through its coefficients $\{\boldsymbol{d}_{T,i}\}$ and $\{d_{E,j}\}$ w.r.t. the standard Lagrangian bases of $\mathcal{P}_{r-1}(T)^2$ and $\mathcal{P}_r(E)$,

$$(5.1) \qquad \boldsymbol{d} = \sum_i \boldsymbol{d}_{T,i}\, \varphi_{T,i} + \sum_j d_{E,j}\, \varphi_{E,j}.$$

Using (3.1) and (3.2), we rewrite the discrete total variation (1.3) in terms of $\boldsymbol{d}$ and adjoin the constraint $\boldsymbol{d} = \nabla u$ by way of an augmented Lagrangian functional,

$$(5.2) \quad \frac{1}{2}\|u - f\|^2_{L^2(\Omega)} + \beta \sum_{T,i} c_{T,i} \left|\boldsymbol{d}_{T,i}\right|_s + \beta \sum_{E,j} |\boldsymbol{n}_E|_s\, c_{E,j}\, |d_{E,j}|$$

$$+ \frac{\lambda}{2} \sum_{T,i} c_{T,i} \left|\boldsymbol{d}_{T,i} - \nabla u(x_{T,i}) - \boldsymbol{b}_{T,i}\right|^2_2 + \frac{\lambda}{2} \sum_{E,j} c_{E,j} \left|d_{E,j} - [\![u]\!](x_{E,j}) - b_{E,j}\right|^2.$$

Here $\boldsymbol{b}$ is an estimate of the Lagrange multiplier associated with the constraint $\boldsymbol{d} = \nabla u \in Y$, and $\boldsymbol{b}$ is naturally discretized in the same way as $\boldsymbol{d}$.

Remark 5.1 (Inner product on Y). So far we have not endowed the space $Y = \prod_T \mathcal{P}_{r-1}(T)^2 \times \prod_E \mathcal{P}_r(E)$ with an inner product. Since elements of Y represent (meausure-valued) gradients of $\mathcal{DG}_r(\Omega)$ functions, the natural choice would be to endow Y with a total variation norm of vector measures, which would amount to

$$\sum_T \int_T |\boldsymbol{d}_T|_s\, \mathrm{d}x + \sum_E |\boldsymbol{n}_E|_s \int_E |d_E|\, \mathrm{d}S$$

for $\boldsymbol{d} \in Y$. Clearly, this L^1-type norm is not induced by an inner product. Therefore we are using, in a sense, the nearest possible inner product, which is the one of L^2. For computational efficiency, it is crucial to consider its lumped version, which amounts to

$$(5.3) \qquad\qquad (\boldsymbol{d},\, \boldsymbol{e})_Y := \sum_{T,i} c_{T,i}\, \boldsymbol{d}_{T,i}\, \boldsymbol{e}_{T,i} + \sum_{E,j} c_{E,j}\, d_{E,j}\, e_{E,j}$$

for $\boldsymbol{d}, \boldsymbol{e} \in Y$. The associated norm is denoted as $\|\boldsymbol{d}\|^2_Y = (\boldsymbol{d},\, \boldsymbol{d})_Y$. Consequently, the augmented Lagrangian functional (5.2) can also be written more concisely as

$$(5.4) \qquad \frac{1}{2}\|u - f\|^2_{L^2(\Omega)} + \beta \sum_{T,i} c_{T,i} \left|\boldsymbol{d}_{T,i}\right|_s + \beta \sum_{E,j} |\boldsymbol{n}_E|_s\, c_{E,j}\, |d_{E,j}| + \frac{\lambda}{2}\|\boldsymbol{d} - \Lambda u - \boldsymbol{b}\|^2_Y.$$

The efficiency of the split Bregman iteration depends on the ability to efficiently minimize (5.2) independently for u, $\boldsymbol{d}$ and $\boldsymbol{b}$, respectively. Let us show that this is the case.

The Gradient Operator Λ. The gradient operator Λ evaluates the cell-wise gradient of $u \in \mathcal{DG}_r(\Omega)$ as well as the edge jump contributions, see (4.2). These are standard operations in any finite element toolbox. For computational efficiency, the matrix realizing $u(x_{T,i})$ and $u(x_{E,j})$ in terms of the coefficients of u can be stored once and for all.

Solving the u-problem. We consider the minimization of (5.2), or equivalently, of

$$(5.5) \quad \frac{1}{2}\|u - f\|^2_{L^2(\Omega)} + \frac{\lambda}{2} \sum_{T,i} c_{T,i} \left|\boldsymbol{d}_{T,i} - \nabla u(x_{T,i}) - \boldsymbol{b}_{T,i}\right|^2_2 + \frac{\lambda}{2} \sum_{E,j} c_{E,j} \left|d_{E,j} - [\![u]\!](x_{E,j}) - b_{E,j}\right|^2$$

w.r.t. $u \in \mathcal{DG}_r(\Omega)$. This problem can be interpreted as a DG finite element formulation of the elliptic partial differential equation $-\lambda \Delta u + u = f + \lambda \operatorname{div}(\boldsymbol{b} - \boldsymbol{d})$ in Ω. More precisely, it constitutes a nonsymmetric interior penalty Galerkin (NIPG) method; compare for instance [41] or [40, Ch. 2.4, 2.6]. Specialized preconditioned solvers for such systems are available, see for instance [3]. However, as proposed in [29], a (block) Gauss–Seidel method may be sufficient. It is convenient to group the unknowns of the same triangle together, which leads to local systems of size $(r + 1)(r + 2)/2$.

Solving the d-problem. The minimization of (5.2), or equivalently, of

$$(5.6) \quad \beta \sum_{T,i} c_{T,i} \left| \boldsymbol{d}_{T,i} \right|_s + \beta \sum_{E,j} |\boldsymbol{n}_E|_s \, c_{E,j} \left| d_{E,j} \right|$$

$$+ \frac{\lambda}{2} \sum_{T,i} c_{T,i} \left| \boldsymbol{d}_{T,i} - \nabla u(x_{T,i}) - \boldsymbol{b}_{T,i} \right|_2^2 + \frac{\lambda}{2} \sum_{E,j} c_{E,j} \left| d_{E,j} - [\![u]\!](x_{E,j}) - b_{E,j} \right|^2$$

decouples into the minimization of

$$(5.7\text{a}) \qquad \qquad \beta \left| \boldsymbol{d}_{T,i} \right|_s + \frac{\lambda}{2} \left| \boldsymbol{d}_{T,i} - \nabla u(x_{T,i}) - \boldsymbol{b}_{T,i} \right|_2^2 \quad \text{w.r.t. } \boldsymbol{d}_{T,i} \in \mathbb{R}^2,$$

$$(5.7\text{b}) \qquad \text{and} \quad \beta |\boldsymbol{n}_E|_s \left| d_{E,j} \right| + \frac{\lambda}{2} \left| d_{E,j} - [\![u]\!](x_{E,j}) - b_{E,j} \right|^2 \quad \text{w.r.t. } d_{E,j} \in \mathbb{R}.$$

It is well known that the scalar problem (5.7b) is solved via

$$d_{E,j} = \operatorname{shrink}\left([\![u]\!](x_{E,j}) + b_{E,j}, \frac{\beta |\boldsymbol{n}_E|_s}{\lambda} \right), \quad \text{where } \operatorname{shrink}(\xi, \gamma) := \max \left\{ |\xi| - \gamma, \, 0 \right\} \operatorname{sgn} \xi,$$

while the minimization of (5.7a) defines the (Euclidean) prox mapping of $| \cdot |_s$ and thus we have

$$\boldsymbol{d}_{T,i} = \operatorname{prox}_{\beta/\lambda | \cdot |_s} \left(\nabla u(x_{T,i}) + \boldsymbol{b}_{T,i} \right), \quad \text{where } \operatorname{prox}_{\beta/\lambda | \cdot |_s}(\boldsymbol{\xi}) = \boldsymbol{\xi} - \frac{\beta}{\lambda} \operatorname{proj}_{B_{| \cdot |_{s*}}} \left(\frac{\lambda}{\beta} \boldsymbol{\xi} \right).$$

Here $\operatorname{proj}_{B_{| \cdot |_{s*}}}$ is the Euclidean orthogonal projection onto the closed $| \cdot |_{s*}$-norm unit ball; see for instance [7, Ex. 6.47]. When $s \in \{1, 2\}$, then we have closed-form solutions of (5.7a):

$$[\boldsymbol{d}_{T,i}]_\ell = \operatorname{shrink}\left(\left[\nabla u(x_{T,i}) + \boldsymbol{b}_{T,i} \right]_\ell, \frac{\beta}{\lambda} \right) \text{ for } \ell = 1, 2 \qquad \text{when } s = 1,$$

$$\boldsymbol{d}_{T,i} = \max \left\{ \left| \nabla u(x_{T,i}) + \boldsymbol{b}_{T,i} \right|_2 - \frac{\beta}{\lambda}, 0 \right\} \frac{\nabla u(x_{T,i}) + \boldsymbol{b}_{T,i}}{\left| \nabla u(x_{T,i}) + \boldsymbol{b}_{T,i} \right|_2} \quad \text{when } s = 2.$$

When $\nabla u(x_{T,i}) + \boldsymbol{b}_{T,i} = 0$, the second formula is understood as $\boldsymbol{d}_{T,i} = 0$. Efficient approaches for $s = \infty$ are also available; see [22].

Updating b. This is simply achieved by replacing the current values for $\boldsymbol{b}_{T,i}$ and $b_{E,j}$ by $\boldsymbol{b}_{T,i} + \nabla u(x_{T,i}) - \boldsymbol{d}_{T,i}$ and $b_{E,j} + [\![u]\!](x_{E,j}) - d_{E,j}$, respectively.

The quantities $\boldsymbol{b}_{T,i}$ and $b_{E,j}$ represent discrete multipliers associated with the components of the constraint $\boldsymbol{d} = \Lambda u$. Here we clarify how these multipliers relate to the dual variable $\boldsymbol{p} \in \mathcal{RT}^0_{r+1}(\Omega)$ in (DTV-L2-D). In fact, let us interpret $\boldsymbol{b}_{T,i}$ as the coefficients of a function $\boldsymbol{b}_T \in \mathcal{P}_{r-1}(T)$ and $b_{E,j}$ as the coefficients of a function $b_E \in \mathcal{P}_r(E)$ w.r.t. the standard nodal bases, just as in (5.1). Moreover, let us define a function $\bar{\boldsymbol{p}} \in \mathcal{RT}^0_{r+1}(\Omega)$ by specifying its coefficients as follows,

$$(5.8) \qquad \boldsymbol{\sigma}_{T,i}(\bar{\boldsymbol{p}}) = \lambda \, \boldsymbol{b}_{T,i} \, c_{T,i} \quad \text{and} \quad \sigma_{E,j}(\bar{\boldsymbol{p}}) = \lambda \, b_{E,j} \, c_{E,j}.$$

Then

$$\int_T \mathcal{I}_T\big\{\bar{\boldsymbol{p}} \cdot (\nabla u - \boldsymbol{d}_T)\big\} \, \mathrm{d}x = \lambda \int_T \mathcal{I}_T\big\{\boldsymbol{b}_T \cdot (\nabla u - \boldsymbol{d}_T)\big\} \, \mathrm{d}x = \lambda \sum_i c_{T,i} \, \boldsymbol{b}_{T,i} \cdot \big(\nabla u(x_{T,i}) - \boldsymbol{d}_{T,i}\big),$$

$$\int_E \mathcal{I}_E\big\{\bar{\boldsymbol{p}} \cdot (\llbracket u \rrbracket - d_E) \, \boldsymbol{n}_E\big\} \, \mathrm{d}S = \lambda \int_E \mathcal{I}_E\big\{b_E (\llbracket u \rrbracket - d_E)\big\} \, \mathrm{d}S = \lambda \sum_j c_{E,j} \, b_{E,j} \, \big(\llbracket u \rrbracket(x_{E,j}) - d_{E,j}\big),$$

and these are precisely the terms appearing in the discrete augmented Lagrangian functional (5.2). Consequently, $\bar{\boldsymbol{p}}$ can be interpreted as the Lagrange multiplier associated with the components of the constraint $\boldsymbol{d} = \Lambda u$, when the latter are adjoined using the lumped $\boldsymbol{L}^2(T)$ and $L^2(E)$ inner products. It can be shown using similar arguments as in Theorem 3.2 and Theorem 4.2 that $\bar{\boldsymbol{p}}$ defined by (5.8) solves the dual problem (TV-L2-D).

For convenience, we specify the split Bregman iteration in Algorithm 5.1.

Algorithm 5.1 Split Bregman algorithm for (DTV-L2) with $s \in [1, \infty]$

1: Set $u^{(0)} := f \in \mathcal{DG}_r(\Omega)$, $\boldsymbol{b}^{(0)} := \boldsymbol{0} \in Y$ and $\boldsymbol{d}^{(0)} := \boldsymbol{0} \in Y$
2: Set $n := 0$
3: **while** not converged **do**
4:      Minimize (5.5) for $u^{(n+1)}$ with data $\boldsymbol{b}^{(n)}$ and $\boldsymbol{d}^{(n)}$
5:      Minimize (5.7) for $\boldsymbol{d}^{(n+1)}$ with data $u^{(n+1)}$ and $\boldsymbol{b}^{(n)}$
6:      Set $\boldsymbol{b}_{T,i}^{(n+1)} := \boldsymbol{b}_{T,i}^{(n)} + \nabla u^{(n+1)}(x_{T,i}) - \boldsymbol{d}_{T,i}^{(n+1)}$ and $b_{E,j}^{(n+1)} := b_{E,j}^{(n)} + \llbracket u^{(n+1)} \rrbracket(x_{E,j}) - d_{E,j}^{(n+1)}$
7:      Set $n := n + 1$
8: **end while**
9: Set $\boldsymbol{p}^{(n)}$ by (5.8) with data $\boldsymbol{b}^{(n)}$

5.2. Chambolle–Pock Method. The method by [14], also known as primal-dual extragradient method, see [30], is based on a reformulation of the optimality conditions in terms of the prox operators pertaining to F and G^*. We recall that F is defined by $F(u) = \frac{1}{2}\|u - f\|^2_{L^2(\Omega)}$ on $U = \mathcal{DG}_r(\Omega)$. Moreover, G^* is defined on $Y^* \cong \mathcal{RT}^0_{r+1}(\Omega)$ by $G^* = I_{\boldsymbol{P}}$, the indicator function of $\boldsymbol{P}$, see (4.1).

Notice that prox operators depend on the inner product in the respective space. We recall that U has been endowed with the (regular, non-lumped) $L^2(\Omega)$ inner product, see the proof

of Theorem 4.1. For the space Y we are using again the lumped inner product defined in (5.3). Exploiting the duality product (4.4) between Y and $Y^* \cong \mathcal{RT}^0_{r+1}(\Omega)$ it is then straightforward to derive the Riesz map $R : Y \ni \boldsymbol{d} \mapsto \boldsymbol{p} \in Y^*$. In terms of the coefficients of $\boldsymbol{p}$, we have

$$(5.9) \qquad \boldsymbol{\sigma}_{T,i}(\boldsymbol{p}) = c_{T,i}\, \boldsymbol{d}_{T,i} \quad \text{and} \quad \sigma_{E,j}(\boldsymbol{p}) = c_{E,j}\, d_{E,j}.$$

Consequently, the induced inner product in $\mathcal{RT}^0_{r+1}(\Omega)$ becomes

$$(5.10) \qquad (\boldsymbol{p},\, \boldsymbol{q})_{Y^*} := \sum_{T,i} \frac{1}{c_{T,i}}\, \boldsymbol{\sigma}_{T,i}(\boldsymbol{p}) \cdot \boldsymbol{\sigma}_{T,i}(\boldsymbol{q}) + \sum_{E,j} \frac{1}{c_{E,j}}\, \sigma_{E,j}(\boldsymbol{p})\, \sigma_{E,j}(\boldsymbol{q}).$$

To summarize, the inner products in Y, Y^* as well as the Riesz map are realized efficiently by simple, diagonal operations on the coefficients.

Solving the F-prox. Let $\sigma > 0$. The prox-operator of σF, denoted by $\mathrm{prox}_{\sigma F}(\bar{u}) : U \to U$, is defined as

$$u = \mathrm{prox}_{\sigma F}(\bar{u}) \quad \Leftrightarrow \quad u = \underset{v \in \mathcal{DG}_r(\Omega)}{\arg\min} \frac{1}{2}\|v - \bar{u}\|^2_{L^2(\Omega)} + \frac{\sigma}{2}\|v - f\|^2_{L^2(\Omega)}.$$

For given data $\bar{u}, f \in \mathcal{DG}_r(\Omega)$, it is easy to see that a necessary and sufficient condition is $u - \bar{u} + \sigma\,(u - f) = 0$, which amounts to the coefficient-wise formula

$$(5.11) \qquad u_{T,k} = \frac{1}{1 + \sigma}\big(\bar{u}_{T,k} + \sigma f_{T,k}\big).$$

Solving the G^*-prox. Let $\tau > 0$. The prox-operator $\mathrm{prox}_{\tau G^*} : Y^* \cong \mathcal{RT}^0_{r+1}(\Omega) \to Y^*$ is defined as

$$\boldsymbol{p} = \mathrm{prox}_{\tau G^*}(\bar{\boldsymbol{p}}) \quad \Leftrightarrow \quad \boldsymbol{p} = \underset{\boldsymbol{q} \in \mathcal{RT}^0_{r+1}(\Omega)}{\arg\min} \frac{1}{2}\|\boldsymbol{q} - \bar{\boldsymbol{p}}\|^2_{Y^*} \text{ s.t. } \boldsymbol{q} \in \boldsymbol{P}.$$

Similarly, the prox operator for $(\beta G)^*$ is obtained by replacing $\boldsymbol{P}$ by $\beta \boldsymbol{P}$, for any $\tau > 0$. Due to the diagonal structure of the inner product in Y^*, this is efficiently implementable. When $\bar{\boldsymbol{p}} \in \mathcal{RT}^0_{r+1}(\Omega)$, then we obtain the solution in terms of the coefficients, similar to (5.7), as
(5.12)
$$\boldsymbol{\sigma}_{T,i}(\boldsymbol{p}) = \mathrm{proj}_{\beta\, c_{T,i} B_{|\cdot|_{s^*}}}\big(\boldsymbol{\sigma}_{T,i}(\bar{\boldsymbol{p}})\big) \quad \text{and} \quad \sigma_{E,j}(\boldsymbol{p}) = \min\big\{|\sigma_{E,j}(\bar{\boldsymbol{p}})|,\ \beta\,|\boldsymbol{n}_E|_s\, c_{E,j}\big\} \frac{\sigma_{E,j}(\bar{\boldsymbol{p}})}{|\sigma_{E,j}(\bar{\boldsymbol{p}})|}.$$

In particular we have

$$\big[\boldsymbol{\sigma}_{T,i}(\boldsymbol{p})\big]_j = \min\big\{\big|[\boldsymbol{\sigma}_{T,i}(\bar{\boldsymbol{p}})]_j\big|,\ \beta\, c_{T,i}\big\}\, \mathrm{sgn}[\boldsymbol{\sigma}_{T,i}(\bar{\boldsymbol{p}})]_j \quad \text{when } s = 1,$$

$$\boldsymbol{\sigma}_{T,i}(\boldsymbol{p}) = \min\{|\boldsymbol{\sigma}_{T,i}(\bar{\boldsymbol{p}})|_2,\ \beta\, c_{T,i}\} \frac{\boldsymbol{\sigma}_{T,i}(\bar{\boldsymbol{p}})}{|\boldsymbol{\sigma}_{T,i}(\bar{\boldsymbol{p}})|_2} \quad \text{when } s = 2.$$

The second formula is understood as $\boldsymbol{\sigma}_{T,i}(\boldsymbol{p}) = 0$ when $|\boldsymbol{\sigma}_{T,i}(\bar{\boldsymbol{p}})|_2 = 0$. An implementation of the Chambolle–Pock method is given in Algorithm 5.2.

Algorithm 5.2 Chambolle–Pock algorithm for (DTV-L2) with $s \in [1, \infty]$

1: Set $u^{(0)} := f \in \mathcal{DG}_r(\Omega)$, $\boldsymbol{p}^{(0)} := \boldsymbol{0} \in \mathcal{RT}^0_{r+1}(\Omega)$ and $\bar{\boldsymbol{p}}^{(0)} := \boldsymbol{0} \in \mathcal{RT}^0_{r+1}(\Omega)$

2: Set $n := 0$

3: **while** not converged **do**

4:     Set $v^{(n+1)} := -\operatorname{div} \bar{\boldsymbol{p}}^{(n)} \in \mathcal{DG}_r(\Omega)$ $/\!\!/ \; v^{(n+1)} = \Lambda^* \bar{\boldsymbol{p}}^{(n)}$

5:     Set $u^{(n+1)} := \operatorname{prox}_{\sigma F}(u^{(n)} + \sigma v^{(n+1)})$, see (5.11) $/\!\!/ \; u^{(n+1)} = \operatorname{prox}_{\sigma F}(u^{(n)} + \sigma \Lambda^* \bar{\boldsymbol{p}}^{(n)})$

6:     Set $\boldsymbol{d}^{(n+1)} := \Lambda u^{(n+1)} \in Y$

7:     Set $\boldsymbol{q}^{(n+1)} := R\,\boldsymbol{d}^{(n+1)} \in \mathcal{RT}^0_{r+1}(\Omega)$, where R is the Riesz map (5.9)

8:     Set $\boldsymbol{p}^{(n+1)} := \operatorname{prox}_{\tau(\beta G)^*}(\boldsymbol{p}^{(n)} - \tau\,\boldsymbol{q}^{(n+1)})$, see (5.12)

                        $/\!\!/ \; \boldsymbol{p}^{(n+1)} = \operatorname{prox}_{\tau(\beta G)^*}(\boldsymbol{p}^{(n)} - \tau\,R\,\Lambda u^{(n+1)})$

9:     Set $\bar{\boldsymbol{p}}^{(n+1)} := \boldsymbol{p}^{(n+1)} + \theta\,(\boldsymbol{p}^{(n+1)} - \boldsymbol{p}^{(n)})$

10:     Set $n := n + 1$

11: **end while**

5.3. Chambolle's Projection Method. Chambolle's method was introduced in [12] and it solves (DTV-L2) via its dual (DTV-L2-D), specifically in the case $s = s^* = 2$. Squaring the constraints pertaining to $\boldsymbol{p} \in \beta \boldsymbol{P}$, we obtain the Lagrangian

$$(5.13) \quad \frac{1}{2}\|\operatorname{div} \boldsymbol{p} + f\|^2_{L^2(\Omega)} + \sum_{T,i} \frac{\alpha_{T,i}}{2}\left(|\boldsymbol{\sigma}_{T,i}(\boldsymbol{p})|^2_2 - \beta^2 c^2_{T,i}\right) + \sum_{E,j} \frac{\alpha_{E,j}}{2}\left(|\sigma_{E,j}(\boldsymbol{p})|^2 - \beta^2 c^2_{E,j}\right),$$

where $\alpha_{T,i}$ and $\alpha_{E,j}$ are Lagrange multipliers. Consequently, the KKT conditions associated with this formulation of (DTV-L2-D) are

$$(5.14) \quad (\operatorname{div} \boldsymbol{p} + f, \operatorname{div} \delta\boldsymbol{p})_{L^2(\Omega)} + \sum_{T,i} \alpha_{T,i}\,\boldsymbol{\sigma}_{T,i}(\boldsymbol{p}) \cdot \boldsymbol{\sigma}_{T,i}(\delta\boldsymbol{p}) + \sum_{E,j} \alpha_{E,j}\,\sigma_{E,j}(\boldsymbol{p})\,\sigma_{E,j}(\delta\boldsymbol{p}) = 0$$

for all $\delta\boldsymbol{p} \in \mathcal{RT}^0_{r+1}(\Omega)$, together with the complementarity conditions

$$(5.15a) \qquad 0 \le \alpha_{T,i} \quad \perp \quad |\boldsymbol{\sigma}_{T,i}(\boldsymbol{p})|_2 - \beta\,c_{T,i} \le 0 \quad \text{for all } T \text{ and } i = 1, \ldots, r\,(r+1)/2,$$

$$(5.15b) \qquad 0 \le \alpha_{E,j} \quad \perp \quad |\sigma_{E,j}(\boldsymbol{p})| - \beta\,c_{E,j} \le 0 \quad \text{for all } E \text{ and } j = 1, \ldots, r+1.$$

Let us observe that the first term in (5.14) can be written as $-\langle\Lambda(\operatorname{div} \boldsymbol{p} + f), \delta\boldsymbol{p}\rangle_{Y,Y^*}$, and hence as

$$-\sum_T \int_T \nabla u|_T \cdot \delta\boldsymbol{p} \,\mathrm{d}x - \sum_E \int_E [\![u]\!]\,(\delta\boldsymbol{p} \cdot \boldsymbol{n}_E)\,\mathrm{d}S,$$

where we set $u := \operatorname{div} \boldsymbol{p} + f$ as an abbreviation in accordance with (4.7). By selecting directions $\delta\boldsymbol{p}$ from the collections $\{\boldsymbol{\psi}^T_i\}$ and $\{\boldsymbol{\psi}^E_j\}$ of $\mathcal{RT}^0_{r+1}(\Omega)$ basis functions, see section 2, we infer that (5.14) is equivalent to

$$(5.16a) \qquad -\nabla u(x_{T,i}) + \alpha_{T,i}\,\boldsymbol{\sigma}_{T,i}(\boldsymbol{p}) = 0 \quad \text{for all } T \text{ and } i = 1, \ldots, r\,(r+1)/2,$$

$$(5.16b) \qquad -[\![u]\!](x_{E,j}) + \alpha_{E,j}\,\sigma_{E,j}(\boldsymbol{p}) = 0 \quad \text{for all } E \text{ and } j = 1, \ldots, r+1.$$

A simple calculation similar as in [12] then shows that (5.15) and (5.16) imply

$$(5.17) \qquad \beta\,\alpha_{T,i}\,c_{T,i} = |\nabla u(x_{T,i})|_2 \quad \text{and} \quad \beta\,\alpha_{E,j}\,c_{E,j} = \big|[\![u]\!](x_{E,j})\big|.$$

In order to re-derive Chambolle's algorithm for the setting at hand, it remains to rewrite the directional derivative (5.14) in terms of the gradient $\boldsymbol{g} \in Y^*$ w.r.t. the Y^* inner product (5.10). We obtain that $\boldsymbol{g}$ is given by its coefficients

$$(5.18a) \qquad \boldsymbol{\sigma}_{T,i}(\boldsymbol{g}) = c_{T,i}\,\big(\alpha_{T,i}\,\boldsymbol{\sigma}_{T,i}(\boldsymbol{p}) - \nabla u(x_{T,i})\big),$$
$$(5.18b) \qquad \sigma_{E,j}(\boldsymbol{g}) = c_{E,j}\,\big(\alpha_{E,j}\,\sigma_{E,j}(\boldsymbol{p}) - [\![u]\!](x_{E,j})\big).$$

Given an iterate for $\boldsymbol{p}$, the main steps of the algorithm are then to update the auxiliary quantity $u = \operatorname{div}\boldsymbol{p} + f$ as well as the multipliers $\alpha_{T,i}$ and $\alpha_{E,j}$ according to (5.17), and take a semi-implicit gradient step with a suitable step length to update $\boldsymbol{p}$. Since all of these steps are inexpensive, Chambolle's method can be implemented just as efficiently as its finite difference version originally given in [12]. For the purpose of comparison, we point out that one step of the method can be written compactly as

$$\boldsymbol{\sigma}_{T,i}(\boldsymbol{p}^{(n+1)}) := \frac{\boldsymbol{\sigma}_{T,i}(\boldsymbol{p}^{(n)}) + \tau\,c_{T,i}\nabla(\operatorname{div}\boldsymbol{p}^{(n)} + f)(x_{T,i})}{1 + \tau\,\beta^{-1}\big|\nabla(\operatorname{div}\boldsymbol{p}^{(n)} + f)(x_{T,i})\big|_2} \quad \text{for all } T \text{ and } i,$$

$$\sigma_{E,j}(\boldsymbol{p}^{(n+1)}) := \frac{\sigma_{E,j}(\boldsymbol{p}^{(n)}) + \tau\,c_{E,j}[\![\operatorname{div}\boldsymbol{p}^{(n)} + f]\!](x_{E,j})}{1 + \tau\,\beta^{-1}\big|[\![\operatorname{div}\boldsymbol{p}^{(n)} + f]\!](x_{E,j})\big|} \quad \text{for all } E \text{ and } j.$$

Let us mention that our variable $\boldsymbol{p}$ differs by a factor of β from the one used in [12]. Moreover, in the implementation given as Algorithm 5.3, we found it convenient to rename $\alpha_{T,i}\,c_{T,i}$ as $\gamma_{T,i}$, and similarly for the edge based quantities. Notice that $\gamma_{T,i}$ and $\gamma_{E,j}$ can be conveniently stored, for instance, as the coefficients of a $\mathcal{DG}_{r-1}(\Omega)$ function, and another $\mathcal{DG}_r$ function on the skeleton of the mesh, i.e., the union of all interior edges.

Algorithm 5.3 Chambolle's algorithm for (DTV-L2) with $s = 2$

1: Set $\boldsymbol{p}^{(0)} := \boldsymbol{0} \in \mathcal{RT}_{r+1}^0(\Omega)$
2: Set $n := 0$
3: **while** not converged **do**
4:    Set $u^{(n)} := \operatorname{div}\boldsymbol{p}^{(n)} + f \in \mathcal{DG}_r(\Omega)$
5:    Set $\gamma_{T,i} := \beta^{-1}|\nabla u^{(n)}(x_{T,i})|_2$ $/\!\!/\ \gamma_{T,i} = \alpha_{T,i}\,c_{T,i}$, see (5.17)
6:    Set $\gamma_{E,j} := \beta^{-1}|[\![u^{(n)}]\!](x_{E,j})|$ $/\!\!/\ \gamma_{E,j} = \alpha_{E,j}\,c_{E,j}$, see (5.17)
7:    Set $\boldsymbol{\sigma}_{T,i}(\boldsymbol{p}^{(n+1)}) := \dfrac{\boldsymbol{\sigma}_{T,i}(\boldsymbol{p}^{(n)}) + \tau\,c_{T,i}\nabla u^{(n)}(x_{T,i})}{1 + \tau\,\gamma_{T,i}}$
8:    Set $\sigma_{E,j}(\boldsymbol{p}^{(n+1)}) := \dfrac{\sigma_{E,j}(\boldsymbol{p}^{(n)}) + \tau\,c_{E,j}[\![u^{(n)}]\!](x_{E,j})}{1 + \tau\,\gamma_{E,j}}$
9:    Set $n := n + 1$
10: **end while**

5.4. Primal-Dual Active Set Method. We consider a primal-dual active set (PDAS) strategy for the dual problem (DTV-L2-D). A similar approach was proposed in [32], however in the context of finite difference approximation and an additional regularization of the dual problem. The PDAS method is closely related to a semi-smooth Newton approach, see [31], and it is based on the associated KKT conditions and a semi-smooth reformulation of the complementarity conditions associated with the constraints $\boldsymbol{p} \in \beta \boldsymbol{P}$. The approach is particularly suitable when $s = 1$ and thus the constraints describing $\boldsymbol{P}$ are simple bounds. We thus focus on the case $s = 1$. Then the KKT conditions associated with (DTV-L2-D) can be written as follows:

$$(5.19) \qquad (\operatorname{div} \boldsymbol{p} + f, \operatorname{div} \delta \boldsymbol{p})_{L^2(\Omega)} + \sum_{T,i} \boldsymbol{\mu}_{T,i} \cdot \boldsymbol{\sigma}_{T,i}(\delta \boldsymbol{p}) + \sum_{E,j} \mu_{E,j} \, \sigma_{E,j}(\delta \boldsymbol{p}) = 0$$

for all $\delta \boldsymbol{p} \in \mathcal{RT}_{r+1}^0(\Omega)$, together with the complementarity conditions

(5.20a)
$$\boldsymbol{\mu}_{T,i} = \max\left\{0, \boldsymbol{\mu}_{T,i} + c\left(\boldsymbol{\sigma}_{T,i}(\boldsymbol{p}) - \beta \, c_{T,i}\, \mathbf{1}\right)\right\} + \min\left\{0, \boldsymbol{\mu}_{T,i} + c\left(\boldsymbol{\sigma}_{T,i}(\boldsymbol{p}) + \beta \, c_{T,i}\, \mathbf{1}\right)\right\},$$
(5.20b)
$$\mu_{E,j} = \max\left\{0, \mu_{E,j} + c\left(\sigma_{E,j}(\boldsymbol{p}) - \beta \, |\boldsymbol{n}_E|_1 \, c_{E,j}\right)\right\} + \min\left\{0, \mu_{E,j} + c\left(\sigma_{E,j}(\boldsymbol{p}) + \beta \, |\boldsymbol{n}_E|_1 \, c_{E,j}\right)\right\},$$

where $c > 0$ is arbitrary. Notice that, as is customary for bound constrained problems, we are using signed multipliers $\boldsymbol{\mu}_{T,i}$ and $\mu_{E,j}$. Moreover, (5.20a) is understood componentwise in $\mathbb{R}^2$. The semi-smooth linearization of (5.20) agrees with a piecewise linearization on the three branches possible per expression. When we write the (non-globalized) semi-smooth Newton method in terms of the subsequent iterate, we arrive at Algorithm 5.4.

Notice that the solution of (5.21) in Algorithm 5.4 is not necessarily unique. This is not an obstacle when (5.21) is solved iteratively, e.g., by the conjugate gradient method. Alternatively, we might add the regularizing term $(\varepsilon/2)\|\boldsymbol{p}\|_{Y*}^2$ to the objective. In this case, also the multiplier update on the active sets must be replaced by

$$\left[\boldsymbol{\mu}_{T,i}^{(n+1)}\right]_{1,2} := \left[\nabla u^{(n+1)}(x_{T,i})\right]_{1,2} - \frac{\varepsilon}{c_{T,i}}\left[\boldsymbol{\sigma}_{T,i}(\boldsymbol{p}^{(n+1)})\right]_{1,2},$$

$$\mu_{E,j}^{(n+1)} := [\![u^{(n+1)}]\!](x_{E,j}) - \frac{\varepsilon}{c_{E,j}}\sigma_{E,j}(\boldsymbol{p}^{(n+1)}).$$

This modification amounts to employing a Huber regularization to $|u|_{DTV(\Omega)}$, see subsection 9.1.

6. Implementation Details. Our implementation was carried out in the finite element framework FEniCS (version 2017.2). We refer the reader to [36, 2] for background reading. FEniCS supports finite elements of various types, including $\mathcal{CG}_r$, $\mathcal{DG}_r$ and $\mathcal{RT}_{r+1}$ elements of arbitrary order. Although we focus on this piece of software, the content of this section will apply to other finite element frameworks as well.

While the bases for the spaces $\mathcal{CG}_r$ and $\mathcal{DG}_r$ in FEniCS are given by the standard nodal basis functions as described in section 2, the implementation of $\mathcal{RT}_{r+1}$ elements in FEniCS

Algorithm 5.4 Primal-dual active set method for (DTV-L2-D) with $s = 1$

1: Set $\boldsymbol{p}^{(0)} := \boldsymbol{0} \in \mathcal{R}\mathcal{T}_{r+1}^0(\Omega)$ and $\boldsymbol{\mu}^{(0)} := \boldsymbol{0}$

2: Set $n := 0$

3: **while** not converged **do**

4:    Determine the active sets

$$\mathcal{A}_T^{\pm,1} := \left\{ (T,i) : \pm[\boldsymbol{\mu}_{T,i}^{(n)} + c\,\boldsymbol{\sigma}_{T,i}(\boldsymbol{p}^{(n)})]_1 > c\,\beta\,c_{T,i} \right\},$$

$$\mathcal{A}_T^{\pm,2} := \left\{ (T,i) : \pm[\boldsymbol{\mu}_{T,i}^{(n)} + c\,\boldsymbol{\sigma}_{T,i}(\boldsymbol{p}^{(n)})]_2 > c\,\beta\,c_{T,i} \right\},$$

$$\mathcal{A}_E^{\pm} := \left\{ (E,j) : \pm[\mu_{E,j}^{(n)} + c\,\sigma_{E,j}(\boldsymbol{p}^{(n)})] > c\,\beta\,|\boldsymbol{n}_E|_1\,c_{E,j} \right\}$$

5:    Solve for $\boldsymbol{p} \in \mathcal{R}\mathcal{T}_{r+1}^0(\Omega)$ and assign the solution to $\boldsymbol{p}^{(n+1)}$

$$\text{Minimize} \quad \frac{1}{2}\|\operatorname{div}\boldsymbol{p} + f\|_{L^2(\Omega)}^2,$$

(5.21)
$$\text{s.t.} \quad \begin{cases} [\boldsymbol{\sigma}_{T,i}(\boldsymbol{p})]_1 = \pm\beta\,c_{T,i} & \text{where } (T,i) \in \mathcal{A}_T^{\pm,1} \\ [\boldsymbol{\sigma}_{T,i}(\boldsymbol{p})]_2 = \pm\beta\,c_{T,i} & \text{where } (T,i) \in \mathcal{A}_T^{\pm,2} \\ \sigma_{E,j}(\boldsymbol{p}) = \pm\beta\,|\boldsymbol{n}_E|_1\,c_{E,j} & \text{where } (E,j) \in \mathcal{A}_E^{\pm} \end{cases}$$

6:    Set $u^{(n+1)} := \operatorname{div}\boldsymbol{p}^{(n+1)} + f \in \mathcal{D}\mathcal{G}_r(\Omega)$

7:    Set

$$[\boldsymbol{\mu}_{T,i}^{(n+1)}]_1 := [\nabla u^{(n+1)}(x_{T,i})]_1 \quad \text{where } (T,i) \in \mathcal{A}_T^{\pm,1} \quad \text{and} \quad [\boldsymbol{\mu}_{T,i}^{(n+1)}]_1 := 0 \quad \text{elsewhere}$$

$$[\boldsymbol{\mu}_{T,i}^{(n+1)}]_2 := [\nabla u^{(n+1)}(x_{T,i})]_2 \quad \text{where } (T,i) \in \mathcal{A}_T^{\pm,2} \quad \text{and} \quad [\boldsymbol{\mu}_{T,i}^{(n+1)}]_2 := 0 \quad \text{elsewhere}$$

$$\mu_{E,j}^{(n+1)} := [\![u^{(n+1)}]\!](x_{E,j}) \quad \text{where } (E,j) \in \mathcal{A}_E^{\pm} \quad \text{and} \quad \mu_{E,j}^{(n+1)} := 0 \quad \text{elsewhere}$$

8:    Set $n := n + 1$

9: **end while**

---

uses degrees of freedom based on point evaluations of $\boldsymbol{p}$ and $\boldsymbol{p} \cdot \boldsymbol{n}_E$, rather than the integral-type dofs in (2.5). Since we wish to take advantage of the simple structure of the constraints in the dual representation (3.3) of $|u|_{DTV(\Omega)}$ however, we rely on the choice of dofs described in (2.5). In order to avoid a global basis transformation, we implemented our own version of the $\mathcal{R}\mathcal{T}_{r+1}$ finite element in FENICS.

Our implementation uses the dofs in (2.5) on the reference cell $\widehat{T}$. As usual in finite element methods, an arbitrary cell T is then obtained via an affine geometry transformation, i.e.,

$$G_T : \widehat{T} \to T, \qquad G_T(\widehat{x}) = B_T\,\widehat{x} + b_T,$$

where $B_T \in \mathbb{R}^{2 \times 2}$ is a non-singular matrix and $b_T \in \mathbb{R}^2$. We mention that B_T need not

necessarily have a positive determinant, i.e., the transformation G_T may not necessarily be orientation preserving. In contrast to $\mathcal{CG}$ and $\mathcal{DG}$ elements, a second transformation is required to define the dofs and basis functions on the world cell T from the dofs and basis functions on $\widehat{T}$. For the ($\boldsymbol{H}(\mathrm{div};\Omega)$-conforming) $\mathcal{RT}$ spaces, this is achieved via the (contravariant) Piola transform; see for instance [24, Ch. 1.4.7] or [42]. In terms of functions $\widehat{\boldsymbol{p}}$ from the local polynomial space, we have

$$P_T : \mathcal{P}_r(\widehat{T})^2 + \widehat{\boldsymbol{x}}\,\mathcal{P}_r(\widehat{T}) \to \mathcal{P}_r(T)^2 + \boldsymbol{x}\,\mathcal{P}_r(T), \qquad P_T(\widehat{\boldsymbol{p}}) = (\det B_T^{-1})\,B_T\,[\widehat{\boldsymbol{p}} \circ G_T^{-1}].$$

The Piola transform preserves tangent directions on edges, as well as normal traces of vector fields, up to edge lengths. It satisfies

(6.1) $$|\widehat{E}|\,\widehat{\boldsymbol{p}} \cdot \widehat{\boldsymbol{n}}_{\widehat{E}} = \pm|E|\,\boldsymbol{p} \cdot \boldsymbol{n}_E \quad \text{and} \quad |\widehat{T}|\,B_T\,\widehat{\boldsymbol{p}} = \pm|T|\,\boldsymbol{p},$$

where $\widehat{E}$ is an edge of $\widehat{T}$, $\widehat{\boldsymbol{n}}_{\widehat{E}}$ is the corresponding unit outer normal, $E = G_T(\widehat{E})$, $\boldsymbol{n}_E$ is a unit normal vector on E with arbitrary orientation, $\boldsymbol{p} = P_T(\widehat{\boldsymbol{p}})$, and $|T|$ is the area of T; see for instance [24, Lem. 1.84].

We denote by $\widehat{\boldsymbol{\sigma}}_{\widehat{T},i}$ and $\widehat{\sigma}_{\widehat{E},j}$ the degrees of freedom as in (2.5), defined in terms of the nodal basis functions $\widehat{\varphi}_{\widehat{T},i} \in \mathcal{P}_{r-1}(\widehat{T})$ and $\widehat{\varphi}_{\widehat{E},j} \in \mathcal{P}_r(\widehat{E})$ on the reference cell. Let us consider how these degrees of freedom act on the world cell. Indeed, the relations above imply

(6.2a) $$\widehat{\boldsymbol{\sigma}}_{\widehat{T},i}(\widehat{\boldsymbol{p}}) := \int_{\widehat{T}} \widehat{\varphi}_{\widehat{T},i}\,\widehat{\boldsymbol{p}}\,\mathrm{d}\widehat{x} = \pm \int_T \varphi_{T,i}\,B_T^{-1}\,\boldsymbol{p}\,\mathrm{d}x =: \pm\widetilde{\boldsymbol{\sigma}}_{T,i}(\boldsymbol{p}),$$

(6.2b) $$\widehat{\sigma}_{\widehat{E},j}(\widehat{\boldsymbol{p}}) := \int_{\widehat{E}} \widehat{\varphi}_{\widehat{E},j}\,(\widehat{\boldsymbol{p}} \cdot \widehat{\boldsymbol{n}}_{\widehat{E}})\,\mathrm{d}\widehat{s} = \pm \int_E \varphi_{E,j}\,(\boldsymbol{p} \cdot \boldsymbol{n}_E)\,\mathrm{d}S = \pm\sigma_{E,j}(\boldsymbol{p}),$$

where we used that Lagrangian basis functions are transformed according to $\varphi_{T,i} = \widehat{\varphi}_{\widehat{T},i} \circ G_T^{-1}$, and similarly for the edge-based quantities. The correct choice of the sign in (6.1) and (6.2) depends on the sign of $\det B_T$ and on the relative orientations of $P_T(\widehat{\boldsymbol{n}}_{\widehat{E}})$ and $\boldsymbol{n}_E$. However the sign is not important since all operations depending on the dofs or coefficients, such as $\boldsymbol{\sigma}_{T,i}(\boldsymbol{p})$, are sign invariant, notably the constraint set in (4.1).

Notice that while (6.2b) agrees (possibly up to the sign) with our preferred set of edge-based dofs (2.5b), the interior dofs $\widetilde{\boldsymbol{\sigma}}_{T,i}$ available through the transformation (6.2a) are related to the desired dofs $\boldsymbol{\sigma}_{T,i}$ from (2.5a) via

(6.3) $$\boldsymbol{\sigma}_{T,i}(\boldsymbol{p}) = \mathrm{sgn}(\det B_T)\,B_T^\top\,\widetilde{\boldsymbol{\sigma}}_{T,i}(\boldsymbol{p}).$$

Notice that this transformation is impossible to avoid since the dofs (2.5a) are not invariant under the Piola transform. However, (6.3) is completely local to the triangle and inexpensive to evaluate. Although not required for our numerical computations, we mention for completeness that the corresponding dual basis functions are related via

(6.4) $$\boldsymbol{\psi}_i^T = \mathrm{sgn}(\det B_T)\,\widetilde{\boldsymbol{\psi}}_i^T\,B_T^{-\top}.$$

To summarize this discussion, functions $p \in \mathcal{RT}_{r+1}(\Omega)$ will be represented in terms of coefficients w.r.t. the dofs $\{\sigma_{E,j}\}$ and $\{\widetilde{\sigma}_{T,i}\}$ in our FENICS implementation of the $\mathcal{RT}$ space. Transformations to and from the desired dofs $\{\sigma_{T,i}\}$ will be performed for all operations manipulating directly the coefficients of an $\mathcal{RT}_{r+1}$ function. For instance, the projection operation in (5.12) (for the Chambolle–Pock Algorithm 5.2) in the case $s = 2$ would be implemented as

$$\widetilde{\sigma}_{T,i}(p) = B_T^{-\top} \min\left\{|B_T^\top \widetilde{\sigma}_{T,i}(\bar{p})|_2, \ \beta\, c_{T,i}\right\} \frac{B_T^\top \widetilde{\sigma}_{T,i}(\bar{p})}{|B_T^\top \widetilde{\sigma}_{T,i}(\bar{p})|_2}.$$

7. Numerical Results for (DTV-L2).

In this section we present some numerical results for (DTV-L2). Our main goal here is to study the impact of the choice of the polynomial degree $r \in \{0, 1, 2\}$ on image quality as well as computational efficiency. We focus on two different test cases. On the one hand, we denoise higher order image data, i.e., images given in $\mathcal{DG}_r(\Omega)$ for some $r \geq 1$. On the other hand, we consider the denoising of $\mathcal{DG}_0$ image data, by first interpolating it into a higher order $\mathcal{DG}_r$ space.
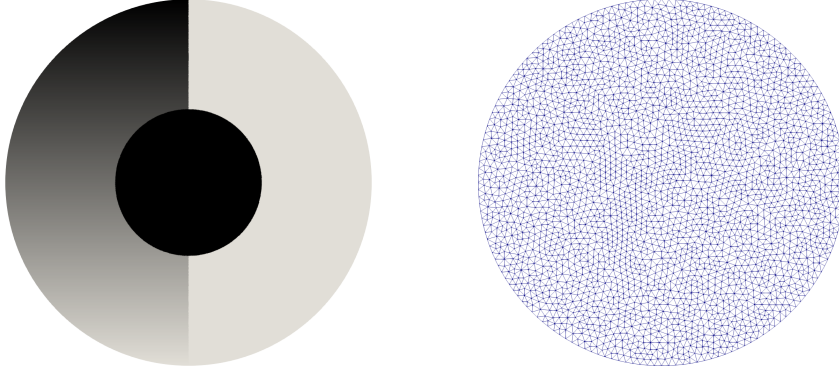
Figure 7.1. *Left: Idealized test picture. Right: Mesh used to represent the picture with varying $\mathcal{DG}$-order.*

In either situation, we use the non-pixelated (continuous) image data with values in the interval $[0, 1]$ and geometry as shown in Figure 7.1. The geometry is a circle of radius 0.5 centered at the origin. The mesh consists of 5460 cells and 2811 points and the number of unknowns for the respective $\mathcal{DG}$ spaces can be found, according to Table 2.1, to be 5460 for $\mathcal{DG}_0$, 16 380 for $\mathcal{DG}_1$, and 32 760 for $\mathcal{DG}_2$. In all of the following tests, noise is added to each degree of freedom in the form of a normally distributed random variable with standard deviation $\sigma = 0.1$ and zero mean.

A stopping criterion for all of Algorithms 5.1 to 5.4 can be based on the primal-dual gap

$$(7.1) \qquad\qquad F(u) + \beta\, G(\Lambda u) + F^*(\Lambda^* p) + \beta\, G^*(p/\beta).$$

Notice that since $G^* = I_P$ is the indicator function of the constraint set P, the last term is either 0 or ∞, and (7.1) can therefore not directly serve as a meaningful stopping criterion. Instead, we omit the last term in (7.1) and instead introduce a distance-to-feasibility measure for p as a second criterion. For the latter, we utilize the difference of p and its Y^*-orthogonal

projection onto $\beta \boldsymbol{P}$, measured in the Y^*-norm squared. This expression can be easily evaluated when $s \in \{1, 2\}$. Straightforward calculations then show that we obtain the following two specific expressions:

$$(7.2a) \qquad \mathrm{GAP}(u, \boldsymbol{p}) := \frac{1}{2}\|u - f\|_{L^2(\Omega)}^2 + \frac{1}{2}\|\mathrm{div}\,\boldsymbol{p} + f\|_{L^2(\Omega)}^2 - \frac{1}{2}\|f\|_{L^2(\Omega)}^2$$

$$+ \beta \sum_T \int_T \mathcal{I}_T\{|\nabla u|_s\}\,\mathrm{d}x + \beta \sum_E \int_E \mathcal{I}_E\{|\llbracket u \rrbracket|_s\}\,\mathrm{d}S,$$

$$(7.2b) \qquad \mathrm{INFEAS}_2(\boldsymbol{p}) := \sum_{T,i} \frac{1}{c_{T,i}} \max\left\{|\boldsymbol{\sigma}_{T,i}(\boldsymbol{p})|_2 - \beta\,c_{T,i},\ 0\right\}^2$$

$$+ \sum_{E,j} \frac{1}{c_{E,j}} \max\left\{|\sigma_{E,j}(\boldsymbol{p})| - \beta\,c_{E,j},\ 0\right\}^2,$$

$$(7.2c) \qquad \mathrm{INFEAS}_1(\boldsymbol{p}) := \sum_{T,i} \frac{1}{c_{T,i}} \sum_{\ell=1}^{2} \max\left\{|[\boldsymbol{\sigma}_{T,i}(\boldsymbol{p})]_\ell| - \beta\,c_{T,i},\ 0\right\}^2$$

$$+ \sum_{E,j} \frac{1}{c_{E,j}} \max\left\{|\sigma_{E,j}(\boldsymbol{p})| - \beta\,|\boldsymbol{n}_E|_s\,c_{E,j},\ 0\right\}^2.$$

We report on the results using Algorithm 5.1 (split Bregman). We focus on the isotropic case $s = 2$ and stopped after 300 iterations.

7.1. Denoising of $\mathcal{DG}_r$-Images. For this test case, we represent (interpolate) the image data displayed in Figure 7.1 in the space $\mathcal{DG}_r(\Omega)$ for $r = 0, 1, 2$. Noise is added as described above. The results of Algorithm 5.1 are shown in Figure 7.2. In either case, the noise is removed successfully and the duality gap (7.2a) is monotonically decreasing for all degrees. The infeasibility criterion (7.2b) was equal to zero throughout the iteration. Although we found the higher order problems slightly slower to converge, they accomodate changes of the image data within each cell, which leads to better resolution and more visually pleasing pictures.

7.2. Denoising of $\mathcal{DG}_0$-Images into Higher-Order Spaces. In this section we interpolate a noisy $\mathcal{DG}_0$ image into $\mathcal{DG}_r(\Omega)$ for $r = 0, 1, 2$ and carry out the denoising in $\mathcal{DG}_r(\Omega)$. The purpose of this experiment is to demonstrate the benefit of using higher-order elements even for low-fidelity image data. The results of Algorithm 5.1 are shown in Figure 7.3. In the case of $\mathcal{DG}_0$, we clearly see the well known staircasing effect, which is not visible in the higher-order cases. Again, we found the algorithm slightly slower to converge, As in the first experiment, the infeasibility criterion (7.2b) was equal to zero throughout the iteration.

8. Solving the (DTV-L1) **Problem.** We briefly discuss the implementation of two algorithms for

$$(\text{DTV-L1}) \qquad \text{Minimize} \quad \sum_T \int_T \mathcal{J}_T\{|u - f|\}\,\mathrm{d}x + \beta\,|u|_{DTV(\Omega)}$$
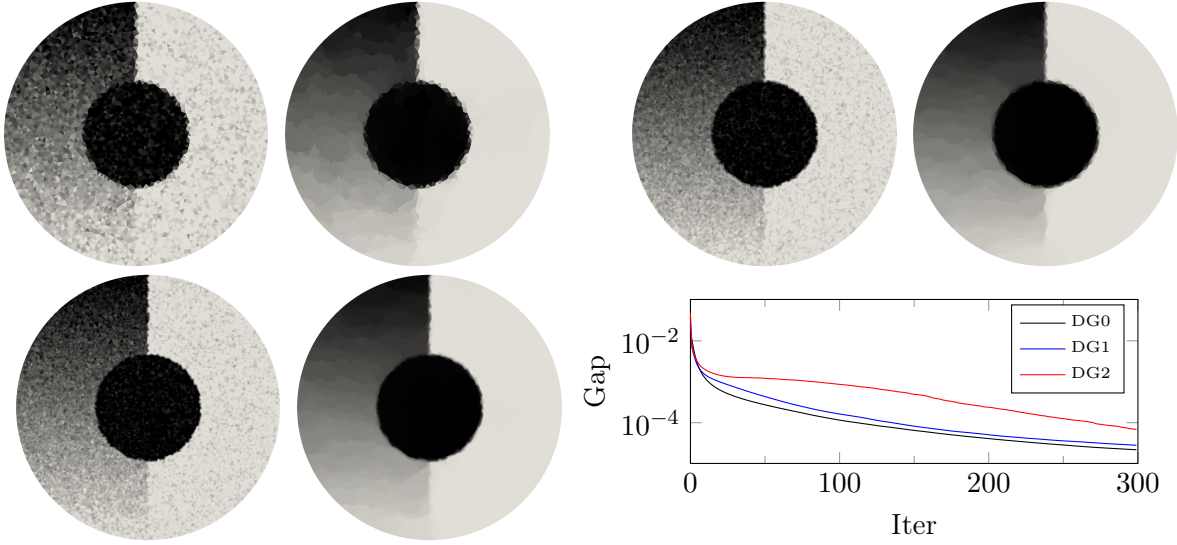
Figure 7.2. *Noisy and denoised images for $\mathcal{DG}_0$ (top left), $\mathcal{DG}_1$ (top right) and $\mathcal{DG}_2$ (bottom left), and duality gap criterion (7.2a), using Algorithm 5.1 (split Bregman) with L^1-penalty parameter $\beta = 0.001$ and Augmented Lagrangian parameter $\lambda = 0.001$ for the isotropic setting ($s = 2$).*

with $s \in [1, \infty]$. They too can be realized equally efficiently as their original counterparts devised for images on Cartesian grids with low-order finite difference approximations of the gradient and divergence. For simplicity, we restrict the discussion to the polynomial degrees $r \in \{0, 1\}$ in this section so that all weights $c_{T,i}$, $c_{E,j}$ as well as $C_{T,k}$ are strictly positive. The cases $r = \{2, 3\}$ can be included provided that zero weights are properly treated and we come back to this in subsection 9.2.

8.1. Chambolle–Pock Method.
We focus on the changes compared to the method for (DTV-L2) discussed in subsection 5.2. As in subsection 4.2, we need to replace F by (4.8) and use the lumped inner product (4.9) in $U = \mathcal{DG}_r(\Omega)$. Due to the diagonal structure of both F and the inner product, the F-prox operator is easily seen to be

$$(8.1) \qquad u = \text{prox}_{\sigma F}(\bar{u}) \quad \Leftrightarrow \quad u_{T,k} = f_{T,k} + \text{shrink}\left(|\bar{u}_{T,k} - f_{T,k}|,\ \sigma\right),$$

similarly as in [14, Sect. 6.2.2]. The remaining steps in Algorithm 5.2 are unaffected.

8.2. ADMM Method.
Finally we consider the Alternating Direction Method of Multipliers (ADMM) for the primal problem (DTV-L1) as in [47]. In our context, similar as for the split Bregman method (subsection 5.1), one introduces variables $u \in \mathcal{DG}_r(\Omega)$ and $\boldsymbol{d}, \boldsymbol{b} \in Y$. A second splitting $e = u - f$ is required, so we additionally introduce $e \in \mathcal{DG}_r(\Omega)$ as well as a
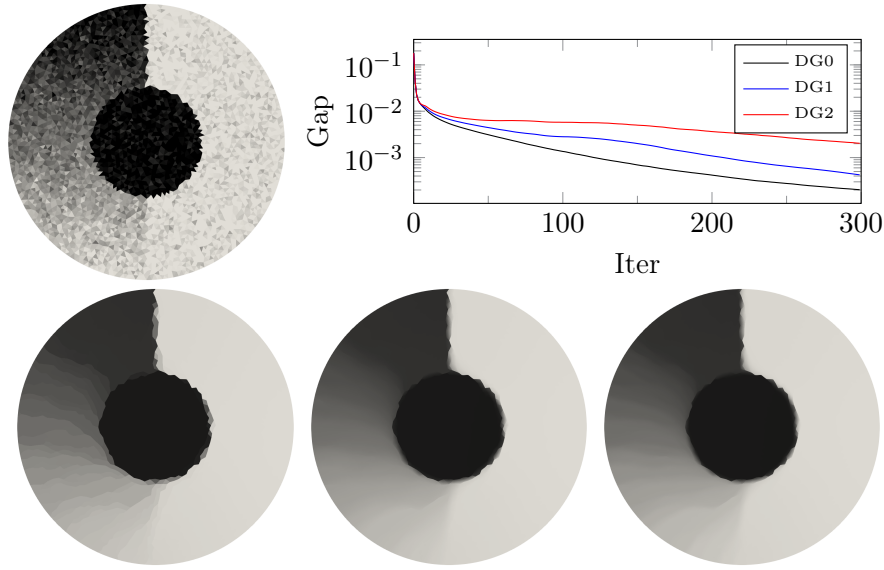
Figure 7.3. *Noisy $\mathcal{DG}_0$ image (top) and denoised images for $\mathcal{DG}_0$ (bottom left), $\mathcal{DG}_1$ (bottom middle) and $\mathcal{DG}_2$ (bottom right), and duality gap criterion (7.2a), using Algorithm 5.1 (split Bregman) with L^1-penalty parameter $\beta = 0.1$ and Augmented Lagrangian parameter $\lambda = 0.01$ for the isotropic setting ($s = 2$).*

multiplier $g \in \mathcal{DG}_r(\Omega)$. The corresponding augmented Lagrangian functional reads

$$(8.2) \quad \sum_{T,k} |e_{T,k}| \, C_{T,k} + \beta \sum_{T,i} c_{T,i} \left| \boldsymbol{d}_{T,i} \right|_s + \beta \sum_{E,j} |\boldsymbol{n}_E|_s \, c_{E,j} \, |d_{E,j}|$$

$$+ \frac{\lambda}{2} \sum_{T,i} c_{T,i} \left| \boldsymbol{d}_{T,i} - \nabla u(x_{T,i}) - \boldsymbol{b}_{T,i} \right|_2^2 + \frac{\lambda}{2} \sum_{E,j} c_{E,j} \left| d_{E,j} - [\![u]\!](x_{E,j}) - b_{E,j} \right|^2$$

$$+ \frac{\lambda}{2} \sum_{T,k} C_{T,k} \left| e_{T,k} - u_{T,k} + f_{T,k} - g_{T,k} \right|_2^2.$$

Let us briefly consider the individual minimization problems w.r.t. u, $\boldsymbol{d}$, and e. The u-problem is to minimize

$$(8.3) \quad \frac{\lambda}{2} \sum_{T,k} C_{T,k} \left| e_{T,k} - u_{T,k} + f_{T,k} - g_{T,k} \right|_2^2$$

$$+ \frac{\lambda}{2} \sum_{T,i} c_{T,i} \left| \boldsymbol{d}_{T,i} - \nabla u(x_{T,i}) - \boldsymbol{b}_{T,i} \right|_2^2 + \frac{\lambda}{2} \sum_{E,j} c_{E,j} \left| d_{E,j} - [\![u]\!](x_{E,j}) - b_{E,j} \right|^2$$

w.r.t. $u \in \mathcal{DG}_r(\Omega)$. This problem is similar to (5.5) and it leads to a coupled linear system for u. The minimization of (8.2) w.r.t. $\boldsymbol{d} \in \mathcal{RT}_{r+1}^0(\Omega)$ is identical to (5.6) and the e-problem is to minimize

$$(8.4) \quad \sum_{T,k} |e_{T,k}| \, C_{T,k} + \frac{\lambda}{2} \sum_{T,k} C_{T,k} \left| e_{T,k} - u_{T,k} + f_{T,k} - g_{T,k} \right|_2^2.$$

This problem can be easily solved via shrinkage, cf. (5.7b). Finally, the multiplier update for $\boldsymbol{b}$ is as in subsection 5.1, and the update for g is similar; see Algorithm 8.1. Once again, the solution $\boldsymbol{p}$ of the dual problem (DTV-L1-D) can be recovered from the multipliers $\boldsymbol{b}_{T,i}$ and $b_{E,j}$ as in (5.8). Moreover, it can be easily checked that

$$(8.5) \qquad \lambda\, g_{T,k} = \frac{1}{C_{T,k}} \int_T (\operatorname{div} \boldsymbol{p})\, \Phi_{T,k} \, \mathrm{d}x$$

holds, where the quantity on the right appears as a constraint in (DTV-L1-D) and thus it satisfies $|\lambda\, g_{T,k}| \leq 1$ in the limit.

Algorithm 8.1 ADMM algorithm for (DTV-L1) with $s \in [1, \infty]$

1: Set $u^{(0)} := f \in \mathcal{DG}_r(\Omega)$, $\boldsymbol{b}^{(0)} := \boldsymbol{0} \in Y$ and $\boldsymbol{d}^{(0)} := \boldsymbol{0} \in Y$
2: Set $e^{(0)} := 0 \in \mathcal{DG}_r(\Omega)$ and $g^{(0)} := 0 \in \mathcal{DG}_r(\Omega)$
3: Set $n := 0$
4: **while** not converged **do**
5:      Minimize (8.3) for $u^{(n+1)}$ with data $\boldsymbol{b}^{(n)}$, $\boldsymbol{d}^{(n)}$, $e^{(n)}$ and $g^{(n)}$
6:      Minimize (5.7) for $\boldsymbol{d}^{(n+1)}$ with data $u^{(n+1)}$ and $\boldsymbol{b}^{(n)}$
7:      Minimize (8.4) for $e^{(n+1)}$ with data $u^{(n+1)}$ and $g^{(n)}$
8:      Set $\boldsymbol{b}_{T,i}^{(n+1)} := \boldsymbol{b}_{T,i}^{(n)} + \nabla u^{(n+1)}(x_{T,i}) - \boldsymbol{d}_{T,i}^{(n+1)}$ and $b_{E,j}^{(n+1)} := b_{E,j}^{(n)} + [\![u^{(n+1)}]\!](x_{E,j}) - d_{E,j}^{(n+1)}$
9:      Set $g_{T,k}^{(n+1)} := g_{T,k}^{(n)} + u_{T,k}^{(n+1)} - f_{T,k} - e_{T,k}^{(n+1)}$
10:     Set $n := n + 1$
11: **end while**
12: Set $\boldsymbol{p}^{(n)}$ by (5.8) with data $\boldsymbol{b}^{(n)}$

---

9. Extensions.
In this section we collect a number of extensions showing that problems more general than those based on the TV-L^2 and TV-L^1 models and discontinuous functions can be dealt with efficiently by generalizations of the respective algorithms to our higher-order finite element setting.

9.1. Huber TV-Seminorm.
We consider the replacement of the TV-seminorm by its 'Huberized' variant; see for instance [35, 39] and [44, Ch. 4]. In the case $s = 2$, on which we focus here, the function G in (4.3) can be written as

$$G(\boldsymbol{d}) = \sum_T \int_T \mathcal{I}_T \{|\boldsymbol{d}_T|_2\} \, \mathrm{d}x + \sum_E \int_E \mathcal{I}_E \{|d_E|\} \, \mathrm{d}S = \sum_{T,i} c_{T,i} \, |\boldsymbol{d}_{T,i}|_2 + \sum_{E,j} c_{E,j} \, |d_{E,j}|.$$

The corresponding Huber functional with parameter $\varepsilon > 0$ then becomes

$$(9.1) \qquad G_\varepsilon(\boldsymbol{d}) = \sum_{T,i} c_{T,i} \, \max\left\{ |\boldsymbol{d}_{T,i}|_2 - \frac{\varepsilon}{2}, \, \frac{1}{2\varepsilon} |\boldsymbol{d}_{T,i}|_2^2 \right\}$$

$$+ \sum_{E,j} c_{E,j} \, \max\left\{ d_{E,j} - \frac{\varepsilon}{2}, \, -d_{E,j} - \frac{\varepsilon}{2}, \, \frac{1}{2\varepsilon} (d_{E,j})^2 \right\}.$$

It can be shown by straightforward calculations that the convex conjugate of G_ε is

$$(9.2) \qquad G_\varepsilon^*(\boldsymbol{p}) = I_{\boldsymbol{P}} + \frac{\varepsilon}{2}\|\boldsymbol{p}\|_{Y^*}^2.$$

We recall that $I_{\boldsymbol{P}}$ is the indicator function of the constraint set $\boldsymbol{P}$ in (4.1).

The 'Huberized' discrete TV-seminorm is thus defined by $G_\varepsilon(\Lambda u)$ where Λ is given in (4.2). It can be combined with both the L^2 and L^1 loss terms,

$$F(u) = \frac{1}{2}\|u - f\|_{L^2(\Omega)}^2 \quad \text{and} \quad F(u) = \sum_T \int_T \mathcal{J}_T\{|u - f|\}\,\mathrm{d}x.$$

We refer to the corresponding primal problems, i.e., the minimization of $F(u) + \beta\,G_\varepsilon(\Lambda u)$, as (DTV$_\varepsilon$-L2) and (DTV$_\varepsilon$-L1). The specific form of corresponding dual problems, where $F^*(-\Lambda^*\boldsymbol{p}) + \beta\,G_\varepsilon^*(\boldsymbol{p}/\beta)$ is minimized, should now also be clear.

As expected, straightforward modifications of the split Bregman method (Algorithm 5.1) in (5.7), of the Chambolle–Pock method (Algorithm 5.2) in (5.12), and of Chambolle's projection method (Algorithm 5.3) in (5.17) and (5.18) apply to tailor them to (DTV$_\varepsilon$-L2) in the case $s = 2$. The modification necessary to the primal-dual active set method (Algorithm 5.4) for $s = 1$ was already discussed in subsection 5.4. Similarly, the methods of section 8 can be adapted to solve (DTV$_\varepsilon$-L1) in case $s = 2$. We leave the details to the reader.

9.2. Polynomial Degrees.
We recall that we restricted the discussion of algorithms for (DTV-L2) and its dual (DTV-L2-D) in section 5 to the cases $r \in \{0, 1, 2, 4\}$, each of which ensures that $c_{T,i}$ and $c_{E,j}$ are strictly positive; see Lemma 3.1. In the case $r = 3$, three of the six weights $c_{T,i}$ on each triangle are zero. This is not a major issue but it requires some care when formulating the algorithms in section 5 in this case. Briefly, when $c_{T,i} = 0$, quantities bearing the same index (T, i) are to be ignored. This applies, in particular, to the inner product $(\cdot, \cdot)_{Y^*}$ in (5.10).

Similarly, we excluded the cases $r \in \{2, 3\}$ in the discussion of algorithms for (DTV-L1) and its dual (DTV-L1-D) in section 8 so that the weights $C_{T,k} := \int_T \Phi_{T,k}\,\mathrm{d}x$ pertaining to the basis $\{\Phi_{T,k}\}$ of $\mathcal{P}_r(T)$ are strictly positive as well. In case $r = 3$, we proceed as discussed above, ignoring terms for which the corresponding weights $c_{T,i} = 0$. When $r = 2$, we instead ignore terms for which $C_{T,k} = 0$ in any of the algorithms in section 8.

9.3. Images in $\mathcal{CG}_r(\Omega)$.
While we believe that the representation of images as discontinuous functions is rather natural, it is certainly useful to consider also the case when $u \in \mathcal{CG}_r(\Omega)$. This situation is meaningful only for $r \geq 1$, and hence we consider $r \in \{1, 2, 3, 4\}$ in this section. Clearly, for $u \in \mathcal{CG}_r(\Omega)$, the TV-seminorm (1.2) and its discrete counterpart (1.3) reduce to

$$(9.3) \qquad |u|_{TV(\Omega)} = \sum_T \int_T |\nabla u|_s\,\mathrm{d}x \quad \text{and} \quad |u|_{DTV(\Omega)} = \sum_T \int_T \mathcal{I}_T\{|\nabla u|_s\}\,\mathrm{d}x$$

since the terms related to edge jumps disappear. As was mentioned in the introduction, the lowest-order case $r = 1$ has been considered in [26, 23, 5, 6, 8, 17]. In this case, $|u|_{TV(\Omega)} = |u|_{DTV(\Omega)}$ holds. Similarly as in Corollary 3.5, a simple convexity argument shows that $|u|_{TV(\Omega)} \le |u|_{DTV(\Omega)}$ holds for all $u \in \mathcal{CG}_2(\Omega)$.

Since $\mathcal{CG}_r(\Omega)$ is a proper subspace of $\mathcal{DG}_r(\Omega)$, it can be expected that it is enough to take the supremum in Theorem 3.2 over a smaller set of test functions. Indeed, as the image of the gradient operator $\Lambda : U = \mathcal{CG}_r(\Omega) \to Y$ reduces to $Y = \prod_T \mathcal{P}_{r-1}(T)^2$, the edge-based dofs of $\boldsymbol{p} \in \mathcal{RT}_{r+1}^0(\Omega)$ that can be dispensed with since no edge jumps need to be measured. We thus obtain the following corollary of Theorem 3.2.

Corollary 9.1 (Dual Representation of $|u|_{DTV(\Omega)}$ for $u \in \mathcal{CG}_r(\Omega)$). *Suppose $r \in \{1, 2, 3, 4\}$. Then for any $u \in \mathcal{CG}_r(\Omega)$, the discrete TV-seminorm* (1.3) *reduces to* (9.3) *and it satisfies*

$$
|u|_{DTV(\Omega)} = \sup \Bigg\{ \int_\Omega u \operatorname{div} \boldsymbol{p} \, \mathrm{d}x : \boldsymbol{p} \in \mathcal{RT}_{r+1}^0(\Omega),
$$

$$
|\boldsymbol{\sigma}_{T,i}(\boldsymbol{p})|_{s^*} \le c_{T,i} \text{ for all triangles } T \text{ and all } i = 1, \ldots, r\,(r+1)/2,
$$

$$
(9.4) \qquad \sigma_{E,j}(\boldsymbol{p}) = 0 \text{ for all interior edges } E \text{ and all } j = 1, \ldots, r+1 \Bigg\}.
$$

It is straightforward to adopt the algorithms presented in sections 5 and 8 to this simpler situation. In a nutshell, all edge-based quantities (such as $d_{E,j}$ and $b_{E,j}$ in the split Bregman method, Algorithm 5.1) can be ignored, and the edge-based coefficients $\sigma_{E,j}(\boldsymbol{p})$ of any function $\boldsymbol{p} \in \mathcal{RT}_{r+1}^0(\Omega)$ would be left at zero.

We remark, however, that the gradient operator is not surjective onto Y unless $r = 1$ holds. In case $r \in \{2, 4\}$, the set of test functions $\boldsymbol{p}$ in (9.4) is unnecessarily large. A more economical formulation for these cases remains open for future investigation.

10. Conclusion and Outlook. In this paper we have introduced a discrete version (DTV) of the TV-seminorm for globally discontinuous ($\mathcal{DG}_r$) and continuous ($\mathcal{CG}_r$) Lagrangian finite element functions on triangular grids in $\mathbb{R}^2$. We have shown that $|\cdot|_{DTV(\Omega)}$ has a convenient dual representation in terms of the supremum over the space of Raviart–Thomas finite element functions, subject to a set of simple constraints. This allows for the efficient realization of a variety of algorithms, e.g., for (DTV-L2-D) and (DTV-L1-D), both with low and higher-order finite element functions available in finite element libraries.

Since we admit higher-order polynomial functions, it would be natural to extend our analysis to a discrete version of the total generalized variation (TGV) functional introduced in [9]. Another generalization that could be of interest is to consider finite element functions defined on more general cells than the simplices considered here. Clearly rectangles are of particular interest in imaging applications, but also hexagons; see [34, 18], as mentioned in the introduction.

The polynomial degree in our study was limited to $0 \leq r \leq 4$ (or $0 \leq r \leq 3$ for (DTV-L1)), which should be sufficient for most applications. The limitation in the degree arises due to the requirement that the quadrature weights, i.e., the integrals over the standard Lagrangian basis functions, have to be non-negative; see Lemma 3.1. Our approach carries over with no changes to tetrahedral grids on 3D domains. In this case the analog of Lemma 3.1 limits the polynomial degrees with non-negative weights to $r \in \{0, 1, 2, 4\}$; see [45, Tab. II]. When (DTV-L1) is considered, only the choices $r \in \{0, 1\}$ remain.

This brings up the question whether a Lagrangian basis for higher-order polynomial functions on triangles or tetrahedra exists, such that the integrals of the basis functions are (strictly) positive. This is answered in the affirmative by results in [49, 48] for the triangle and [27, 51], where interpolatory quadrature formulas with positive weights are constructed. However, it remains to be investigated whether a Lagrangian finite element with a modified basis admits an appropriate Raviart–Thomas type counterpart such that a dual representation of $|\cdot|_{DTV(\Omega)}$ parallel to Theorem 3.2 continues to hold. Moreover, such non-standard finite element spaces certainly incur an overhead in implementation.

One may also envision applications where it would be beneficial to allow for locally varying polynomial degrees and mesh sizes in imaging applications, so that the resolution can be chosen adaptively. Finally, we mention possible extensions to vectorial TV-seminorms, see for instance [28]. These topics remain for future research.

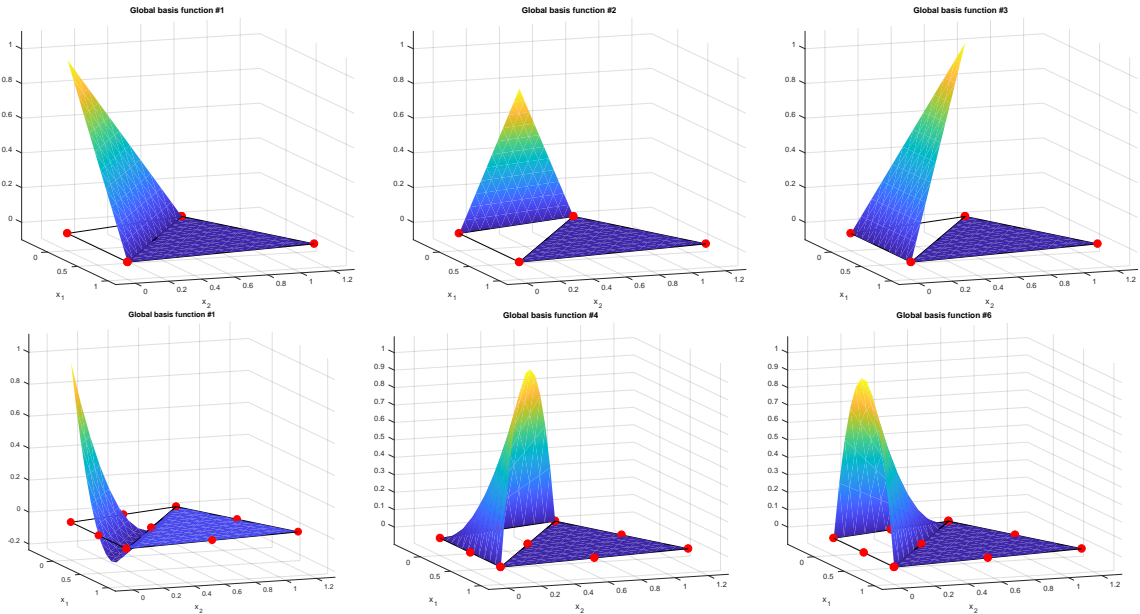## Appendix A. Examples of Finite Element Basis Functions.

Figure A.1. *Some basis functions* $\{\Phi_{T,k}\}$ *of* $\mathcal{DG}_r(\Omega)$ *for* $r = 1$ *(top row) and* $r = 2$ *(bottom row).*
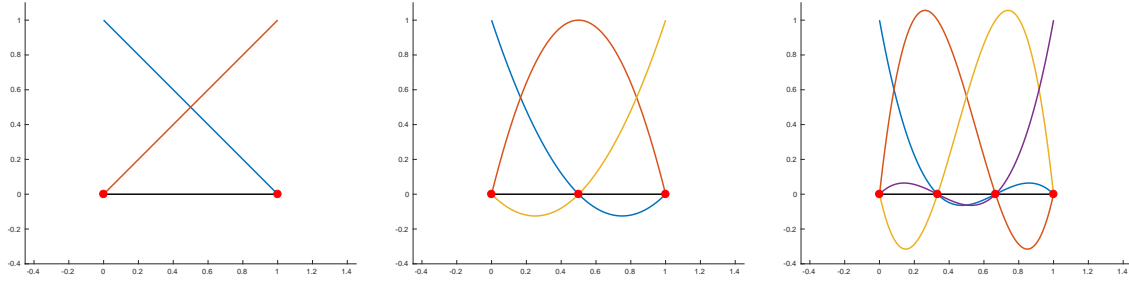
Figure A.2. *Complete set of basis functions* $\{\varphi_{E,j}\}$ *of* $\mathcal{P}_r(E)$ *for* $r \in \{1, 2, 3\}$.
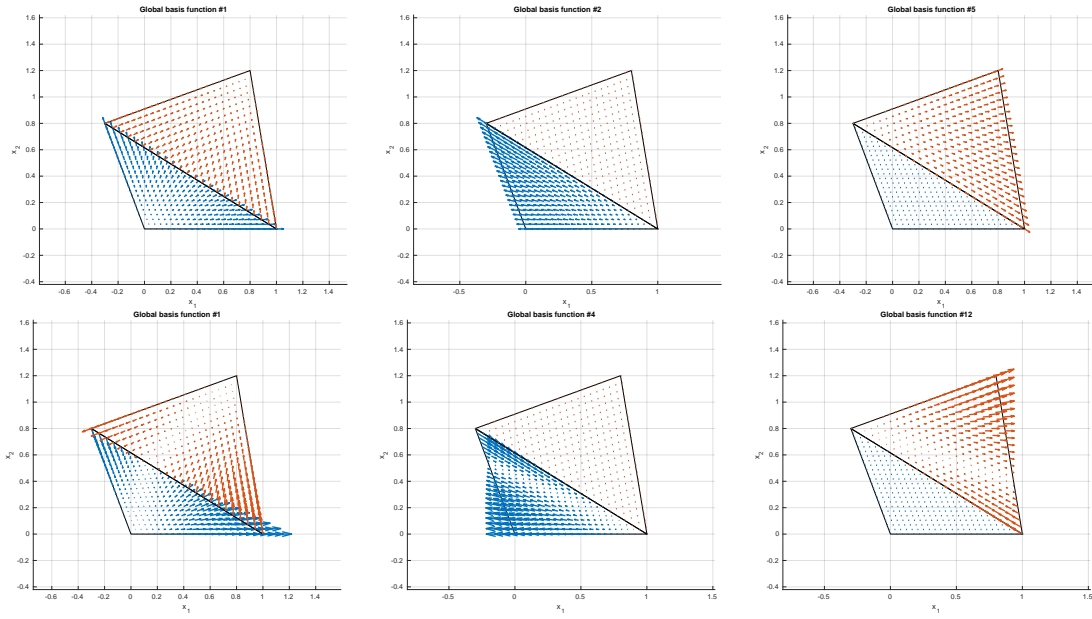

Figure A.3. *Some basis functions* $\{\boldsymbol{\psi}_j^E\}$ *of* $\mathcal{RT}_{r+1}(\Omega)$ *for* $r = 0$ *(top row) and* $r = 1$ *(bottom row).*

quadrature schemes in FEniCS. Part of this research was contrived while the second author was visiting the University of British Columbia, Vancouver. He would like to thank the Department of Computer Science for their hospitality.

REFERENCES

[1] M. ALKÄMPER AND A. LANGER, *Using DUNE-ACFem for non-smooth minimization of bounded variation functions*, Archive of Numerical Software, 5 (2017), pp. 3–19, https://doi.org/10.11588/ans.2017.1. 27475.

[2] M. ALNÆS, J. BLECHTA, J. HAKE, A. JOHANSSON, B. KEHLET, A. LOGG, C. RICHARDSON, J. RING, M. E. ROGNES, AND G. N. WELLS, *The FEniCS project version 1.5*, Archive of Numerical Software, 3 (2015), pp. 9–23, https://doi.org/10.11588/ans.2015.100.20553.

[3] P. F. ANTONIETTI AND B. AYUSO, *Schwarz domain decomposition preconditioners for discontinuous Galerkin approximations of elliptic problems: non-overlapping case*, M2AN. Mathematical Modelling

and Numerical Analysis, 41 (2007), pp. 21–54, https://doi.org/10.1051/m2an:2007006.

[4] M. Bachmayr and M. Burger, *Iterative total variation schemes for nonlinear inverse problems*, Inverse Problems, 25 (2009), pp. 105004, 26, https://doi.org/10.1088/0266-5611/25/10/105004.

[5] S. Bartels, *Total variation minimization with finite elements: convergence and iterative solution*, SIAM Journal on Numerical Analysis, 50 (2012), pp. 1162–1180, https://doi.org/10.1137/11083277X.

[6] S. Bartels, R. H. Nochetto, and A. J. Salgado, *Discrete total variation flows without regularization*, SIAM Journal on Numerical Analysis, 52 (2014), pp. 363–385, https://doi.org/10.1137/120901544.

[7] A. Beck, *First-Order Methods in Optimization*, Society for Industrial and Applied Mathematics, Philadelphia, PA, 2017, https://doi.org/10.1137/1.9781611974997.

[8] B. Berkels, A. Effland, and M. Rumpf, *A posteriori error control for the binary Mumford-Shah model*, Mathematics of Computation, 86 (2017), pp. 1769–1791, https://doi.org/10.1090/mcom/3138.

[9] K. Bredies, K. Kunisch, and T. Pock, *Total generalized variation*, SIAM Journal on Imaging Sciences, 3 (2010), pp. 492–526, https://doi.org/10.1137/090769521.

[10] C. Carstensen, *Quasi-interpolation and a posteriori error analysis in finite element methods*, M2AN. Mathematical Modelling and Numerical Analysis, 33 (1999), pp. 1187–1202, https://doi.org/10.1051/m2an:1999140.

[11] E. Casas, R. Herzog, and G. Wachsmuth, *Approximation of sparse controls in semilinear equations by piecewise linear functions*, Numerische Mathematik, 122 (2012), pp. 645–669, https://doi.org/10.1007/s00211-012-0475-7.

[12] A. Chambolle, *An algorithm for total variation minimization and applications*, Journal of Mathematical Imaging and Vision, 20 (2004), pp. 89–97, https://doi.org/10.1023/B:JMIV.0000011325.36760.1e. Special issue on mathematics and image analysis.

[13] A. Chambolle, V. Caselles, D. Cremers, M. Novaga, and T. Pock, *An introduction to total variation for image analysis*, in Theoretical foundations and numerical methods for sparse recovery, vol. 9 of Radon Series on Computational and Applied Mathematics, Walter de Gruyter, Berlin, 2010, pp. 263–340, https://doi.org/10.1515/9783110226157.263.

[14] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, Journal of Mathematical Imaging and Vision, 40 (2011), pp. 120–145, https://doi.org/10.1007/s10851-010-0251-1.

[15] T. F. Chan and S. Esedoḡlu, *Aspects of total variation regularized L^1 function approximation*, SIAM Journal on Applied Mathematics, 65 (2005), pp. 1817–1837, https://doi.org/10.1137/040604297.

[16] T. F. Chan and X.-C. Tai, *Level set and total variation regularization for elliptic inverse problems with discontinuous coefficients*, Journal of Computational Physics, 193 (2004), pp. 40–66, https://doi.org/10.1016/j.jcp.2003.08.003.

[17] C. Clason, F. Kruse, and K. Kunisch, *Total variation regularization of multi-material topology optimization*, ESAIM. Control, Optimisation and Calculus of Variations, (to appear), https://doi.org/10.1051/m2an/2017061.

[18] S. Coleman, B. Scotney, and B. Gardiner, *Tri-directional gradient operators for hexagonal image processing*, Journal of Visual Communication and Image Representation, 38 (2016), pp. 614–626, https://doi.org/10.1016/j.jvcir.2016.04.001.

[19] L. Condat, *Discrete total variation: New definition and minimization*, SIAM Journal on Imaging Sciences, 10 (2017), pp. 1258–1290, https://doi.org/10.1137/16m1075247.

[20] G. Dahlquist and A. Björck, *Numerical methods in scientific computing. Vol. I*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2008, https://doi.org/10.1137/1.9780898717785.

[21] P. J. Davis and P. Rabinowitz, *Methods of numerical integration*, Computer Science and Applied Mathematics, Academic Press, Inc., Orlando, FL, 2 ed., 1984.

[22] J. Duchi, S. Shalev-Shwartz, Y. Singer, and T. Chandra, *Efficient projections onto the ℓ_1-ball for learning in high dimensions*, in Proceedings of the 25th international conference on Machine learning ICML 2008, ACM Press, 2008, pp. 272–279, https://doi.org/10.1145/1390156.1390191.

[23] C. M. Elliott and S. A. Smitheman, *Numerical analysis of the TV regularization and H^{-1} fidelity model for decomposing an image into cartoon plus texture*, IMA Journal of Numerical Analysis, 29 (2009), pp. 651–689, https://doi.org/10.1093/imanum/drn025.

[24] A. Ern and J.-L. Guermond, *Theory and Practice of Finite Elements*, Springer, Berlin, 2004.

[25] S. Esedoḡlu and S. J. Osher, *Decomposition of images by the anisotropic Rudin-Osher-Fatemi model*, Communications on Pure and Applied Mathematics, 57 (2004), pp. 1609–1626, https://doi.org/10.1002/cpa.20045.

[26] X. Feng and A. Prohl, *Analysis of total variation flow and its finite element approximations*, M2AN. Mathematical Modelling and Numerical Analysis, 37 (2003), pp. 533–556, https://doi.org/10.1051/m2an:2003041.

[27] M. Gellert and R. Harbord, *Moderate degree cubature formulas for 3-d tetrahedral finite-element approximations*, Communications in Applied Numerical Methods, 7 (1991), pp. 487–495, https://doi.org/10.1002/cnm.1630070609.

[28] B. Goldluecke, E. Strekalovskiy, and D. Cremers, *The natural vectorial total variation which arises from geometric measure theory*, SIAM Journal on Imaging Sciences, 5 (2012), pp. 537–564, https://doi.org/10.1137/110823766.

[29] T. Goldstein and S. Osher, *The split Bregman method for L1-regularized problems*, SIAM Journal on Imaging Sciences, 2 (2009), pp. 323–343, https://doi.org/10.1137/080725891.

[30] B. He and X. Yuan, *Convergence analysis of primal-dual algorithms for a saddle-point problem: from contraction perspective*, SIAM Journal on Imaging Sciences, 5 (2012), pp. 119–149, https://doi.org/10.1137/100814494.

[31] M. Hintermüller, K. Ito, and K. Kunisch, *The primal-dual active set strategy as a semismooth Newton method*, SIAM Journal on Optimization, 13 (2002), pp. 865–888.

[32] M. Hintermüller and K. Kunisch, *Total bounded variation regularization as a bilaterally constrained optimization problem*, SIAM Journal on Applied Mathematics, 64 (2004), pp. 1311–1333, https://doi.org/10.1137/S0036139903422784.

[33] J. R. Jensen, *Introductory Digital Image Processing: A Remote Sensing Perspective*, Prentice-Hall Inc., Englewood Cliffs, N.J., 4th ed., 2015.

[34] M. Knaup, S. Steckmann, O. Bockenbach, and M. Kachelriess, *CT image reconstruction using hexagonal grids*, in 2007 IEEE Nuclear Science Symposium Conference Record, IEEE, 2007, pp. 3074–3076, https://doi.org/10.1109/nssmic.2007.4436779.

[35] H. R. Künsch, *Robust priors for smoothing and image restoration*, Annals of the Institute of Statistical Mathematics, 46 (1994), pp. 1–19, https://doi.org/10.1007/BF00773588.

[36] A. Logg, K.-A. Mardal, G. N. Wells, et al., *Automated Solution of Differential Equations by the Finite Element Method*, Springer, 2012, https://doi.org/10.1007/978-3-642-23099-8.

[37] L. D. López Pérez, *Régularisation d'images sur des surfaces non planes*, PhD thesis, Université de Nice-Sophia Antipolis, 2006, https://tel.archives-ouvertes.fr/tel-00141417v1.

[38] M. Nikolova, *A variational approach to remove outliers and impulse noise*, Journal of Mathematical Imaging and Vision, 20 (2004), pp. 99–120, https://doi.org/10.1023/B:JMIV.0000011920.58935.9c. Special issue on mathematics and image analysis.

[39] R. Pan and S. J. Reeves, *Efficient Huber-Markov edge-preserving image restoration*, IEEE Transactions on Image Processing, 15 (2006), pp. 3728–3735, https://doi.org/10.1109/tip.2006.881971.

[40] B. Rivière, *Discontinuous Galerkin Methods for Solving Elliptic and Parabolic Equations*, Society for Industrial and Applied Mathematics, jan 2008, https://doi.org/10.1137/1.9780898717440.

[41] B. Rivière, M. F. Wheeler, and V. Girault, *Improved energy estimates for interior penalty, constrained and discontinuous Galerkin methods for elliptic problems. I*, Computational Geosciences, 3 (1999), pp. 337–360 (2000), https://doi.org/10.1023/A:1011591328604.

[42] M. E. Rognes, R. C. Kirby, and A. Logg, *Efficient assembly of H(div) and H(curl) conforming finite elements*, SIAM Journal on Scientific Computing, 31 (2009), pp. 4130–4151, https://doi.org/10.1137/08073901X.

[43] L. I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Physica D, 60 (1992), pp. 259–268, https://doi.org/10.1016/0167-2789(92)90242-F.

[44] O. Scherzer, M. Grasmair, H. Grossauer, M. Haltmeier, and F. Lenzen, *Variational methods in imaging*, vol. 167 of Applied Mathematical Sciences, Springer, New York, 2009.

[45] P. Silvester, *Symmetric quadrature formulae for simplexes*, Mathematics of Computation, 24 (1970), pp. 95–100, https://doi.org/10.2307/2004880.

[46] B. Stamm and T. P. Wihler, *A total variation discontinuous Galerkin approach for image restoration*, International Journal of Numerical Analysis and Modeling, 12 (2015), pp. 81–93.

[47] M. Tao and J. Yang, *Alternating direction algorithms for total variation deconvolution in image reconstruction*, Tech. Report 2463, Optimization Online, 2009, http://www.optimization-online.org/DB_HTML/2009/11/2463.html.

[48] M. A. Taylor, B. A. Wingate, and L. P. Bos, *A cardinal function algorithm for computing multivariate quadrature points*, SIAM Journal on Numerical Analysis, 45 (2007), pp. 193–205, https://doi.org/10.1137/050625801.

[49] S. Wandzura and H. Xiao, *Symmetric quadrature rules on a triangle*, Computers & Mathematics with Applications. An International Journal, 45 (2003), pp. 1829–1840, https://doi.org/10.1016/S0898-1221(03)90004-6.

[50] C. Wu, J. Zhang, Y. Duan, and X.-C. Tai, *Augmented Lagrangian method for total variation based image restoration and segmentation over triangulated surfaces*, Journal of Scientific Computing, 50 (2012), pp. 145–166, https://doi.org/10.1007/s10915-011-9477-3.

[51] L. Zhang, T. Cui, and H. Liu, *A set of symmetric quadrature rules on triangles and tetrahedra*, Journal of Computational Mathematics, 27 (2009), pp. 89–96.