ITERATIVE TV MINIMIZATION ON THE GRAPH*

JAPHET NIYOBUHUNGIRO†, ERIC SETTERQVIST‡, FREDDIE ÅSTRÖM§, AND
GEORGE BARAVDISH¶

Abstract. We define the space of functions of bounded variation (BV) on the graph. Using the notion of divergence of flows on graphs, we show that the unit ball of the dual space to BV in the graph setting can be described as the image of the unit ball of the space ℓ^∞ by the divergence operator. Based on this result, we propose a new iterative algorithm to find the exact minimizer for the total variation (TV) denoising problem on the graph. The proposed algorithm is provable convergent and its performance on image denoising examples is compared with the Split Bregman and Primal-Dual algorithms as benchmarks for iterative methods and with BM3D as a benchmark for other state-of-the-art denoising methods. The experimental results show highly competitive empirical convergence rate and visual quality for the proposed algorithm.

Keywords. Total variation; ROF model on the graph; Split Bregman; Primal-dual; BM3D.

AMS subject classifications. 65K10; 93B40; 68U10.

1. Introduction

1.1. Background. Removing or reducing the noise from obtained and observed images is a fundamental image processing problem known as *denoising* appearing in many application areas. The image noise η considered here is additive which means that the observed image data u_0 is related to the underlying true image u according to the linear model

$$u_0 = u + \eta. \tag{1.1}$$

The noise component η is further assumed to be normally, independent and identically distributed. We study in this work the total variation (abbreviated as TV) denoising problem. In this problem a noisy image $u_0 \in L^2(\Omega)$, where the open set $\Omega \subset \mathbb{R}^2$ is the image domain, is observed and the denoised image approximating the original image is then defined as the solution u_{opt} of the optimization problem

$$\inf_{u \in BV(\Omega)} \left(\frac{1}{2} \|u_0 - u\|_{L^2(\Omega)}^2 + t \|u\|_{BV(\Omega)} \right), \tag{1.2}$$

where $t > 0$ is called the regularization parameter and $BV(\Omega)$ is the space of functions of bounded variation. The TV denoising model 1.2 was introduced in 1992 by Rudin, Osher and Fatemi [24] and is now also widely known in the image processing community as the ROF model. The space $BV(\Omega)$ is defined as follows

DEFINITION 1.1.

$$BV(\Omega) = \big\{ u \in L^1(\Omega) : \|u\|_{BV} < \infty \big\}$$

where the bounded variation or total variation seminorm of u is given by

DEFINITION 1.2.

$$\|u\|_{BV(\Omega)} = \int_\Omega |Du| = \sup \int_\Omega u(x)\,\mathrm{div}\ \vec{g}(x)dx,$$

where the supremum is taken over all $\vec{g} \in C_c^1\big(\Omega, \mathbb{R}^2\big)$ such that $\sup_{x\in\Omega} \sqrt{g_1^2(x)+g_2^2(x)} \leq 1$.

Note that if u is a differentiable function then $\|u\|_{BV(\Omega)} = \int_\Omega |\nabla u(x)|\,dx$. An important feature of the BV term in the minimization problem 1.2 is that it discourages the solution from having oscillations and at the same time allowing it to have discontinuities.

Since its appearance in 1992, the ROF model has received a large amount of popularity for its effeciency in denoising images without smoothing out the boundaries, and it has also been applied to a multitude of other imaging problems (see for example the book [12]). We choose next to highlight a few selected works from the vast literature on the ROF model and TV minimization which are related to our approach.

An early work on total variation minimization based on dual formulation is [11]. In 2004, Chambolle provided an iterative algorithm related to [11] and proved its convergence, see [7]. We remark that the works [1,9] also proposed efficient projection algorithms for total variation minimization. The papers [16] and [3] adapted Chambolle's algorithm from [7] to handle linear operators in the ROF model, such as convolution operators representing blurring.

After the appearance of [7], several iterative algorithms have been developed which can be used to solve TV minimization problems. Bregman iteration was shown in [19] to be an efficient and fast way to solve TV problems among other L^1-regularized optimization problems. In particular, a *split Bregman* method was proposed in [19] and subsequently used to compute the ROF minimizer. The Primal-Dual algorithm proposed in [10] is another general purpose iterative algorithm which can be efficiently applied to solve TV minimization problems. The fast iterative shrinkage-thresholding algorithm (FISTA) for linear inverse problems, see [4] and [5], is also known to be able to solve TV minimization problems efficiently.

For anisotropic total variation minimization of quantized images, i.e. the pixels of the image take values in a prescribed finite set because the observed image is decomposed into a prescribed number of level sets, graph cut algorithms have been developed that exactly compute the minimizer up to machine precision. Foundational works in this direction are the algorithms of Chambolle [8], Darbon and Sigelle [14] and Goldfarb and Yin [18]. These algorithms are not iterative and in terms of speed, they are very fast.

Based on the fact that an image has a locally sparse representation in transform domain and that this sparsity is enhanced by grouping similar 2D image patches into 3D groups, a paper on collaborative filtering or BM3D grouping and filtering procedure was written [13] and later analysed and implemented in [21]. Though this method is not directly designed to solve the TV minimization problem, it is one of the state-of-the-art filtering methods applicable to the denoising problem.

Image decomposition models into a piecewise-smooth and oscillating components that usually researchers refer to as cartoon and textures (or textures + noise) respectively, have received great interest in the image processing community. For example $u_0 \in L^2(\Omega)$ is decomposable as

$$u_0 = u_{opt} + (u_0 - u_{opt}).$$

This is the decomposition of u_0 into the piecewise-smooth component $u_{opt} \in BV(\Omega)$ satisfying 1.2 and the component $(u_0 - u_{opt}) \in L^2(\Omega)$ which contains textures and noise. The original theoretical model for such an image decomposition was introduced in 2001 by Meyer in [22] by using the total variation to model the piecewise-smooth component and an appropriate dual space G which is the Banach space composed of the distributions $f = \partial_1 g_1 + \partial_2 g_2 = \text{div}\,\vec{g}$, where g_1 and g_2 are in $L^\infty(\Omega)$ and $\|f\|_G = \inf \|\vec{g}\|_{L^\infty(\Omega;\mathbb{R}^2)}$ where the infimum is taken over all $\vec{g}$ such that $f = \text{div}\,\vec{g}$ and $\|\vec{g}\|_{L^\infty(\Omega;\mathbb{R}^2)} = \text{ess sup}_{x \in \Omega} \sqrt{|g_1(x)|^2 + |g_2(x)|^2}$, to model the oscillating component. Some of the works proposed in the literature for numerically solving Meyer's model or its variants include for instance [2] that proposed to split the image into three components, a geometrical component modeled by the total variation, a texture component modeled by a negative Sobolev norm and a noise component modeled by a negative Besov norm. Furthermore, [17] designed an algorithm by using split Bregman iterations and the duality used by Chambolle to find the minimizer of a functional based on Meyer's G-norm. Other works based on the G-norm include for example [25] and [23].

1.2. Summary of main contributions and motivation. We present an iterative method for solving the discrete analogue of the TV minimization problem 1.2 on finite graphs. The algorithmic representation of the method is given in Algorithm 1 and is proved to converge to the exact minimizer. Further, the algorithm can be run on a parallel computer architecture and is thereby suitable to handle large graphs and data sets. The proof of the convergence result Theorem 3.2 is based on duality principles from convex analysis and Theorem 3.1 which characterizes, in the graph setting, the unit ball of the dual space to BV as the image of the unit ball of the space ℓ^∞ by the divergence operator. We note that the strength of a graph representation is when considering non-Euclidean metric spaces via manifold representations, for example when the image is a map in spherical geometry, which could be the case in many applications. Our approach also illustrates the properties of the optimal decomposition of the image data into a piecewise-smooth image component and a noise component and gives its geometrical interpretation. Experimental results confirm that our method is a highly competitive TV denoising algorithm in terms of both convergence rate and visual quality.

1.3. Overview. This paper is organized as follows. In Section 2 we present some needed notation, definitions and simple results from interpolation theory and algebra. Next, in Section 3 the TV minimization problem on the graph is formulated, the proposed algorithm is given and its convergence is proved. Thereafter, in Section 4, we present numerical experiments in order to compare the performance of the proposed algorithm with other iterative TV denoising algorithms and the BM3D image denoising method. Finally, discussion and concluding remarks are given in Section 5.

2. Notation and definitions

In this section we briefly introduce the necessary mathematical theory and notation needed for presentation of the proposed algorithm.

2.1. Interpolation theory. Let X_0 and X_1 be two Banach spaces. They form a Banach couple (X_0, X_1) if there exists a Hausdorff topological vector space $\mathcal{H}$ in which both X_0 and X_1 are linearly and continuously embedded. For an introduction to the theory of interpolation, we refer to the book [6]. When (X_0, X_1) is a Banach couple, then the sum $X_0 + X_1$ given by $X_0 + X_1 = \{x \in \mathcal{H}: \ x = x_0 + x_1, \ x_j \in X_j, \ j = 0, 1\}$ is well defined, and can be shown to be a Banach space under the norm $\|x\|_{X_0 + X_1} = \inf \left(\|x_0\|_{X_0} + \|x_1\|_{X_1}, \ x = x_0 + x_1, \ x_j \in X_j, \ j = 0, 1 \right)$. Furthermore, given a Banach couple (X_0, X_1), an element $u_0 \in X_0 + X_1$ and a positive parameter t the Peetre's K-functional is defined by

$$K(t, u_0; X_0, X_1) = \inf_{u \in X_1} \left(\|u_0 - u\|_{X_0} + t \|u\|_{X_1} \right).$$

The K-functional is very important for the so-called K-method of real interpolation which generates families of real interpolation spaces between X_0 and X_1. The K-functional is a particular case of the more general L-functional which, for given $1 \leq p_0, p_1 < \infty$, is defined by

$$L_{p_0, p_1}(t, u_0; X_0, X_1) = \inf_{u \in X_1} \left(\frac{1}{p_0} \|u_0 - u\|_{X_0}^{p_0} + \frac{t}{p_1} \|u\|_{X_1}^{p_1} \right). \tag{2.1}$$

We need the following definitions of exact minimizers and optimal decomposition.

DEFINITION 2.1 (**Exact minimizers**).
We say that the element $u_{opt} \in X_1$ is an exact minimizer for the functional 2.1 if

$$\frac{1}{p_0} \|u_0 - u_{opt}\|_{X_0}^{p_0} + \frac{t}{p_1} \|u_{opt}\|_{X_1}^{p_1} = L_{p_0, p_1}(t, u_0; X_0, X_1).$$

DEFINITION 2.2 (**Optimal decomposition**). *If $u_{opt} \in X_1$ is an exact minimizer for 2.1, then we call $u_0 = u_{opt} + (u_0 - u_{opt})$ an optimal decomposition for 2.1.*

REMARK 2.1. It is important to note that an exact minimizer, and therefore an optimal decomposition, does not always exist.

The L_{p_0, p_1}-functional appears in regularization of inverse problems where the second term in the expression 2.1 is called a penalty term or regularization term. Note that the total variation regularization functional 1.2 above is a particular case of the L-functional 2.1 for $p_0 = 2$, $p_1 = 1$ and for the spaces $X_0 = L^2(\Omega)$ and $X_1 = BV(\Omega)$.

2.2. Some algebra. We start with the definition of the notion of annihilator.

DEFINITION 2.3. *Let X be a Banach space and let Z be a subspace of X. The annihilator of Z denoted $\mathrm{ann}(Z)$ is the set of bounded linear functionals that vanish on Z. That is the set defined by*

$$\mathrm{ann}(Z) = \{x^* \in X^*: \ \langle x^*, z \rangle = 0, \ for \ all \ z \in Z\},$$

where X^* is the dual space of X and $\langle x^*, z \rangle$ denotes the action of the bounded linear functional $x^* \in X^*$ on the element $z \in Z$.

We will make use of the following result in the sequel.

LEMMA 2.1. *Let X be a Banach space with dual space X^*, $x_0 \in X$ and let Z be a finite-dimensional subspace of X. Then*

$$\inf_{z \in Z} \|x_0 + z\|_X = \sup_{x^* \in \mathcal{B}_{X^*} \cap \mathrm{ann}(Z)} \langle x^*, x_0 \rangle,$$

where $\mathcal{B}_{X^*}$ is the unit ball of X^*.

Proof. The case $x_0 \in Z$ is obvious. From now on we suppose $x_0 \notin Z$. Let us take an arbitrary $x^* \in \mathcal{B}_{X^*} \cap \mathrm{ann}(Z)$. Then we have

$$\langle x^*, x_0 \rangle = \langle x^*, x_0 + z \rangle \leq \|x^*\|_{X^*} \|x_0 + z\|_X \leq \|x_0 + z\|_X, \ \forall z \in Z.$$

Therefore since $x^* \in \mathcal{B}_{X^*} \cap \mathrm{ann}(Z)$ and $z \in Z$ are arbitrary, we have that

$$\inf_{z \in Z} \|x_0 + z\|_X \geq \sup_{x^* \in \mathcal{B}_{X^*} \cap \mathrm{ann}(Z)} \langle x^*, x_0 \rangle. \tag{2.2}$$

In order to prove the reverse inequality, let us consider the space W, which is the algebraic sum of the span of x_0 and the space Z:

$$W = \{x_0\} + Z = \{w \in X : w = \lambda x_0 + z, \ z \in Z \text{ and } \lambda \in \mathbb{R}\},$$

and take $z_0 \in Z$ such that $\inf_{z \in Z} \|x_0 + z\|_X = \|x_0 + z_0\|_X$. The existence of such z_0 follows from the assumption that Z is a finite-dimensional subspace of X. Without loss of generality we can assume that $\|x_0 + z_0\|_X = 1$. Since W is a normed vector space, it is possible to consider its dual space. Further, as Z is a linear subspace of W $x_0 + z_0 \in W$ and $x_0 + z_0 \notin Z$, the Hahn-Banach Theorem (see for example Corollary II.3.13 in [15]) gives that there exists a bounded linear functional $x_0^* \in W^*$ such that

$$\langle x_0^*, z \rangle = 0 \text{ for all } z \in Z, \text{and } \langle x_0^*, x_0 + z_0 \rangle = 1.$$

It follows that

$$x_0^* \in \mathrm{ann}(Z) \text{ and } \langle x_0^*, x_0 \rangle = 1. \tag{2.3}$$

Let us now investigate the action of x_0^* on W. Let $w = \lambda x_0 + z$ be an element of W for some $\lambda \in \mathbb{R}$ and $z \in Z$. Then we have

$$\begin{aligned} \langle x_0^*, w \rangle &= \langle x_0^*, \lambda x_0 + z \rangle = \langle x_0^*, \lambda x_0 + \lambda z_0 + z - \lambda z_0 \rangle \\ &= \lambda \langle x_0^*, x_0 + z_0 \rangle + \langle x_0^*, z - \lambda z_0 \rangle = \lambda, \end{aligned} \tag{2.4}$$

because $\langle x_0^*, x_0 + z_0 \rangle = 1$ and $\langle x_0^*, z - \lambda z_0 \rangle = 0$ since $x_0^* \in \mathrm{ann}(Z)$ and $z - \lambda z_0 \in Z$. Let us now describe the unit ball $\mathcal{B}_W$ of W. Suppose that $w = \lambda x_0 + z \in \mathcal{B}_W$ where $\lambda \neq 0$. We have that

$$1 \geq \|w\|_X = \|\lambda x_0 + z\|_X = |\lambda| \left\| x_0 + \frac{z}{\lambda} \right\|_X \geq |\lambda| \|x_0 + z_0\|_X = |\lambda|. \tag{2.5}$$

Therefore $w = \lambda x_0 + z \in \mathcal{B}_W$ implies that $|\lambda| \leq 1$. From 2.4 and 2.5, it follows that

$$\|x_0^*\|_{W^*} = \sup_{w \in \mathcal{B}_W} \langle x_0^*, w \rangle = \sup_{w \in \mathcal{B}_W} \lambda = \sup_{|\lambda| \leq 1} \lambda = 1.$$

By invoking the Hahn-Banach theorem (see for example Theorem II.3.11 in [15]), we can extend the functional x_0^* to a functional $\widetilde{x_0^*} \in X^*$ such that $\widetilde{x_0^*}|_W = x_0^*$ and $\left\| \widetilde{x_0^*} \right\|_{X^*} = \|x_0^*\|_{W^*} = 1$. From this and 2.3 we conclude that $\widetilde{x_0^*} \in \mathcal{B}_{X^*} \cap \mathrm{ann}(Z)$. It follows that

$$\inf_{z \in Z} \|x_0 + z\|_X = \|x_0 + z_0\|_X = \langle x_0^*, x_0 + z_0 \rangle = \langle \widetilde{x_0^*}, x_0 + z_0 \rangle$$

$$\leq \sup_{x^* \in \mathcal{B}_{X^*} \cap \mathrm{ann}(Z)} \langle x^*, x_0 + z_0 \rangle = \sup_{x^* \in \mathcal{B}_{X^*} \cap \mathrm{ann}(Z)} \langle x^*, x_0 \rangle. \tag{2.6}$$

Putting 2.2 and 2.6 together, we obtain

$$\inf_{z \in Z} \|x_0 + z\|_X = \sup_{x^* \in \mathcal{B}_{X^*} \cap \mathrm{ann}(Z)} \langle x^*, x_0 \rangle$$

which concludes the proof. $\qquad\qquad\square$

3. Introducing iterative TV minimization on the graph

3.1. A graph specific problem formulation.   Suppose we have an observed noisy image $u_0 \in L^2$ defined on the domain $\Omega = (0,1)^2 \subset \mathbb{R}^2$ which is a degraded version of the original true image $u \in BV(\Omega)$ according to the linear model 1.1. The ROF model suggests to take as an approximation to the original image u the function $u_{opt} \in BV$ which is the exact minimizer for the $L_{2,1}$-functional of the couple (L^2, BV):

$$L_{2,1}\left(t, u_0; L^2, BV\right) = \inf_{u \in BV} \left(\frac{1}{2} \|u_0 - u\|_{L^2}^2 + t \|u\|_{BV} \right), \text{ for some } t > 0. \qquad (3.1)$$

We will use the following anisotropic BV seminorm:

$$\|u\|_{BV(\Omega)} = \int_0^1 \text{var}_x\, u(y)\, dy + \int_0^1 \text{var}_y\, u(x)\, dx,$$

where

$$\text{var}_x\, u(y) = \sup_{0 \le x_1 \le \ldots \le x_n \le 1} \sum_{j=1}^{n-1} |u(x_{j+1}, y) - u(x_j, y)|$$

is the total variation of u along the horizontal axis for a given y, and

$$\text{var}_y\, u(x) = \sup_{0 \le y_1 \le \ldots \le y_n \le 1} \sum_{i=1}^{n-1} |u(x, y_{i+1}) - u(x, y_i)|$$

is the total variation of u along the vertical axis for a given x. The reason for choosing this BV seminorm is that it suggests a convenient formulation of total variation in the graph setting, see 3.7a and 3.7b.

We use a standard approach when discretizing the functional 3.1, i.e., we divide Ω into $N \times N$ square cells and, instead of the space $L^2(\Omega)$, consider its finite-dimensional subspace S_N consisting of functions that are piecewise constant on each cell. Throughout, we consider our discretization grid as a 2D Cartesian coordinate in *screen space*, i.e., the same way matrices are represented on the computer. We define

$$S_N = \left\{ u = \sum_{i,j=1}^N u_{ij} \chi_{ij}, \ \chi_{ij}(x,y) = \begin{cases} 1 \text{ if } \frac{j-1}{N} < x < \frac{j}{N} \text{ and } \frac{i-1}{N} < y < \frac{i}{N} \\ 0 \text{ otherwise.} \end{cases} \right\}.$$

It is clear that the BV seminorm of a function $u \in S_N$ is equal to

$$\|u\|_{BV(S_N)} = \frac{1}{N} \left(\sum_{i=1}^N \sum_{j=1}^{N-1} |u_{i,j+1} - u_{ij}| + \sum_{j=1}^N \sum_{i=1}^{N-1} |u_{i+1,j} - u_{ij}| \right).$$

Therefore the discrete analogue of the functional 3.1 can be written as

$$L_{2,1}\left(t, u_0; L^2, BV\right) = \inf_{u \in S_N} \left(\frac{1}{2N^2} \left(\sum_{i,j=1}^N \left(u_{0_{ij}} - u_{ij} \right)^2 \right) \right.$$

$$\left. + \frac{t}{N} \left(\sum_{i=1}^N \sum_{j=1}^{N-1} |u_{i,j+1} - u_{ij}| + \sum_{j=1}^N \sum_{i=1}^{N-1} |u_{i+1,j} - u_{ij}| \right) \right). \qquad (3.2)$$

3.1.1. Graph notations.   We now turn to the framework of graph which generalizes the problem 3.2. Let $G = (V, E)$ be a finite, directed and connected[1] graph with N vertices $V = \{v_1, v_2, \ldots, v_N\}$ and M directed edges $E = \{e_1, e_2, \ldots, e_M\}$ where each edge is determined by a pair of vertices, i.e. $e_k = (v_i, v_j)$ for some $i, j \in \{1, 2, \ldots, N\}$ and $k = 1, 2, \ldots, M$. We assume that the edge $e_k = (v_i, v_j)$ is directed from the vertex v_i to the vertex v_j. Let $S_V = \{f \colon f : V \to \mathbb{R}\}$ denote the N-dimensional space of real-valued functions defined on V and let $S_E = \{g \colon g : E \to \mathbb{R}\}$ denote the M-dimensional space of real-valued functions defined on E.

The *gradient* operator $\mathrm{grad} : S_V \to S_E$ is defined by

$$\mathrm{grad}\, f(e) = f(v_j) - f(v_i), e = (v_i, v_j) \in E.$$

We define inner products on S_E and S_V according to

$$\langle f_1, f_2 \rangle_{S_E} = \sum_{e \in E} f_1(e) f_2(e)$$

and

$$\langle g_1, g_2 \rangle_{S_V} = \sum_{v \in V} g_1(v) g_2(v).$$

It is easy to show that the *divergence* operator $\mathrm{div} : S_E \to S_V$ given by

$$\mathrm{div}\, g(v_j) = \sum_{i : (v_i, v_j) \in E} g((v_i, v_j)) - \sum_{k : (v_j, v_k) \in E} g((v_j, v_k)).$$

is conjugate to grad, i.e.

$$\langle \mathrm{div}\, g, f \rangle_{S_V} = \langle g, \mathrm{grad}\, f \rangle_{S_E}, \ \forall f \in S_V, \forall g \in S_E. \tag{3.3}$$

If we consider elements of S_E as flows on the graph $G = (V, E)$, the divergence at a vertex can be interpreted as the difference between the total incoming flows and the total outgoing flows.

The graph in Figure 3.1 illustrates the definition for the gradient and the divergence operator by an example for the case $N = 6$ and $M = 10$. For example, $\mathrm{grad}\, f(e_6) = f(v_5) - f(v_2)$ is the gradient at $e_6 = (v_2, v_5)$ and $\mathrm{div}\, g(v_3) = g((v_1, v_3)) + g((v_4, v_3)) + g((v_6, v_3)) - g((v_3, v_2)) - g((v_3, v_5))$ is the divergence at v_3.

REMARK 3.1.   The operator grad has a kernel given by

$$\ker(\mathrm{grad}) = \{f \in S_V : f = C, \text{ for some } C \in \mathbb{R}\},$$

and its orthogonal complement coincides with its annihilator and is given by

$$(\ker(\mathrm{grad}))^\perp = \mathrm{ann}(\ker(\mathrm{grad})) = \left\{ F \in S_V : \sum_{v \in V} F(v) = 0 \right\}. \tag{3.4}$$

Since div is the conjugate operator of grad, the fundamental theorem of linear algebra ensures that

$$\mathrm{im}(\mathrm{div}) = \mathrm{ann}(\ker(\mathrm{grad})) \text{ and } \mathrm{im}(\mathrm{grad}) = \mathrm{ann}(\ker(\mathrm{div})) \tag{3.5}$$

---

[1]The analysis can be extended to disconnected graphs as the components of the graph are considered separately. For convenience, we have therefore chosen to only consider connected graphs in this paper.
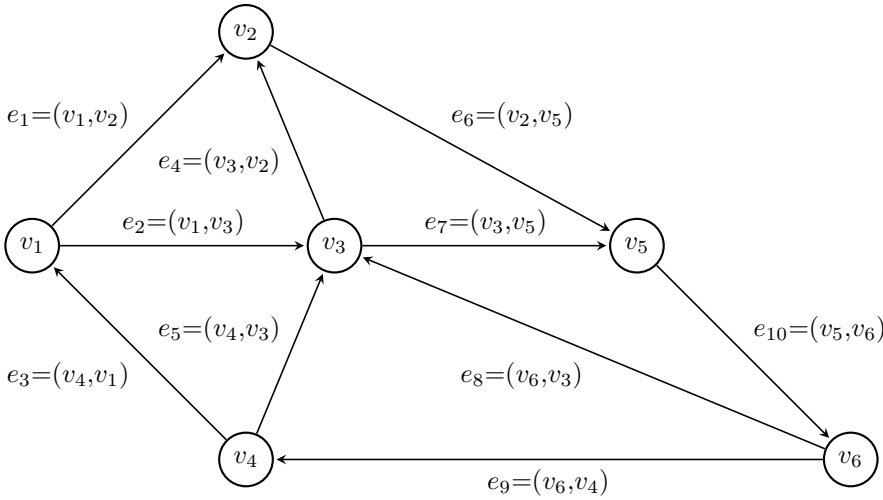
Fig. 3.1: *Graph illustrating the notation and node relations for the discrete gradient and divergence operators.*

where $\text{im}(A)$ denotes the image of the operator A.

An observed image $u_0 \in S_N$ can be considered as an element of S_V for a graph $G = (V, E)$ where the cells are represented by the vertices in V and pairs of adjacent cells are represented by edges in E (any direction of the edges can be chosen). The functional 3.2 can then be written as

$$L_{2,1}\left(t, u_0; L^2, BV\right) = \inf_{u \in S_V} \left(\frac{1}{2N^2} \|u_0 - u\|^2_{\ell^2(S_V)} + \frac{t}{N} \|\text{grad}\, u\|_{\ell^1(S_E)} \right). \qquad (3.6)$$

It is clear that the exact minimizer of 3.6 coincides with the exact minimizer of

$$L_{2,1}\left(s, u_0; \ell^2(S_V), BV(S_V)\right) = \inf_{u \in S_V} \left(\frac{1}{2} \|u_0 - u\|^2_{\ell^2(S_V)} + s \|\text{grad}\, u\|_{\ell^1(S_E)} \right), \quad s = Nt.$$

This observation leads to the following analogue of the ROF model on a general finite, connected and directed graph.

PROBLEM 3.1.

Suppose that we know the function $u_0 \in S_V$. For given $t > 0$, find the exact minimizer of the functional

$$L_{2,1}\left(t, u_0; \ell^2(S_V), BV(S_V)\right) = \inf_{u \in BV(S_V)} \left(\frac{1}{2} \|u_0 - u\|^2_{\ell^2(S_V)} + t \|u\|_{BV(S_V)} \right),$$

where

$$\|u\|_{\ell^2(S_V)} = \left(\sum_{v \in V} (u(v))^2 \right)^{\frac{1}{2}}, \quad \|u\|_{BV(S_V)} = \|\text{grad}\, u\|_{\ell^1(S_E)}, \qquad (3.7a)$$

$$and \qquad \|\psi\|_{\ell^1(S_E)} = \sum_{e \in E} |\psi(e)|. \qquad (3.7b)$$

3.1.2. Description of the ball of dual space to $BV(S_V)$.   In order to describe our algorithm for Problem 3.1, we first need a description of the ball of the dual space to $BV(S_V)$.

It was shown in [20] that the exact minimizer u_{opt} for the $L_{2,1}$- functional for the couple (ℓ^2, X), where X is a Banach space,

$$L_{2,1}\left(t, u_0; \ell^2, X\right) = \inf_{u \in X} \left(\frac{1}{2} \|u_0 - u\|_{\ell^2}^2 + t \|u\|_X\right),$$

is equal to the difference between u_0 and the nearest element to u_0 of the ball of radius $t > 0$ of the space X^*, i.e., $u_{opt} = u_0 - \underset{\psi \in t\mathcal{B}_{X^*}}{\arg\min} \|u_0 - \psi\|_{\ell^2}$. Figure 3.2 provides a geometrical illustration of the optimal decomposition.
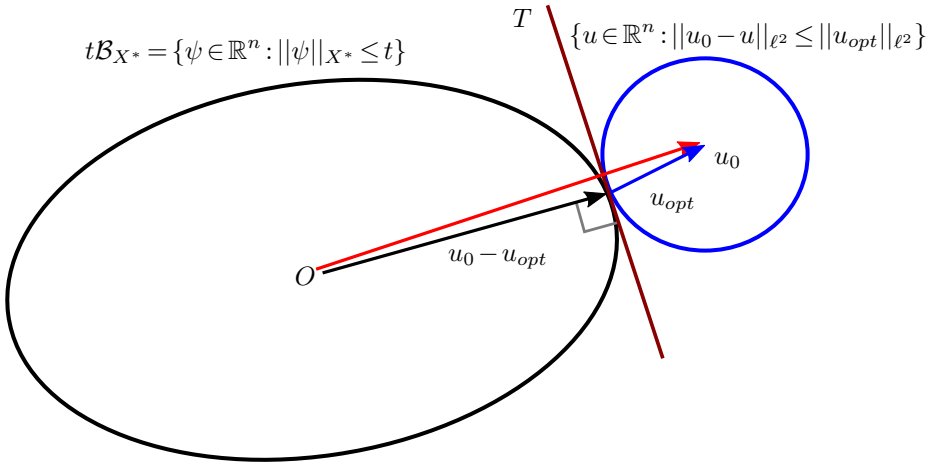

Fig. 3.2: *Geometrical illustration of the ball of dual space and the position of the element of best approximation $u_{opt} = u_0 - \underset{\psi \in t\mathcal{B}_{X^*}}{\arg\min} \|u_0 - \psi\|_{\ell^2}$. The hyperplane T is orthogonal to $u_0 - u_{opt}$.*

Consider now $X = BV$. As $\|\cdot\|_{BV(S_V)}$ is a seminorm on S_V, we restrict to the subspace $(\ker(\mathrm{grad}))^\perp$ where $\|\cdot\|_{BV(S_V)}$ is a norm. The dual space $BV^*(S_V)$ is then $(\ker(\mathrm{grad}))^\perp$ equipped with the norm defined by

$$\|\psi\|_{BV^*(S_V)} = \sup_{\|h\|_{BV(S_V)} \leq 1} \langle \psi, h \rangle_{S_V}. \tag{3.8}$$

We have the following characterization of the unit ball of $BV^*(S_V)$:

THEOREM 3.1.   *The unit ball of the space $BV^*(S_V)$ is equal to the image of the unit ball of the space $\ell^\infty(S_E)$ under the operator* div, *i.e.,*

$$\mathcal{B}_{BV^*(S_V)} = \mathrm{div}\left(\mathcal{B}_{\ell^\infty(S_E)}\right).$$

Proof.   Let us consider an arbitrary $\psi \in BV^*(S_V)$. From relations 3.4 and 3.5, we conclude that $BV^*(S_V) = \mathrm{im}(\mathrm{div})$. Therefore for all $\psi \in BV^*(S_V)$, there exists at least one $g \in S_E$ such that $\psi = \mathrm{div}\, g$. Fix $g_0 \in S_E$ such that $\psi = \mathrm{div}\, g_0$. We have

$$\inf_{\psi = \mathrm{div}\, g} \|g\|_{\ell^\infty(S_E)} = \inf_{\varphi \in \ker(\mathrm{div})} \|g_0 + \varphi\|_{\ell^\infty(S_E)}. \tag{3.9}$$

By applying Lemma 2.1 and using expression 3.5, together with 3.3, $\psi = \operatorname{div} g_0$, 3.7a and 3.8 we derive

$$\inf_{\psi = \operatorname{div} g} \|g\|_{\ell^\infty(S_E)} = \inf_{\varphi \in \ker(\operatorname{div})} \|g_0 + \varphi\|_{\ell^\infty(S_E)} = \sup_{f \in \mathcal{B}_{\ell^1(S_E)} \cap \operatorname{im}(\operatorname{grad})} \langle f, g_0 \rangle_{S_E}$$

$$= \sup_{\|\operatorname{grad} h\|_{\ell^1(S_E)} \leq 1} \langle \operatorname{grad} h, g_0 \rangle_{S_E} = \sup_{\|\operatorname{grad} h\|_{\ell^1(S_E)} \leq 1} \langle h, \operatorname{div} g_0 \rangle_{S_V}$$

$$= \sup_{\|h\|_{BV(S_V)} \leq 1} \langle h, \psi \rangle_{S_V} = \|\psi\|_{BV^*(S_V)}.$$

From this it follows that

$$\|\psi\|_{BV^*(S_V)} \leq 1 \text{ if and only if } \inf_{\psi = \operatorname{div} g} \|g\|_{\ell^\infty(S_E)} \leq 1.$$

So, it is clear that $BV^*(S_V) \supset \operatorname{div}\big(\mathcal{B}_{\ell^\infty(S_E)}\big)$. Note next that the infimum in 3.9 is attained because $\ker(\operatorname{div})$ is a subspace of the finite-dimensional space S_E. Therefore, for each $\psi \in \mathcal{B}_{BV^*(S_V)}$ there exists an element

$$g_\psi \in g_0 + \ker(\operatorname{div}), \text{ such that } \|g_\psi\|_{\ell^\infty(S_E)} \leq 1 \text{ and } \operatorname{div} g_\psi = \psi.$$

We conclude that

$$\mathcal{B}_{BV^*(S_V)} = \operatorname{div}\big(\mathcal{B}_{\ell^\infty(S_E)}\big).$$

$\square$

3.2. Algorithm. Algorithm 1 below embodies our algorithmic contribution for computing the ROF-minimizer u_{opt} and we will now describe its construction in detail.

The core of the algorithm is the construction of the element $\widetilde{\psi} = (u_0 - u_{opt}) \in t\mathcal{B}_{BV^*(S_V)}$ that satisfies

$$\left\|u_0 - \widetilde{\psi}\right\|_{\ell^2(S_V)} = \inf_{\psi \in t\mathcal{B}_{BV^*(S_V)}} \|u_0 - \psi\|_{\ell^2(S_V)}.$$

From Theorem 3.1, this is equivalent to constructing a flow $g_{\widetilde{\psi}} \in t\mathcal{B}_{\ell^\infty(S_E)}$ such that

$$\left\|u_0 - \operatorname{div} g_{\widetilde{\psi}}\right\|_{\ell^2(S_V)} = \inf_{g \in t\mathcal{B}_{\ell^\infty(S_E)}} \|u_0 - \operatorname{div} g\|_{\ell^2(S_V)},$$

and put $\widetilde{\psi} = \operatorname{div} g_{\widetilde{\psi}}$. Once this is done, $u_{opt} = u_0 - \widetilde{\psi}$.

We now describe the steps of the algorithm in detail. Let u_0 be defined on $G = (V, E)$ with vertex set $V = \{v_1, \ldots, v_N\}$ and edge set $E = \{e_1, \ldots, e_M\}$. The parameter t denotes a regularization parameter and N_{iter} denotes the maximum number of iterations. The edge set is specifically defined as

$$e_k = (v_i, v_j) \in E, \ k = 1, 2, \ldots, M; \text{ for some } i, j \in \{1, 2, \ldots, N\}.$$

Introduce the operator $T : t\mathcal{B}_{\ell^\infty(S_E)} \to t\mathcal{B}_{\ell^\infty(S_E)}$ given by

$$T = T_M T_{M-1} T_{M-2} \ldots T_2 T_1 \tag{3.10}$$

where for $k=1,2,\dots,M$, the operator $T_k: t\mathcal{B}_{\ell^\infty(S_E)} \to t\mathcal{B}_{\ell^\infty(S_E)}$ is defined as follows

$$T_k g(e) = \begin{cases} \begin{cases} Kg(e_k) & \text{,if } Kg(e_k) \in [-t,+t]; \\ -t & \text{,if } Kg(e_k) < -t; \\ +t & \text{,if } Kg(e_k) > +t. \end{cases} & \text{, if } e = e_k; \\ g(e) & \text{, if } e \neq e_k. \end{cases} \qquad (3.11)$$

Here

$$Kg(e_k) = \frac{[u_0(v_j) - \mathrm{div}\,\backslash_{e_k} g(v_j)] - [u_0(v_i) - \mathrm{div}\,\backslash_{e_k} g(v_i)]}{2}$$

and

$$\begin{cases} \mathrm{div}_{\backslash e_k}\, g(v_i) = \mathrm{div}\, g(v_i) + g(e_k); \\ \mathrm{div}_{\backslash e_k}\, g(v_j) = \mathrm{div}\, g(v_j) - g(e_k); \\ \mathrm{div}_{\backslash e_k}\, g(v_\ell) = \mathrm{div}\, g(v_\ell),\ \forall \ell \neq i,j, \end{cases}$$

i.e. $\mathrm{div}_{\backslash e_k}$ is the divergence operator div without taking into account the flow on the edge e_k.

The constructed operator T depends on the enumeration of the edges in E. However the results concerning T, i.e. Proposition 3.1 and Theorem 3.2 below, hold regardless of the specific enumeration of the edges. We will now point out a certain construction of T which leads to a version of Algorithm 1 suitable for parallel computer architectures.

Colour the set of edges E such that incident edges, i.e. edges that share a common vertex, have different colours. Denote by $E_1,\dots,E_L$ the resulting disjoint subsets of E, $\cup_{i\in\{1,\dots,L\}} E_i = E$, from such a colouring with usage of L different colours. Let $e_{i,1},\dots,e_{i,M_i}$ denote the edges of E_i and define $T_{E_i} = T_{i,M_i} T_{i,M_i-1} \dots T_{i,1}$ where

$$T_{i,k} g(e) = \begin{cases} \begin{cases} Kg(e_{i,k}) & \text{,if } Kg(e_{i,k}) \in [-t,+t]; \\ -t & \text{,if } Kg(e_{i,k}) < -t; \\ +t & \text{,if } Kg(e_{i,k}) > +t. \end{cases} & \text{, if } e = e_{i,k}; \\ g(e) & \text{, if } e \neq e_{i,k}. \end{cases}$$

Because the edges of E_i are non-incident, it follows that the applications of $T_{i,k}$, $k = 1,\dots,M_i$, can be done in arbitrary order without affecting the resulting update $T_{E_i} g$ of g. The associated computations can therefore be done in parallel. With T_{E_i}, $i=1,\dots,L$, given, the operator T is then constructed according to

$$T = T_{E_L} T_{E_{L-1}} \dots T_{E_1}.$$

3.3. Convergence results.   For Algorithm 1, convergence is established in Theorem 3.2. Its proof is based on the following proposition.

PROPOSITION 3.1.   *The operator $T: t\mathcal{B}_{\ell^\infty(S_E)} \to t\mathcal{B}_{\ell^\infty(S_E)}$ given by 3.10-3.11 is continuous and satisfies the following two conditions*

(1) For any $g \in t\mathcal{B}_{\ell^\infty(S_E)}$, $\mathrm{div}\, g = \widetilde{\psi}$ if and only if $Tg = g$;

(2) For any $g \in t\mathcal{B}_{\ell^\infty(S_E)}$, if $\mathrm{div}\, g \neq \widetilde{\psi}$ then $\|u_0 - \mathrm{div}\,(Tg)\|_{\ell^2(S_V)} < \|u_0 - \mathrm{div}\, g\|_{\ell^2(S_V)}$.

Proof.   Each operator T_k is continuous because by definition, it is clear that small changes of $g \in t\mathcal{B}_{\ell^\infty(S_E)}$ leads to small changes of T_k and therefore T is continuous as a product of continuous operators.

Algorithm 1 :ROF model on the graph

$n \leftarrow 0$
choose initial $g \in t\mathcal{B}_{\ell^\infty(S_E)}$
$g_n \leftarrow g$
while $n < N_{iter}$ **do**
  if $Tg_n = g_n$ **then**
    stop
  else $\{Tg_n \neq g_n\}$
    Compute $g_{n+1} = Tg_n$
  end if
  $n = n+1$
end while
Compute $\widetilde{\psi} = \text{div}(g_n)$
return $u_{opt} = u_0 - \widetilde{\psi}$

We now prove condition (1). Let $g \in t\mathcal{B}_{\ell^\infty(S_E)}$ and assume that $\text{div}\, g = \widetilde{\psi}$. Take $e_k = (v_i, v_j) \in E$. We note that $u(e_k)$ appears only in the following two terms of $\|u_0 - \text{div}\, g\|^2_{\ell^2(S_V)}$:

$$[u_0(v_j) - \text{div}\, g(v_j)]^2 + [u_0(v_i) - \text{div}\, g(v_i)]^2$$
$$= \left[u_0(v_j) - \left(\text{div}_{\backslash e_k} g(v_j) + g(e_k)\right)\right]^2 + \left[u_0(v_i) - \left(\text{div}_{\backslash e_k} g(v_i) - g(e_k)\right)\right]^2. \qquad (3.12)$$

Since $\text{div}\, g = \widetilde{\psi}$, $g(e_k)$ in particular must minimize 3.12 in the interval $[-t, t]$. By Jensen's inequality we note that

$$\xi(g(e_k)) = \left[u_0(v_j) - \left(\text{div}_{\backslash e_k} g(v_j) + g(e_k)\right)\right]^2 + \left[u_0(v_i) - \left(\text{div}_{\backslash e_k} g(v_i) - g(e_k)\right)\right]^2$$
$$\geq 2\left(\frac{[u_0(v_j) - \text{div}_{\backslash e_k} g(v_j)] + [u_0(v_i) - \text{div}_{\backslash e_k} g(v_i)]}{2}\right)^2.$$

Equality holds if and only if

$$u_0(v_j) - \left(\text{div}_{\backslash e_k} g(v_j) + g(e_k)\right) = u_0(v_i) - \left(\text{div}_{\backslash e_k} g(v_i) - g(e_k)\right),$$

or equivalently

$$g(e_k) = \frac{[u_0(v_j) - \text{div}_{\backslash e_k} g(v_j)] - [u_0(v_i) - \text{div}_{\backslash e_k} g(v_i)]}{2} =: Kg(e_k).$$

Moreover, $\xi(x)$ is strictly convex and therefore strictly decreasing for $x < Kg(e_k)$ and strictly increasing for $x > Kg(e_k)$. So the minimal value of $\xi(x)$ on the interval $[-t, t]$ is only attained at
    (i) the point $Kg(e_k)$ if $Kg(e_k) \in [-t, t]$,
    (ii) the point $-t$ if $Kg(e_k) < -t$,
    (iii) the point t if $Kg(e_k) > t$.

The assumption $\text{div}\, g = \widetilde{\psi}$ then implies that $g(e_k)$ must be the nearest point in the interval $[-t, t]$ to $Kg(e_k)$, implying that $T_k g(e_k) = g(e_k)$. Since $e_k \in E$ was arbitrary, it follows that $T_k g(e_k) = g(e_k)$ for all $k = 1, ..., M$. Therefore $T_k g = g$ for all $k = 1, ..., M$ and we conclude that $Tg = g$.

Conversely, let us assume that $g \in t\mathcal{B}_{\ell^\infty(S_E)}$ and $Tg = g$. Then for any edge $e \in E$, $g(e)$ coincides with the point of the interval $[-t,t]$ which is nearest to $Kg(e)$. As $\|u_0 - \text{div}(\cdot)\|_{\ell^2(S_V)}$ is a convex function on $t\mathcal{B}_{\ell^\infty(S_E)}$, it is enough to show that g minimizes $\|u_0 - \text{div}(\cdot)\|_{\ell^2(S_V)}$ locally, i.e. it is enough to show that for some small $\varepsilon > 0$ we have

$$\|u_0 - \text{div}\, g\|_{\ell^2(S_V)} = \inf_{\omega \in D_\varepsilon} \|u_0 - \text{div}\, \omega\|_{\ell^2(S_V)},$$

where D_ε is the tubular set given by $D_\varepsilon = \left\{ \omega \in t\mathcal{B}_{\ell^\infty(S_E)} : \|g - \omega\|_{\ell^\infty(S_E)} \le \varepsilon \right\}$. Note that for any $\omega \in D_\varepsilon$ and $e \in E$ we have $\omega(e) \in [-t,t] \cap [g(e) - \varepsilon, g(e) + \varepsilon]$. The set D_ε is a compact subset of S_E and it therefore contains a function $\omega_\varepsilon \in D_\varepsilon$ such that

$$\|u_0 - \text{div}\, \omega_\varepsilon\|_{\ell^2(S_V)} = \inf_{\omega \in D_\varepsilon} \|u_0 - \text{div}\, \omega\|_{\ell^2(S_V)}. \tag{3.13}$$

So, we will need to prove that

$$\|u_0 - \text{div}\, g\|_{\ell^2(S_V)} = \|u_0 - \text{div}\, \omega_\varepsilon\|_{\ell^2(S_V)}.$$

We first note that it follows from the necessity direction proved above that for any edge $e \in E$, $\omega_\varepsilon(e)$ will coincide with the point of the interval $[-t,t] \cap [g(e) - \varepsilon, g(e) + \varepsilon]$, which is nearest to $K\omega_\varepsilon(e)$.

Let us now decompose the edge set E into two parts. The first part denoted by Ω_g consists of the edges for which $Kg(e)$ does not belong to the interval $[-t,t]$, i.e. $\Omega_g = \{e \in E : Kg(e) \notin [-t,t]\}$. As $g(e)$ is the nearest point in the interval $[-t,t]$ to $Kg(e)$ we have

$$g(e) = \begin{cases} -t \text{ if } Kg(e) < -t \\ +t \text{ if } Kg(e) > +t \end{cases} \text{ for edges } e \in \Omega_u.$$

If the number $\varepsilon > 0$ is small enough, it follows from $\|g - \omega\|_{\ell^\infty(S_E)} \le \varepsilon$ that on $e \in \Omega_g$ where we have $Kg(e) < -t$ we will also have $K\omega_\varepsilon(e) < -t$ and therefore $\omega_\varepsilon(e) = -t = g(e)$. Analogously, on $e \in \Omega_g$ where $Kg(e) > t$ we will have $K\omega_\varepsilon(e) > t$ and therefore $\omega_\varepsilon(e) = t = g(e)$. So we have

$$\omega_\varepsilon(e) = g(e) \text{ for all } e \in \Omega_g. \tag{3.14}$$

Next, we consider the remaining edges $E \backslash \Omega_g$. Let $G' = (V, E \backslash \Omega_g)$, i.e. the graph G with the edges in Ω_g removed. The graph G' is the union of several connected components (V_k, E_k), $k = 1, \dots, \ell$ so that we have $V_1 \cup \dots \cup V_\ell = V$ and $E_1 \cup \dots \cup E_\ell = E \backslash \Omega_g$. Note that it is possible that some of the graphs (V_k, E_k) consist of just one single vertex. For these graphs there is nothing to prove because $E_k = \emptyset$. Let us now consider a subgraph (V_k, E_k) where $E_k \ne \emptyset$. On each $e \in E_k$ we have $Kg(e) \in [-t,t]$ and therefore $g(e) = Kg(e)$, i.e. if $e = (v_i, v_j)$ then

$$g(e) = Kg(e) = \frac{\left[u_0(v_j) - \text{div}_{\backslash e}\, u(v_j)\right] - \left[u_0(v_i) - \text{div}_{\backslash e}\, u(v_i)\right]}{2},$$

or equivalently, in view of the definition of $\text{div}_{\backslash e}\, g(\cdot)$, we get that

$$u_0(v_j) - \text{div}\, g(v_j) = u_0(v_i) - \text{div}\, g(v_i).$$

Note that operators K, div and $\mathrm{div}_{\backslash u(e)}$ are considered in the original setting of $G = (V, E)$. Therefore, for all $v \in V_k$ the values of $u_0(v) - \mathrm{div}\, g(v)$ are equal. It follows that

$$\sum_{v \in V_k} [u_0(v) - \mathrm{div}\, g(v)]^2 = |V_k| \left(\frac{\sum_{v \in V_k} [u_0(v) - \mathrm{div}\, g(v)]}{|V_k|} \right)^2. \tag{3.15}$$

For ω_ε we can with Jensen's inequality derive the corresponding inequality

$$\sum_{v \in V_k} [u_0(v) - \mathrm{div}\, \omega_\varepsilon(v)]^2 \geq |V_k| \left(\frac{\sum_{v \in V_k} [u_0(v) - \mathrm{div}\, \omega_\varepsilon(v)]}{|V_k|} \right)^2. \tag{3.16}$$

Now, note that flows on edges in E_k are canceled in the sums $\sum_{v \in V_k} [u_0(v) - \mathrm{div}\, g(v)]$ and $\sum_{v \in V_k} [u_0(v) - \mathrm{div}\, \omega_\varepsilon(v)]$. Therefore, only flows on edges in Ω_g remain in these sums. It then follows from (3.14) that

$$\sum_{v \in V_k} [u_0(v) - \mathrm{div}\, \omega_\varepsilon(v)] = \sum_{v \in V_k} [u_0(v) - \mathrm{div}\, g(v)].$$

Therefore, taking into account (3.15) and (3.16), we obtain

$$\sum_{v \in V_k} [u_0(v) - \mathrm{div}\, \omega_\varepsilon(v)]^2 \geq \sum_{v \in V_k} [u_0(v) - \mathrm{div}\, g(v)]^2.$$

Summing over all V_k gives

$$\|u_0 - \mathrm{div}\, \omega_\varepsilon\|_{\ell^2(S_V)}^2 \geq \|u_0 - \mathrm{div}\, g\|_{\ell^2(S_V)}^2,$$

and we conclude from the definition of ω_ε, recall (3.13), that

$$\|u_0 - \mathrm{div}\, \omega_\varepsilon\|_{\ell^2(S_V)}^2 = \|u_0 - \mathrm{div}\, g\|_{\ell^2(S_V)}^2.$$

So, g minimizes $\|u_0 - \mathrm{div}(\cdot)\|_{\ell^2(S_V)}$ on D_ε and therefore, by convexity, on $t\mathcal{B}_{\ell^\infty(S_E)}$. Therefore, $Tg = g$ implies $\mathrm{div}\, g = \widetilde{\psi}$ and we have now established condition (1).

Finally, we prove condition (2). Note that by definition for $\forall g \in t\mathcal{B}_{\ell^\infty(S_E)}$, the operators T_k, $k = 1, ..., M$ satisfy

$$\|u_0 - \mathrm{div}(T_k u)\|_{\ell^2(S_V)} \leq \|u_0 - \mathrm{div}\, u\|_{\ell^2(S_V)},$$

with equality if and only if $T_k g(e_k) = g(e_k)$. This implies that

$$\|u_0 - \mathrm{div}(Tg)\|_{\ell^2(S_V)} \leq \|u_0 - \mathrm{div}\, g\|_{\ell^2(S_V)},$$

with equality if and only if $Tg = g$ which in turn by condition (1) is equivalent to $\mathrm{div}\, g = \widetilde{\psi}$. Hence for any $g \in t\mathcal{B}_{\ell^\infty(S_E)}$, if $\mathrm{div}\, g \neq \widetilde{\psi}$ then

$$\|u_0 - \mathrm{div}(Tg)\|_{\ell^2(S_V)} < \|u_0 - \mathrm{div}\, g\|_{\ell^2(S_V)}.$$

$\square$

We are now ready to show the following theorem which establishes that Algorithm 1 converges to the ROF-minimizer u_{opt}.

THEOREM 3.2.   *Let $g \in t\mathcal{B}_{\ell^\infty(S_E)}$ and $T : t\mathcal{B}_{\ell^\infty(S_E)} \to t\mathcal{B}_{\ell^\infty(S_E)}$ be the operator given by* 3.10-3.11. *Then*

$$\operatorname{div}(T^n g) \to \widetilde{\psi} = u_0 - u_{opt} \ as \ n \to +\infty.$$

Proof.   From Proposition 3.1, it follows that T is continuous and satisfies the conditions (1) and (2).   These conditions in turn give that the sequence $\left(\|u_0 - \operatorname{div}(T^n g)\|_{\ell^2(S_V)} \right)_{n \in \mathbb{N}}$ is monotonically decreasing and bounded below by $\left\| u_0 - \widetilde{\psi} \right\|_{\ell^2(S_V)}$. Therefore it converges. Let us now consider the sequence $(T^n g)_{n \in \mathbb{N}} \subset t\mathcal{B}_{\ell^\infty(S_E)}$. The ball $t\mathcal{B}_{\ell^\infty(S_E)}$ is a compact set and therefore has $(T^n g)_{n \in \mathbb{N}}$ a convergent subsequence in $t\mathcal{B}_{\ell^\infty(S_E)}$, say $(T^{n_k} g)_{k \in \mathbb{N}}$:

$$\lim_{k \to \infty} T^{n_k} g = g_\psi \in t\mathcal{B}_{\ell^\infty(S_E)}.$$

Since T, div and $\|\cdot\|_{\ell^2(S_V)}$ are continuous operators, we have

$$\|u_0 - \operatorname{div}(T g_\psi)\|_{\ell^2(S_V)} = \left\| u_0 - \operatorname{div}\left(T\left(\lim_{k \to \infty} T^{n_k} g \right) \right) \right\|_{\ell^2(S_V)}$$
$$= \lim_{k \to \infty} \|u_0 - \operatorname{div}(T(T^{n_k} g))\|_{\ell^2(S_V)}$$
$$= \lim_{k \to \infty} \|u_0 - \operatorname{div}(T^{n_k+1} g)\|_{\ell^2(S_V)}.$$

As $T^{n_{k+1}} g = T^m T^{n_k+1} g$ for some $m \in \{0, 1, 2, \dots\}$, Proposition 3.1 implies

$$\lim_{k \to \infty} \|u_0 - \operatorname{div}(T^{n_k+1} g)\|_{\ell^2(S_V)} \geq \lim_{k \to \infty} \|u_0 - \operatorname{div}(T^{n_{k+1}} g)\|_{\ell^2(S_V)}.$$

The continuity of div and $\|\cdot\|_{\ell^2(S_V)}$ then gives

$$\|u_0 - \operatorname{div}(T g_\psi)\|_{\ell^2(S_V)} \geq \lim_{k \to \infty} \|u_0 - \operatorname{div}(T^{n_{k+1}} g)\|_{\ell^2(S_V)} = \|u_0 - \operatorname{div} g_\psi\|_{\ell^2(S_V)}.$$

Applying Proposition 3.1 again, we conclude that

$$\operatorname{div} g_\psi = \widetilde{\psi}$$

and therefore, by the continuity of div,

$$\lim_{k \to \infty} \operatorname{div}(T^{n_k} g) = \widetilde{\psi}. \tag{3.17}$$

The final step is to show the convergence of the entire sequence $(\operatorname{div}(T^n g))_{n \in \mathbb{N}}$. From 3.17 it follows that

$$\lim_{k \to \infty} \|u_0 - \operatorname{div}(T^{n_k} g)\|_{\ell^2(S_V)} = \left\| u_0 - \widetilde{\psi} \right\|_{\ell^2(S_V)}.$$

Since the subsequence $\left(\|u_0 - \operatorname{div}(T^{n_k} g)\|_{\ell^2(S_V)} \right)_{k \in \mathbb{N}}$ must converge to the same limit as the convergent sequence $\left(\|u_0 - \operatorname{div}(T^n g)\|_{\ell^2(S_V)} \right)_{n \in \mathbb{N}}$, we conclude that

$$\lim_{n \to \infty} \|u_0 - \operatorname{div}(T^n g)\|_{\ell^2(S_V)} = \left\| u_0 - \widetilde{\psi} \right\|_{\ell^2(S_V)}.$$

Therefore, as $\widetilde{\psi}$ is the unique nearest element to u_0 in $t\mathcal{B}_{BV^*(S_V)} = t\,\mathrm{div}\left(\mathcal{B}_{\ell^\infty(S_E)}\right)$, we have

$$\lim_{n\to\infty} \mathrm{div}\left(T^n g\right) = \widetilde{\psi}.$$

□

4. Numerical results

In order to evaluate the performance of the proposed method, we make a numerical comparison with other efficient iterative TV minimization methods and one of the best known state-of-the-art denoising methods, namely BM3D. We have tested different types of images and made comparisons in terms of convergence rate, peak signal to noise ratio (PSNR), running time and visual quality. All experiments were performed on a *Windows 7 Professional 32-bit* computer with a *Intel(R) Core(TM) i5-2400 CPU, 3.1 GHz* Processor and a RAM of *4096 MB*.

4.1. Comparison with other iterative TV minimization methods.   In this subsection the proposed algorithm is compared numerically with two state-of-the-art iterative algorithms for TV denoising, the Split-Bregman algorithm [19] and the Primal-Dual algorithm [10]. We include numerical results obtained by testing different types of images and various noise levels. More specifically we consider a denoising scenario of natural and cartoon images aimed to numerically evaluate and illustrate the proposed algorithm's convergence rate and PSNR. In our experiments we used Gaussian noise with standard deviation 10, 20 and 30.

- The implementation of the Split-Bregman algorithm was obtained from [26]. To find the best performing regularization parameter λ we performed a brute-force optimization in the interval $[1,30]$ uniformly quantized into 100 values.

- The implementation for the Primal-Dual algorithm was obtained from the publicly available repository GPU4Vision `https://github.com/VLOGroup/primal-dual-toolbox`. In this implementation $\tau = \sigma = 1/\sqrt{8}, \gamma = 0.7\lambda$, the value of λ was optimized in the same range as the regularization parameter in the Split-Bregman algorithm. The parameter θ was dynamically updated at each iteration by the rule $\theta \leftarrow 1/\sqrt{1+2\gamma\tau}$ as well as $\tau \leftarrow \tau\theta$ and $\sigma \leftarrow \sigma/\theta$.

- We implemented the proposed ROF model on the graph Algorithm 1 and optimized the regularization parameter using the same parameter space as the Split-Bregman algorithm.

The stopping criteria for all approaches was set to $||u^k - u^{k-1}||/||u^k|| < 10^{-5}$, where $||\cdot||$ is the Frobenius norm and u^k is the current iterate of the numerical scheme.

Cartoon image denoising. In this example we have the *exact* ground truth image data u available which makes an objective evaluation of the methods possible. Figures 4.1, 4.3 and 4.5 show the qualitative results for each evaluated method and noise level with the corresponding best PSNR values. The proposed method produces results with the best PSNR value in all cases. The visual quality of the results produced by all methods is comparable. Panels (a) of Figures 4.2, 4.4 and 4.6 show the respective peak signal to noise ratio (PSNR) curves for the highest obtained PSNR values, obtained after a dense parameter grid search as previously described, for the respective methods and noise levels. Panels (b) of the same figures show the descent towards the stopping criteria. Each algorithm was terminated when the normalized difference between the current iterate and the previous iterate became smaller than 10^{-5}, this is illustrated in panels (b). From the same figures it is clear that up till 30 (resp. 20) iterations for
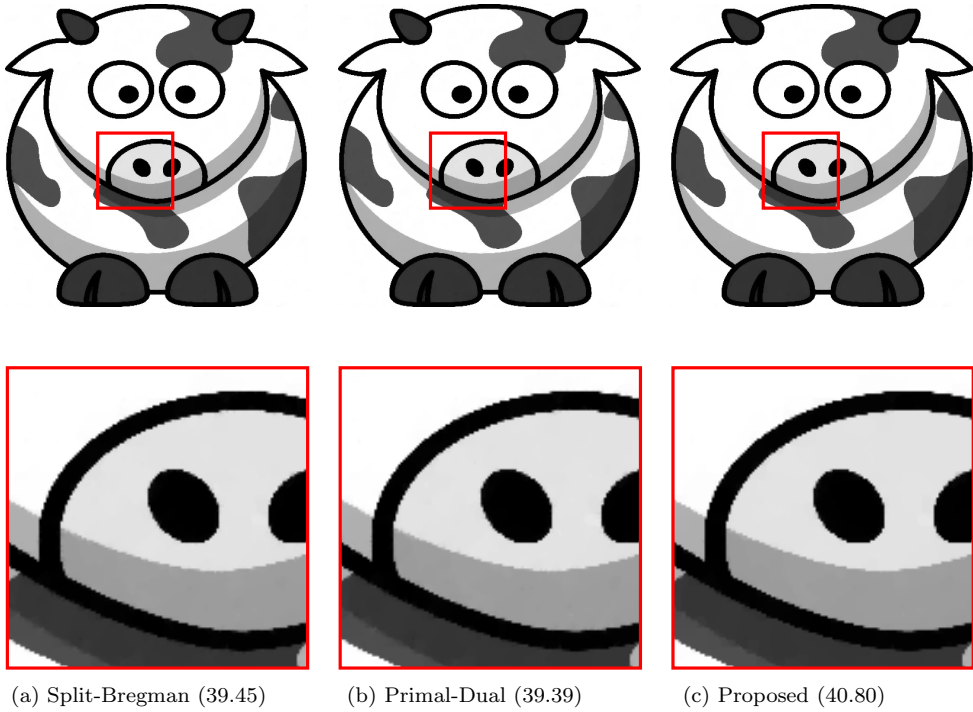
(a) Split-Bregman (39.45)   (b) Primal-Dual (39.39)   (c) Proposed (40.80)

Fig. 4.1: *Reconstructions when the original image was corrupted by Gaussian noise of standard deviation 10. Panels (a)-(c) show the images obtained at the best PSNR value for a brute-force parameter optimization strategy described in the main text. Obtained PSNR values are shown in parenthesis. Reaching the stopping criteria of 10^{-5}, the proposed solution strategy shows improved PSNR compared to the other methods. We refer to Figure 4.2 for empirical convergence results.*
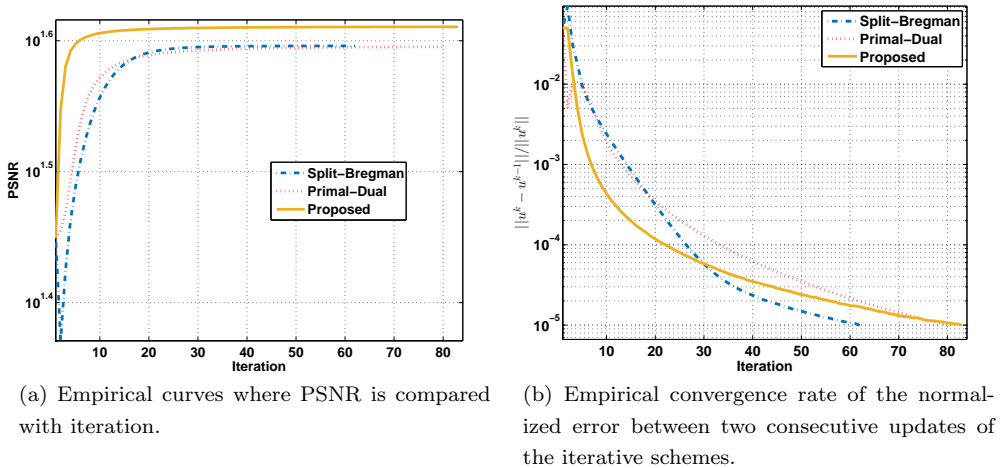

(a) Empirical curves where PSNR is compared with iteration.

(b) Empirical convergence rate of the normalized error between two consecutive updates of the iterative schemes.

Fig. 4.2: *Panels (a)-(b) show the PSNR and empirical convergence rates for the solution images in Figure 4.1. Panel (b) depicts that after 30 iterations the Split-Bregman scheme shows a smaller relative update between two consecutive update steps, this indicates earlier convergence for said method. However, after 30 iterations the improvement of the iterates is negligible compared to the ground truth data as seen in panel (a).*
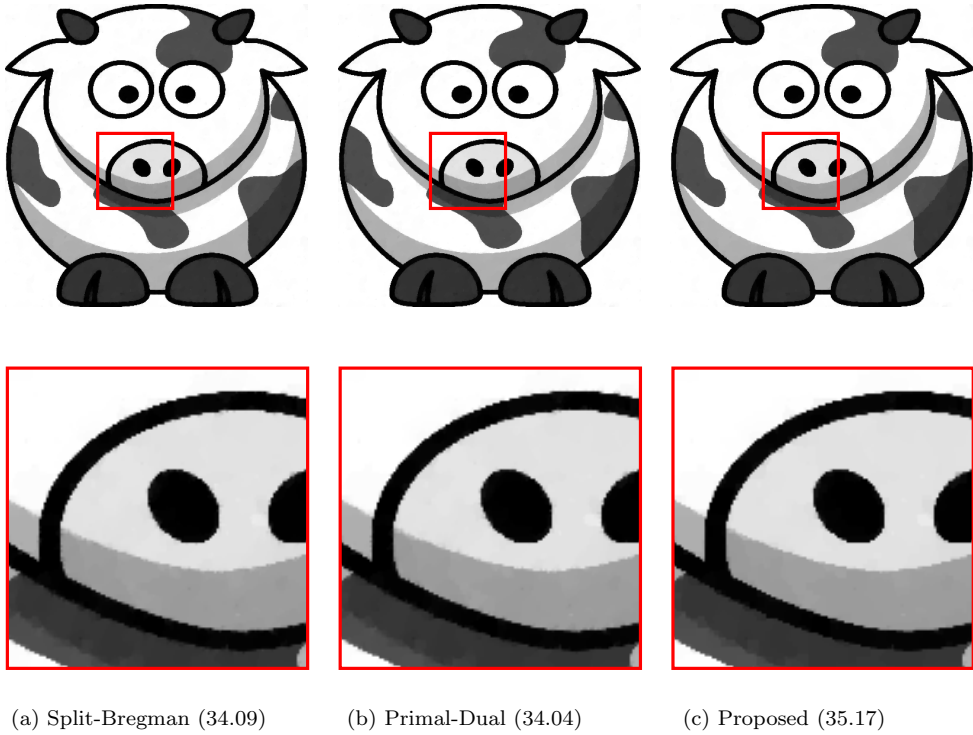
(a) Split-Bregman (34.09)   (b) Primal-Dual (34.04)   (c) Proposed (35.17)

Fig. 4.3: *Reconstructions when the original image was corrupted by Gaussian noise of standard deviation 20. Panels (a)-(c) show the images obtained at the best PSNR value for a brute-force parameter optimization strategy described in the main text. Obtained PSNR values are shown in parenthesis. Reaching the stopping criteria of 10^{-5}, the proposed solution strategy shows improved PSNR compared to the other methods. We refer to Figure 4.4 for empirical convergence results.*
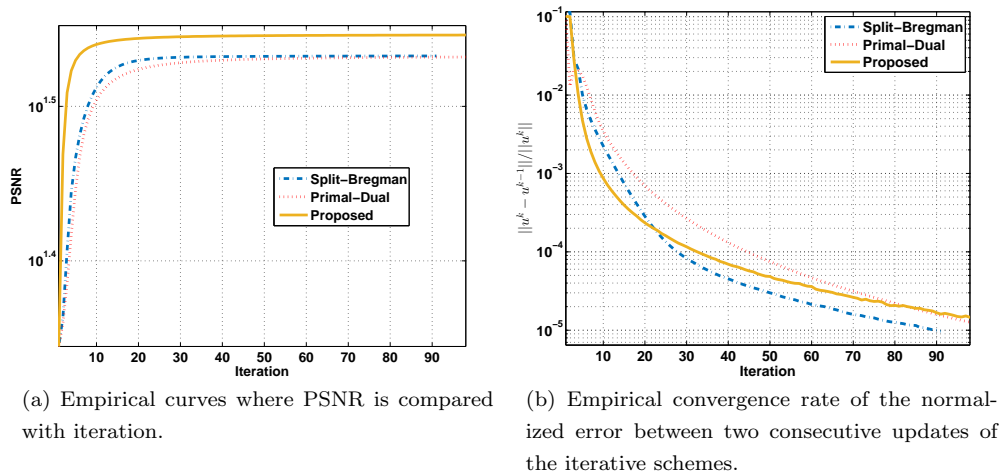

(a) Empirical curves where PSNR is compared with iteration.

(b) Empirical convergence rate of the normalized error between two consecutive updates of the iterative schemes.

Fig. 4.4: *Panels (a)-(b) show the PSNR and empirical convergence rates for the solution images in Figure 4.3. Panel (b) depicts that after 20 iterations the Split-Bregman scheme shows a smaller relative update between two consecutive update steps, this indicates earlier convergence for said method. However, after 20 iterations the improvement of the iterates is negligible compared to the ground truth data as seen in panel (a).*
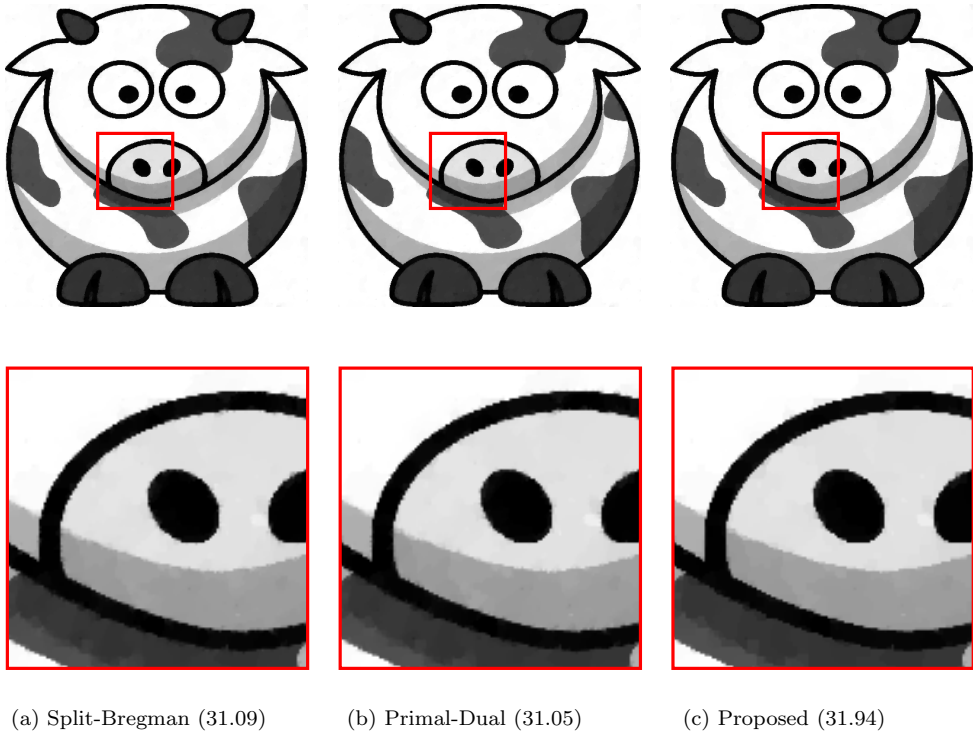
(a) Split-Bregman (31.09)   (b) Primal-Dual (31.05)   (c) Proposed (31.94)

Fig. 4.5: *Reconstructions when the original image was corrupted by Gaussian noise of standard deviation 30. Panels (a)-(c) show the images obtained at the best PSNR value for a brute-force parameter optimization strategy described in the main text. Obtained PSNR values are shown in parenthesis. Reaching the stopping criteria of 10^{-5}, the proposed solution strategy shows improved PSNR compared to the other methods. We refer to Figure 4.6 for empirical convergence results.*

(a) Empirical curves where PSNR is compared with iteration.

(b) Empirical convergence rate of the normalized error between two consecutive updates of the iterative schemes.

Fig. 4.6: *Panels (a)-(b) show the PSNR and empirical convergence rates for the solution images in Figure 4.5. Panel (b) depicts that after 20 iterations the Split-Bregman scheme shows a smaller relative update between two consecutive update steps, this indicates earlier convergence for said method. However, after 20 iterations the improvement of the iterates is negligible compared to the ground truth data as seen in panel (a).*
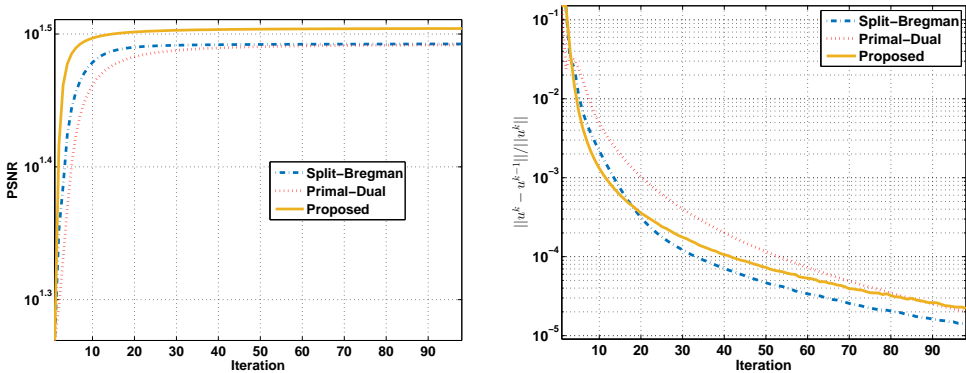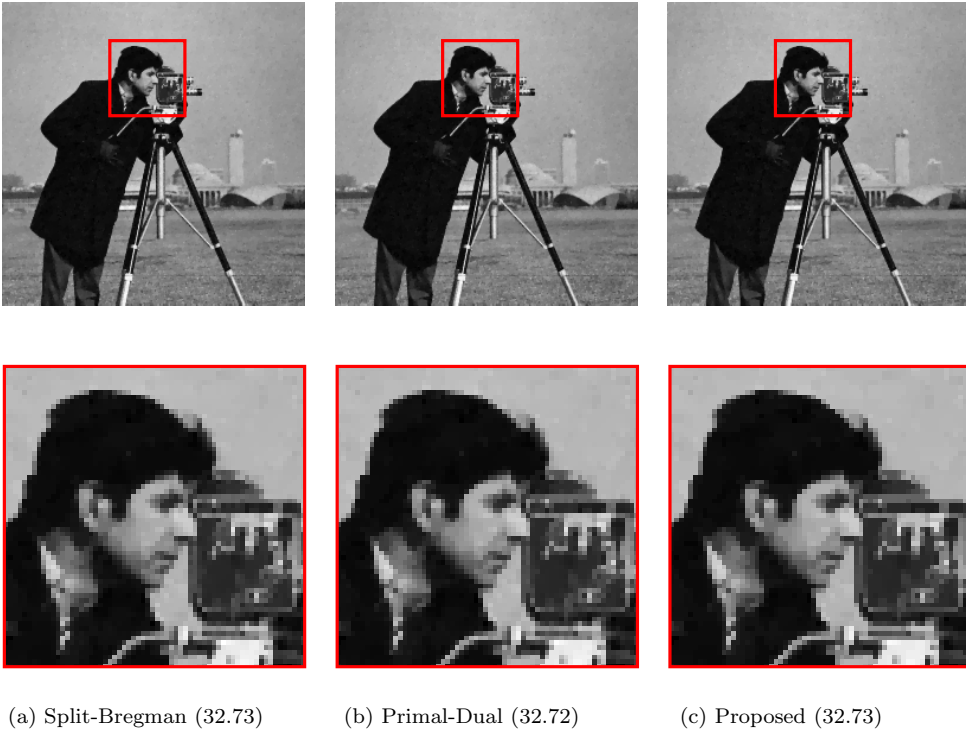
(a) Split-Bregman (32.73)   (b) Primal-Dual (32.72)   (c) Proposed (32.73)

Fig. 4.7: *Reconstructions when the original image was corrupted by Gaussian noise of standard deviation 10. Panels (a)-(c) show the images obtained at the best PSNR value for a brute-force parameter optimization strategy described in the main text. Obtained PSNR values are shown in parenthesis. The methods produce near identical PSNR values. We refer to Figure 4.8 for empirical convergence results.*

(a) Empirical curves where PSNR is compared with iteration.

(b) Empirical convergence rate of the normalized error between two consecutive updates of the iterative schemes.
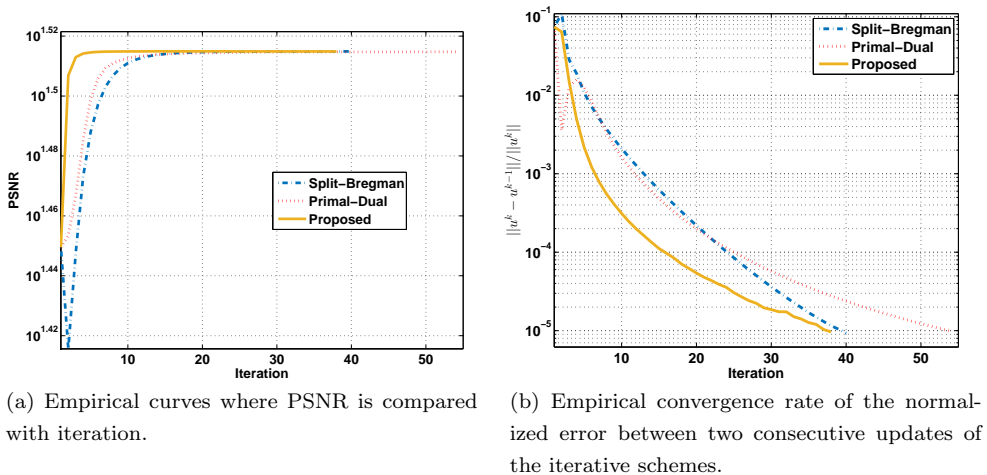
Fig. 4.8: *Panels (a)-(b) show the PSNR and empirical convergence rates to obtain the solution images in Figure 4.7. In this example, the proposed solution scheme shows an improved convergence rate compared to the Split-Bregman and the Primal-Dual approaches, yet resulting in near identical error values as seen in panel (a).*

(a) Split-Bregman (28.93)   (b) Primal-Dual (28.92)   (c) Proposed (28.93)

Fig. 4.9: *Reconstructions when the original image was corrupted by Gaussian noise of standard deviation 20. Panels (a)-(c) show the images obtained at the best PSNR value for a brute-force parameter optimization strategy described in the main text. Obtained PSNR values are shown in parenthesis. The methods produce near identical PSNR values. We refer to Figure 4.10 for empirical convergence results.*

(a) Empirical curves where PSNR is compared with iteration.

(b) Empirical convergence rate of the normalized error between two consecutive updates of the iterative schemes.

Fig. 4.10: *Panels (a)-(b) show the PSNR and empirical convergence rates to obtain the solution images in Figure 4.9. In this example, the proposed solution scheme shows an improved convergence rate compared to the Split-Bregman and the Primal-Dual approaches, yet resulting in near identical error values as seen in panel(a).*
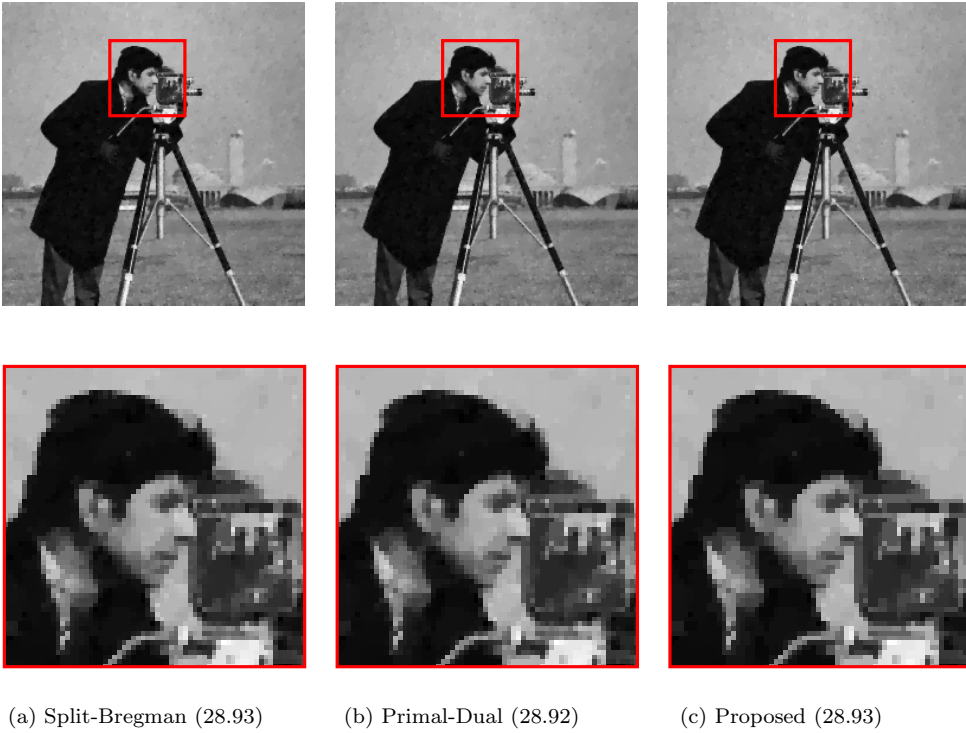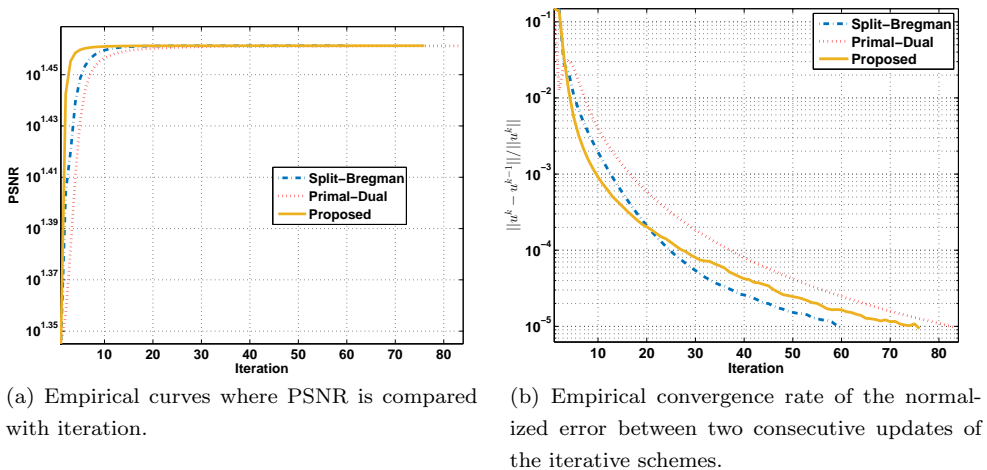
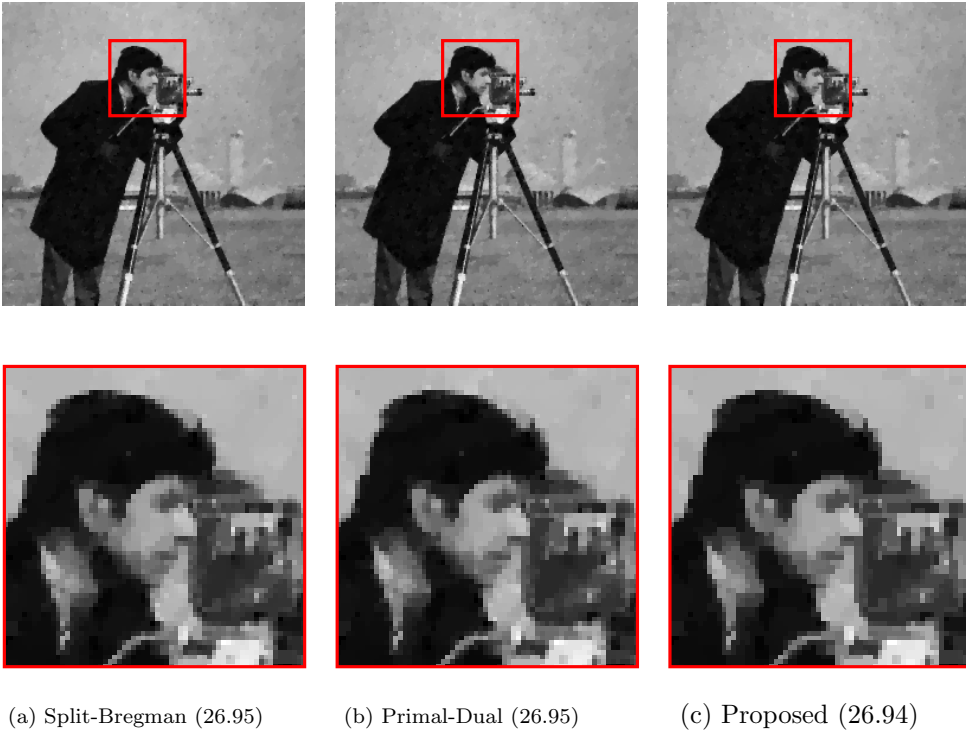(a) Split-Bregman (26.95)   (b) Primal-Dual (26.95)   (c) Proposed (26.94)

Fig. 4.11: *Reconstructions when the original image was corrupted by Gaussian noise of standard deviation 30. Panels (a)-(c) show the images obtained at the best PSNR value for a brute-force parameter optimization strategy described in the main text. Obtained PSNR values are shown in parenthesis. The methods produce near identical PSNR values. We refer to Figure 4.12 for empirical convergence results.*

(a) Empirical curves where PSNR is compared with iteration.

(b) Empirical convergence rate of the normalized error between two consecutive updates of the iterative schemes.
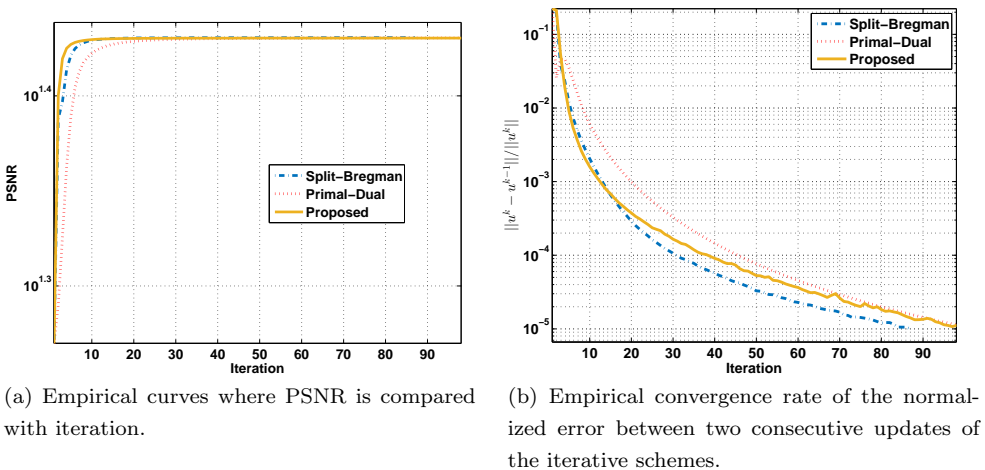
Fig. 4.12: *Panels (a)-(b) show the PSNR and empirical convergence rates to obtain the solution images in Figure 4.11. In this example, the proposed solution scheme shows an improved convergence rate compared to the Split-Bregman and the Primal-Dual approaches, yet resulting in near identical error values as seen in panel (a).*

(a) Man

(b) Couple

(c) Hill

(d) Boat

Fig. 4.13: *Images used to compare the proposed algorithm with the BM3D in terms of PSNR, running time, and visual quality.*

noise level 10 (resp. for noise levels 20 and 30) the proposed method shows a faster convergence rate. Note, however, that after these number of iterations any further updates of the iterative schemes have an negligible effect to the end result.

Natural image denoising. In this imaging scenario, we denoise the "cameraman" image. We have the *exact* ground truth image data u available so that it is possible to evaluate the methods objectively. Figures 4.7, 4.9 and 4.11 show the qualitative results for each evaluated method and noise level with the corresponding best PSNR values. The proposed method produces results with the best or comparable PSNR value in all cases. The visual quality of the results produced by all methods is also compara-

(a) Man: Noisy with $\sigma = 20$

(b) Couple: Noisy with $\sigma = 20$

(c) Man: BM3D with PSNR 30.59 dB

(d) Couple: BM3D with PSNR 30.76 dB

(e) Man: Proposed with PSNR 29.42 dB

(f) Couple: Proposed with PSNR 28.90 dB

Fig. 4.14: *Subjective visual quality comparison between denoised images by BM3D and the proposed method. Noise with $\sigma = 20$ has been added to the images shown in 4.13a and 4.13b. In both cases it is clear that though there is a gap in PSNR values, the methods generate reconstructions of similar visual quality. The same conclusion holds for the other images shown in Figure 4.13 and different noise levels.*

| | Man | | Couple | | Hill | | Boat | |
| | 512×512 | | 512×512 | | 512×512 | | 512×512 | |
σ	BM3D	Proposed	BM3D	Proposed	BM3D	Proposed	BM3D	Proposed
5	37.82	36.80	37.52	36.50	37.13	36.40	37.06	36.27
10	33.98	32.77	34.04	32.44	33.62	32.57	33.89	32.49
15	31.93	30.73	32.11	30.31	31.86	30.70	32.10	30.50
20	30.59	29.34	30.76	28.90	30.72	29.50	30.83	29.17
25	29.62	28.48	29.72	27.90	29.85	28.63	29.85	28.18
30	28.86	27.75	28.87	27.09	29.15	27.94	29.04	27.41
35	28.22	27.16	28.15	26.45	28.49	27.39	28.32	26.78
40	27.65	26.67	27.48	25.93	27.88	26.94	27.63	26.25
45	27.17	26.25	26.91	25.09	27.43	26.56	27.12	25.81
50	26.81	25.88	26.46	24.75	27.19	26.23	26.67	25.43
55	26.44	25.56	26.01	24.48	26.74	25.94	26.27	25.10
60	26.14	25.27	25.66	24.45	26.52	25.68	25.90	24.80
65	25.90	25.01	25.29	24.18	26.12	25.44	25.56	24.52
70	25.56	24.77	25.00	23.94	25.93	25.23	25.25	24.27
75	25.32	24.56	24.70	23.72	25.68	25.03	24.97	24.04
80	25.06	24.36	24.42	23.52	25.43	24.85	24.70	23.83
85	24.86	24.17	24.21	23.33	25.10	24.68	24.45	23.64
90	24.63	24.00	23.94	23.16	24.98	24.52	24.22	23.46
95	24.39	23.83	23.67	23.01	24.66	24.38	24.01	23.29
100	24.22	23.67	23.51	22.86	24.58	24.24	23.80	23.13

Table 4.1: *PSNR (dB) results of the proposed method and BM3D method*

ble. In accordance with the previous example, panels (a) of Figures 4.8, 4.10 and 4.12 show the respective peak signal to noise ratio (PSNR) curves for the highest obtained PSNR values, obtained after a dense parameter grid search as previously described, for the respective methods and noise levels. Further, panels (b) of the same figures show the descent towards the stopping criteria. Each algorithm was terminated when the normalized difference between the current iterate and the previous iterate became smaller than 10^{-5}, this is illustrated in panels (b). From the same figures it is clear that similar to the cartoon image, the relative convergence rate seen in (b) is initially faster for the proposed method than for the compared methods, and then it is overcome by the Split-Bregman algorithm at a certain number of iterations after which the improvement of the image quality is insignificant for all methods. Panels (b) show that the best error rates for the proposed method are obtained at comparable or fewer iterations than the Split-Bregman and the Primal-Dual approaches.

4.2. Comparison with BM3D. In this subsection we compare numerically the proposed algorithm with the BM3D denoising method introduced in [13]. This method is a state-of-the-art filtering method applicable to denoising among other important problems in image processing. It is therefore interesting to make a comparison although BM3D is not directly designed to solve the TV minimization problem.

In our experiments we used different types of images and Gaussian noise with standard deviation ranging from 5 to 100. More precisely, we have reproduced and compared results obtained by BM3D and by the proposed method by testing all images available at the online repository http://www.cs.tut.fi/~foi/GCF-BM3D/. However, since all results obtained lead to the same conclusions, we have only reported results obtained for test images illustrated in Figure 4.13. Comparison with BM3D has been made in terms of PSNR, running time and visual quality.

	Man 512×512		Couple 512×512		Hill 512×512		Boat 512×512	
σ	BM3D	Proposed	BM3D	Proposed	BM3D	Proposed	BM3D	Proposed
5	7.50	1.24	7.40	0.99	8.10	0.98	11.4	0.99
10	8.00	2.37	7.80	2.04	8.50	2.15	8.4	2.21
15	8.40	3.40	8.20	3.09	8.90	3.52	8.7	3.28
20	8.60	4.66	8.50	4.06	9.20	4.67	11.6	4.15
25	8.90	5.42	8.70	5.14	9.50	5.72	9.0	5.08
30	9.00	6.47	9.00	6.05	9.50	6.89	9.1	6.23
35	9.10	7.57	9.00	7.02	9.50	7.77	9.0	6.94
40	8.80	8.55	8.60	7.99	9.30	8.96	8.8	8.18
45	10.50	9.69	10.50	8.88	10.70	9.89	10.6	9.01
50	10.60	10.84	10.60	9.66	10.80	10.70	10.7	9.62
55	10.70	11.62	10.70	10.42	10.90	12.12	10.8	10.54
60	10.80	12.46	10.70	11.48	11.00	13.20	10.9	11.54
65	10.80	14.00	10.80	12.14	11.10	13.61	11.0	12.42
70	10.90	14.35	10.90	13.26	11.10	15.13	13.4	13.15
75	11.00	15.53	11.00	13.87	11.20	15.46	11.3	14.30
80	11.00	16.60	11.10	15.08	11.20	16.02	11.3	15.20
85	11.10	17.40	11.10	16.02	11.40	17.65	11.4	15.85
90	11.20	17.99	11.20	16.76	11.40	18.34	11.4	16.44
95	11.20	18.97	11.30	17.93	11.50	19.54	11.4	17.22
100	11.20	19.32	11.30	18.32	11.50	20.49	11.4	17.38

Table 4.2: *Running time (seconds) results of the proposed method and BM3D method*

- The implementation of the BM3D was obtained from [13] and the online repository http://www.cs.tut.fi/~foi/GCF-BM3D/.
- We implemented the proposed Algorithm 1 and optimized the regularization parameter t by performing a brute-force optimization with uniformly quantized values in suitable sub-intervals of the interval $[1, 100]$ corresponding to noise levels ranging from 5, 10, …, 100.

Obtained results in terms of PSNR, running time and visual quality are shown in Table 4.1, Table 4.2 and Figure 4.14 respectively. As might be expected, the PSNR values for the test images is higher for the BM3D method as shown in Table 4.1. The difference in PSNR values is in the range of 1 to 2 dB. In terms of running time, as can be seen in Table 4.2, the proposed algorithm is the most competitive in the low noise regime. However, it is also more sensitive to the noise level than the BM3D method. In terms of visual quality, the methods produce comparable results.

5. Conclusion

In this work we proposed an iterative algorithm for total variation minimization on graphs and proved its convergence. The algorithm that is presented can be viewed as a coordinate descent on dual space and can be run on a parallel computer architecture, which makes it suitable to handle large graphs and data sets. The algorithm is simple, easy to implement and converges to the exact minimizer with fewer iterations compared to the Split-Bregman and Primal-Dual algorithms. Furthermore, in order to compare the proposed method with other state-of-the-art denoising methods, BM3D was chosen as a benchmark and obtained results still show competitive performance for the proposed method. In a follow-up work we intend to further study the convergence rate of the algorithm and include additional imaging scenarios.

REFERENCES

[1] J.F. Aujol, G. Aubert, L. Blanc-Feraud, and A. Chambolle, *Image decomposition into a bounded variation component and an oscillating component*, J. Math. Imaging Vision, 22:71–88, 2005. 1.1

[2] J.F. Aujol and A. Chambolle, *Dual norms and image decomposition models*, Int. J. Comput. Vision, 63(1):85–104, 2005. 1.1

[3] J.F. Aujol, G. Gilboa, T. Chan, and S. Osher, *Structure-texture image decomposition – Modeling, algorithms, and parameter Selection*, Int. J. Comput. Vision, 67(1):111–136, 2006. 1.1

[4] A. Beck and M. Teboulle, *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci., 2(1):183–202, 2009. 1.1

[5] A. Beck and M. Teboulle, *Fast gradient-based algorithms for constrained total variation image denoising and deblurring problems*, IEEE Trans. Image Process, 18(11):2419–2434, 2009. 1.1

[6] J. Bergh and J. Löfström, *Interpolation Spaces. An Introduction*, Springer, Berlin, 1976. 2.1

[7] A. Chambolle, *An algorithm for total variation minimization and applications*, J. Math. Imaging Vision, 20:89–97, 2004. 1.1

[8] A. Chambolle, *Total variation minimization and a class of binary MRF models*, in Energy Minimization Methods in Computer Vision and Pattern Recognition, Springer-Verlag, Berlin, 136–152, 2005. 1.1

[9] A. Chambolle and J. Darbon, *Algorithms for total variation minimization*, in Mathematical Methods for Multichannel Image Processing, Beijing , 7–8, 2006. 1.1

[10] A. Chambolle and T. Pock, *A first-order primal-dual algorithm for convex problems with applications to imaging*, J. Math. Imaging Vision, 40:120–145, 2011. 1.1, 4.1

[11] T.F. Chan, G.H. Golub, and P. Mulet, *A nonlinear primal–dual method for total variation–based image restoration*, SIAM J. Sci. Comput., 20(6):1964–1977, 1999. 1.1

[12] T.F. Chan and J. Shen, *Image Processing and Analysis. Variational, PDE, Wavelet, and Stochastic Methods*, SIAM, Philadelphia, PA., 2005. 1.1

[13] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, *Image denoising by sparse 3-D transform-domain collaborative filtering*, IEEE Trans. Image Process., 16(8):2080–2095, 2007. 1.1, 4.2

[14] J. Darbon and M. Sigelle, *Image restoration with discrete constrained total variation part I: Fast and exact optimization*, J. Math. Imaging Vision, 26:261–276, 2006. 1.1

[15] N. Dunford and J.T. Schwartz, *Linear Operators. Part I. General Theory*, Wiley, New York, Reprint of the 1958 Edition, 1988. 2.2, 2.2

[16] J. Gilles, *Noisy image decomposition: a new structure, texture and noise model based on local adaptivity*, J. Math. Imaging Vision, 28:285–295, 2007. 1.1

[17] J. Gilles and S. Osher, *Bregman implementation of Meyer's G-norm for cartoon + textures decomposition*, UCLA CAM Report, 11-73, 2011. 1.1

[18] D. Goldfarb and W. Yin, *Parametric maximum flow algorithms for fast total variation minimization*, SIAM J. Sci. Comput., 31(5):3712–3743, 2009. 1.1

[19] T. Goldstein and S. Osher, *The split Bregman Method for L^1–regularized problems*, SIAM J. Imaging Sci., 2(2):323–343, 2009. 1.1, 4.1

[20] N. Kruglyak and J. Niyobuhungiro, *Characterization of optimal decompositions in real interpolation*, J. Approx. Theory, 185:1–11, 2014. 3.1.2

[21] M. Lebrun, *An analysis and implementation of the BM3D image denoising method*, Image Processing On Line, 2:175–213, 2012. 1.1

[22] Y. Meyer, *Oscillating Patterns in Image Processing and Nonlinear Evolution Equations*, University Lecture Series, AMS, Providence, RI., 22, 2001. 1.1

[23] S. Osher and O. Scherzer, *G–norm properties of bounded variation regularizaton*, Commun. Math. Sci., 2(2):237–254, 2004. 1.1

[24] L.I. Rudin, S. Osher, and E. Fatemi, *Nonlinear total variation based noise removal algorithms*, Phys. D, 60:259–268, 1992. 1.1

[25] D.M. Strong, J.F. Aujol, and T.F. Chan, *Scale recognition, regularization parameter selection and Meyer's G–norm in total variation regularization*, Multiscale Model. Simul., 5(1):273–303,

2006. 1.1

[26] B. Tremoulheac, *Matlab central, Split-Bregman algorithm*, File ID: #36278. Accessed: 2016-02-04.
      4.1