DUAL CONSTRAINED TV-BASED REGULARIZATION ON GRAPHS

CAMILLE COUPRIE [*], LEO GRADY [†], LAURENT NAJMAN [‡], JEAN-CHRISTOPHE PESQUET [§], AND HUGUES TALBOT [‡]

Abstract. Algorithms based on Total Variation (TV) minimization are prevalent in image processing. They play a key role in a variety of applications such as image denoising, compressive sensing and inverse problems in general. In this work, we extend the TV dual framework that includes Chambolle's and Gilboa-Osher's projection algorithms for TV minimization. We use a flexible graph data representation that allows us to generalize the constraint on the projection variable. We show how this new formulation of the TV problem may be solved by means of fast parallel proximal algorithms. On denoising and deblurring examples, the proposed approach is shown not only to perform better than recent TV-based approaches, but also to perform well on arbitrary graphs instead of regular grids. The proposed method consequently applies to a variety of other inverse problems including image fusion and mesh filtering.

Key words. Proximal methods, inverse problems, image denoising, convex optimization, parallel algorithms, image restoration, bounded variation, mesh smoothing, discrete calculus.

1. Introduction.

1.1. Problem statement. In this paper, we consider a weighted graph $G = (V, E)$ with vertices $v \in V$ and edges $e \in E \subseteq V \times V$, with cardinalities $n = \operatorname{card}(V)$ and $m = \operatorname{card}(E)$.

Let $H \in \mathbb{R}^{q \times n}$, for some $q \in \mathbb{N}^*$, be the matrix associated with a linear observation process. A data vector $f = (f_i)_{1 \leq i \leq q} \in \mathbb{R}^q$ is assumed to be obtained from $\bar{x} = (\bar{x}_i)_{1 \leq i \leq n} \in \mathbb{R}^n$ through the model

$$f = H\bar{x} + b, \tag{1.1}$$

where $b \in \mathbb{R}^q$ is a realization of a random noise. Our purpose is to reconstruct a vector close to $\bar{x}$ from f. In order to do so, we propose to solve an optimization problem taking the form

$$\operatorname*{minimize}_{x \in \mathbb{R}^n} \left(\underbrace{\sup_{F \in C} F^\top Dx}_{\text{regularization}} + \underbrace{\frac{1}{2}(Hx - f)^\top \Lambda^{-1}(Hx - f)}_{\text{data fidelity}} \right) \tag{1.2}$$

where Λ is a symmetric positive-definite weighting matrix, D is the so-called *weighted incidence matrix of the graph* G (see below), and C is the intersection of closed balls defined with weighted semi-norms. The regularization term corresponds to a generalization of the combinatorial formulation of the TV dual problem, as we are going to detail shortly. We name (1.2) the Dual Constrained Total Variation (DCTV) problem[1]. We note that any solution $\hat{x}$ to (1.2) is parameterized by the weighted graph G

---

[*]IFP Energies nouvelles, Technology, Computer Science and Applied Mathematics Division, Rueil Malmaison, France (camille.couprie@ifpen.fr). A major part of the work was performed at Université Paris Est, LIGM, France

[†] HeartFlow, Inc., Redwood City, California, USA (leograd@yahoo.com).

[‡]Université Paris-Est, Laboratoire d'Informatique Gaspard-Monge, ESIEE Paris (93160 Noisy-le-Grand, France) (h.talbot@esiee.fr, l.najman@esiee.fr).

[§]Université Paris-Est, Laboratoire d'Informatique Gaspard Monge (77454 Marne-la-vallée, France) (jean-christophe.pesquet@univ-paris-est.fr).

[1]A preliminary version of this work was presented in [1].

2

and that the components of $\widehat{x}$ can be viewed as optimal values which are attached to each node of the graph. In Section 2, we show how Problem (1.2) can be efficiently solved thanks to proximal methods.

A typical applicative context of such a model is image recovery. In this case, the nodes of the graph usually represent the pixels of an image and the edges a neighborhood relationship between pixels. However, there is no reason to restrict the model to this particular framework, as there is no limitation on the type of graph our model can deal with. The matrix H may simply be the identity matrix for image denoising, a convolution operator in restoration tasks, or a projection matrix in reconstruction problems. The matrix Λ may be proportional to the covariance matrix of the noise corrupting the data, as is commonly used in weighted least squares approaches. As detailed in Section 1.4, improving the regularization term by exploiting duality is a highly relevant problem investigated by a number of researchers, for instance Osher *et al.* [2], Bredies *et al.* [3, 4], Lenzen *et al.* [5, 6], Hidane *et al.* [7].

1.2. Discrete calculus interpretation of Problem (1.2). Discrete calculus [8, 9] has been employed in recent years to produce a combinatorial reformulation of continuous problems onto a graph in such a manner that the solution behaves analogously to the continuous formulation (e.g., [10, 11]). Additionally, it has been shown that the combinatorial reformulation of classical energies is able to provide new avenues for performing efficient combinatorial optimization [12, 10, 11, 13].

An edge spanning two vertices, v_i and v_j, with $1 \leq i \leq n$ and $1 \leq j \leq n$, is denoted by $e_{i,j}$. We deal with weighted graphs that include weights on both the edges and nodes. An edge weight is a value assigned to each edge $e_{i,j}$, and it is denoted by $w_{i,j}$. We assume that $w_{i,j} \in]0, +\infty[$. The weight of a node v_i is denoted by g_i. We also assume that $g_i \in]0, +\infty[$.

The incidence matrix of a graph is a fundamental operator for defining combinatorial formulations of variational problems. Specifically, the edge-node incidence matrix $A \in \mathbb{R}^{m \times n}$ defines the discrete calculus analogue of the gradient, while $-A^\top$ is known to define the discrete calculus analogue of the divergence (see [9] and the references therein). The incidence matrix can be viewed as an operator mapping functions on nodes (analogue to a scalar field) to functions on edges (analogue to a vector field) and its elements are defined as

$$A_{e_{i,j}v_k} = \begin{cases} -1 & \text{if } i = k, \\ +1 & \text{if } j = k, \\ 0 & \text{otherwise,} \end{cases} \tag{1.3}$$

for every vertex v_k ($k \in \{1, \ldots, n\}$) and edge $e_{i,j}$. An example of a graph represented with its incidence matrix is given in Fig. 1.1.

We introduce the weighted incidence matrix $D = \text{diag}(\sqrt{w})A$, where $\text{diag}(\sqrt{w})$ is a diagonal matrix whose diagonal elements correspond to the square roots of the components of edge weight vector w.

In the context of discrete calculus, the vector $F \in \mathbb{R}^m$ in (1.2) is a vector associated with the edges of G, that we name "flow variable". Indeed, as in the context of graph cuts, F represents the amount of flow that is allowed to pass through an edge.

For an arbitrary matrix M, $|M|$ designates the matrix composed of the absolute values of each entry individually, we denote by $\cdot$ the Hadamard product and by $M^{\cdot 2}$ the product $M \cdot M$.

$$A = \begin{array}{c|cccccc} & v_1 & v_2 & v_3 & v_4 & v_5 & v_6 \\ \hline e_{1,2} & -1 & 1 & 0 & 0 & 0 & 0 \\ e_{2,3} & 0 & -1 & 1 & 0 & 0 & 0 \\ e_{4,5} & 0 & 0 & 0 & -1 & 1 & 0 \\ e_{5,6} & 0 & 0 & 0 & 0 & -1 & 1 \\ e_{1,4} & -1 & 0 & 0 & 1 & 0 & 0 \\ e_{2,5} & 0 & -1 & 0 & 0 & 1 & 0 \\ e_{3,6} & 0 & 0 & -1 & 0 & 0 & 1 \end{array}$$
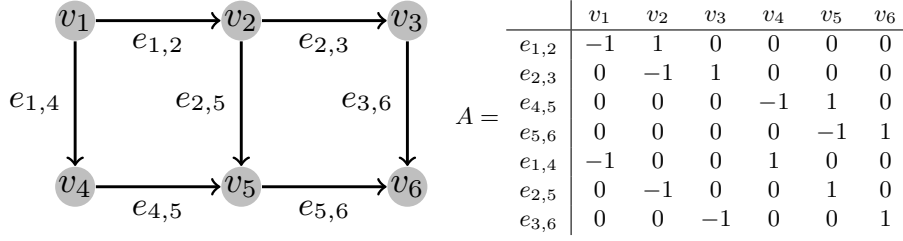
FIG. 1.1. *A graph and its incidence matrix $A \in \mathbb{R}^{m \times n}$ with $m = 7$ and $n = 6$.*

In this work we propose to address nonempty convex sets C defined as

$$C = \{F \in \mathbb{R}^m \mid (\forall i \in \{1, \dots, n\}) \; \|\theta^{(i)} \cdot F\|_\alpha \leq g_i\}, \tag{1.4}$$

where $g = (g_i)_{1 \leq i \leq n}$ is a vector of $[0, +\infty[^n$, $\|\cdot\|_\alpha$ is the ℓ_α norm of $\mathbb{R}^m$ with $\alpha \in [1, +\infty]$ and $\theta^{(i)} \in]0, +\infty[^m$ is a vector of multiplicative constants $\theta^{(i)}_{i',j}$ associated with every couple of nodes (i', j) with $i' \in J$ and $j \in N_{i'}$. Here, for every $i' \in \{1, \dots, n\}$, $N_{i'} = \{j \in \{1, \dots, n\} \mid e_{i',j} \in E\}$ and $J = \{i' \in \{1, \dots, n\} \mid N_{i'} \neq \varnothing\}$. The flow variable is thus constrained by edge weights through the weighted incidence matrix D and by node weights through g.

1.3. Particular cases of interest. We now give three examples of how our formulation may be used to improve recovery results, by considering various convex sets and various graphs. The first example justifies the link with usual TV and hence the name DCTV. In the second example, we provide a convex set that can serve to define a useful local constraint on the flow in many practical situations. The third example illustrates one of the benefits of a combinatorial graph-based approach to regularization problems, in this instance non-local image denoising. In these three preliminary examples, H is always the identity matrix.

EXAMPLE 1.1. *Assume that $J = \{1, \dots, n\}$, consider as convex set,*

$$C = \{(F_{i,j})_{i,j} \mid (\forall i \in \{1, \dots, n\}) \sum_{j \in N_i} F^2_{i,j} \leq 1\}, \tag{1.5}$$

and set $\Lambda = \lambda I$ with $\lambda \in]0, +\infty[$. Problem. (1.2) then reads

$$\underset{x}{\text{minimize}} \; \max_{F \in C} F^\top Dx + \frac{1}{2\lambda} \|x - f\|^2. \tag{1.6}$$

The problem can be reformulated as

$$\underset{x}{\text{minimize}} \; \max_{\|F\|_\infty \leq 1} F^\top ((Ax) \cdot \sqrt{w}) + \frac{1}{2\lambda} \|x - f\|^2 \tag{1.7}$$

where $\|F\|_\infty = \max_{i \in \{1, \dots, n\}} \left(\sum_{j \in N_i} F^2_{i,j} \right)^{1/2}$.

We recognize in (1.7) the combinatorial primal-dual formulation [14] of the discrete weighted TV model as defined in [15, 14, 10]:

$$\underset{x}{\text{minimize}} \; \sum_{i=1}^{n} \left(\sum_{j \in N_i} w_{i,j}(x_j - x_i)^2 \right)^{1/2} + \frac{1}{2\lambda} \sum_{i=1}^{n} (x_i - f_i)^2. \tag{1.8}$$

(a) Two constraints at nodes i and j in a local graph

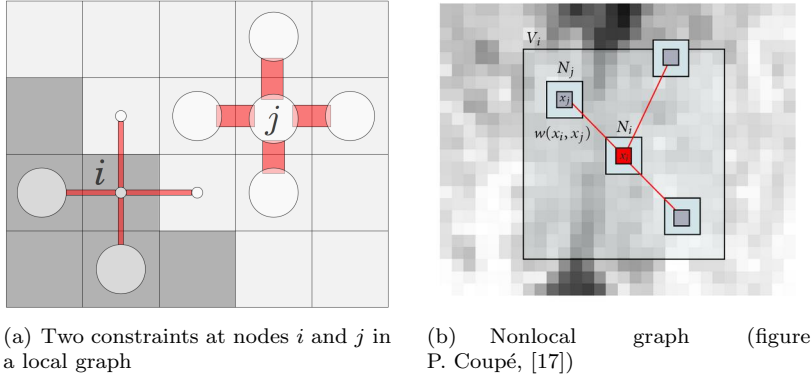(b) Nonlocal graph (figure P. Coupé, [17])

FIG. 1.2. *(a) Illustration of the flow bottleneck created in a contour area (at node i) in a 4-connected lattice. Circles sizes are proportional to the value of g in each node (e.g. small in the presence of contour, large otherwise). The flow is represented by the red pipes. (b) Illustration of some connections in a non-local graph – a node is connected to node centers of similar patches.*

The weighted TV problem appeared previously as a primal-dual formulation in the continuous setting. In a functional framework, given a planar domain Ω, and denoting by u and u' two arbitrary points of Ω, the weighted TV model [14] is given by

$$\operatorname*{minimize}_{x} \quad \int_\Omega \Big(\int_\Omega w(u,u')\big(x(u')-x(u)\big)^2 du' \Big)^{1/2} du + \frac{1}{2\lambda} \int_\Omega \big(x(u)-f(u)\big)^2 du, \quad (1.9)$$

where w is now a non-negative valued function defined on Ω^2. As shown by Chan et al. in [16] the TV minimization problem (1.9) is equivalent to the min-max problem

$$\operatorname*{minimize}_{x} \Big(\max_{||F||_\infty \leq 1} \int_{\Omega^2} w(u,u')^{1/2}\big(x(u')-x(u)\big)F(u,u') du\, du' +$$

$$\frac{1}{2\lambda} \int_\Omega \big(x(u)-f(u)\big)^2 du \Big), \quad (1.10)$$

where F is a two-variable function and

$$\|F\|_\infty = \sup_{u\in\Omega} \Big(\int_\Omega F(u,u')^2 du' \Big)^{1/2}.$$

We note two essential differences betweeen this previous approach and the proposed model: our ability to use extended constraints on F and the combinatorial formulation, which allows our model to be valid on arbitrary graphs.

EXAMPLE 1.2. *The main idea of this second case of interest is to create flow bottlenecks in controlled areas. For example, bounding the flow in high gradient areas of an image will help us to preserve contours. The flow constraints are defined on vertices through the form of g, while the flow itself is defined on edges. We can use the operator $|A^\top|$ to average the flow exiting a vertex in order to link the two. Specifically, C may be defined as*

$$C = \{F \in \mathbb{R}^m \quad | \quad g^{\cdot 2} - |A^\top|F^{\cdot 2} \in [0,+\infty[^n\}, \quad (1.11)$$

where $g = (g_i)_{1 \le i \le n}$. We note that this particular case corresponds in (1.4) *to the case when $\alpha = 2$, and $(\forall i \in \{1, \ldots, n\})$ $\theta^{(i)}$ is the i-th line of matrix $|A^\top|$. For image filtering applications, in addition to intensity information, g may be used to penalize changes in other relevant image quantities such as color or texture. The graph may be in this case locally 4-connected, as illustrated in Fig. 1.2(a).*

The constraint given in the convex set of (1.11) *also appears in the Combinatorial Continuous Maximum Flow (CCMF) problem [12]. The problem studied in the present work may thus be seen as an extension of the CCMF problem (applied to clustering problems in graphs, such as image segmentation) to multi-label problems.*

EXAMPLE 1.3. *Instead of using only locally neighboring nodes in the regularization, we can also consider non-local connections to improve recovery results. This idea was originally developed in the non-local means denoising framework [18], and expressed as a total variation regularization in [19]. The DCTV model is adapted to this particular case, by building a graph where each node is connected to center nodes of its most similar patches. An illustration is given in Fig. 1.2(b). The main difference with the work of [19] is the definition of a discrete energy which is then optimized without any approximation.*

1.4. Related work. The Total Variation (TV) model was introduced as a regularization criterion for image denoising in the seminal work by Rudin, Osher and Fatemi (ROF) [20], and has been shown to be quite efficient for smoothing images while preserving contours. Total variation minimization has been used for solving a large number of inverse problems, including deconvolution [21], inpainting, MRI reconstruction [22], optical flow estimation [23], stereo vision [24], among others. A major advantage of the ROF-TV minimization problem is that it is convex. However, the non-differentiability of the objective function requires the use of specific optimization techniques, particularly with respect to speed and efficiency. Much progress has been achieved by employing primal-dual approaches [16, 25], the dual approach of Chambolle [26], and fast first order methods [27]. Other investigations have been pursued to develop anisotropic TV models, e.g. [28, 2].

Chan *et al.* expressed the TV problem on graphs in [29]. Improvements in the quality of the recovery were later obtained by introducing a weighted model [30, 15, 10, 14]. In this model, a discrete TV energy is optimized on edge-weighted graphs. The methods described in [30] and [14] employ a non-local weighting strategy following the non-local means concept [18]. Furthermore, Gilboa and Osher provided an efficient dual algorithm in [14], based on the method in [26] to address this problem. Currently, one of the fastest methods for optimizing classical and weighted TV is an alternating direction of multipliers method [31, 32, 33], belonging to the class of augmented Lagrangian approaches [34, 35, 36], and also called the split-Bregman algorithm. Another (node) weighting strategy and a fast primal-dual algorithm were proposed in [37].

Recently, several approaches were proposed to reduce the staircaising effect of total variation methods. The Total Generalized Variation (TGV) method of Bredies *et al.* [3, 4], following [38] which introduced the inf-convolution idea in TV, employs higher order derivatives of the function to regularize. The benefit of TGV is to replace the staircase effect by smoothed variations, while still preserving contours. The work by Jalalzai and Chambolle aims at achieving the same goal [39]. In the hybrid TV approach of Combettes and Pesquet [21] (see [40] for further investigations) an energy combining wavelets and TV is optimized to jointly reduce the staircase effect of both TV and wavelet artifacts. In addition to the staircase effect, the various TV models

typically result in some loss of contrast even when solving the model exactly [41]. A few methods have been proposed [42, 43] for contrast preservation.

By formulating our restoration problem as in (1.2), we improve on the state-of-the art. More specifically, our main contributions are the following:

(i). By using an appropriate reformulation and decomposition of our energy functional in (1.2), we are able to optimize it efficiently, owing to the use of fast parallel proximal algorithms.

(ii). By taking into account both weights on nodes and edges of the graph, we obtain a more flexible and accurate regularization, particularly for thin structures.

(iii). These improvements lead to better results, in particular a better contrast preservation and a more detailed recovery in image processing examples.

(iv). By considering a graph formulation of the primal-dual total variation problem, we are able to use our regularization method on graphs that are not uniformly sampled, for example on a graph representation of the fovea or actually any arbitrary mesh.

TV-based schemes are often optimized using one of the many proximal methods. Proximal methods provide efficient solutions to convex optimization problems where the objective function can be split into a sum of convex functions [44, 45]. This core of methods is applicable to large size problems, as well as to non-smooth and non-finite functions. The projection algorithms in [26, 14] are based on a relatively simple local constraint on the norm of the projection variable. Problem (1.2) extends the constraint on this variable while making the optimization problem more challenging. This extension allows us to better capture the local information on a graph.

Recently, other proposals for different constrained formulations have been suggested in the litterature. Combettes and Pesquet [46] introduced a total variation bound in a contrained convex energy minimization problem. Malgouyres [47] defined a constraint in order to combine the ROF-TV model with wavelet-based soft thresholding methods. Closer to what we propose is a work by Lenzen *et al.* [5, 6] where the idea of introducing a convex set constraining the dual variable in a total variation formulation is proposed. However in [5, 6], no linear operator is allowed in the degradation model, and more importantly, the formulation being given in a purely continuous setting, it does not benefit from the versatility of the graph-based framework. In their work, Hidane *et al.* [7] propose a constrained TV based regularization term using the $G-$norms of Meyer [48], that is used as a first scale in the construction of a multiscale representation of data embedded in undirected graphs.

2. Proposed algorithms. We show in this section that it is possible to efficiently solve Problem (1.2) by using recent proximal algorithms.

The choice of proximal methods [44] is motivated by the non-differentiability of our objective function. Contrary to many other proximal algorithms that could be used, such as the forward-backward algorithm (and its variants e.g. [49]), the Alternating Direction Method of Multipliers, and the inexact Uzawa method [50, 51], the algorithms we propose in Sections 2.3 and 2.5 avoid inner iterations. Furthermore, the proximity operators necessary for solving our problem can be computed in parallel, offering additional speed gains.

2.1. Convex analysis tools. We introduce in this section some convex analysis tools used in the core of this work. For a detailed overview of the main results of convex analysis, we refer to the self-contained book of Bauschke and Combettes [52].

Let $\Gamma_0(\mathbb{R}^n)$ be the class of convex functions φ from $\mathbb{R}^n$ to $]-\infty, +\infty]$ that are proper (such that $\operatorname{dom}\varphi = \{z \in \mathbb{R}^n \mid \varphi(z) < +\infty\} \neq \varnothing$) and lower semi-continuous. The proximity operator of a convex function is defined as follows:

DEFINITION 2.1. *For every $y \in \mathbb{R}^n$, the function $z \mapsto \varphi(z) + \|z - y\|^2/2$, where $\varphi \in \Gamma_0(\mathbb{R}^n)$, achieves its infimum at a* unique point *denoted by* $\operatorname{prox}_\varphi y$*, i.e.,*

$$(\forall y \in \mathbb{R}^n) \qquad \operatorname{prox}_\varphi y = \operatorname*{argmin}_{z \in \mathbb{R}^n} \; \varphi(z) + \frac{1}{2}\|z - y\|^2. \qquad (2.1)$$

The proximity operator of a convex function is a natural extension of the notion of projection onto a convex set.

DEFINITION 2.2. *Let C be a nonempty closed convex set of $\mathbb{R}^n$, the* indicator function *of C is*

$$\imath_C \colon z \mapsto \begin{cases} 0 & \text{if } z \in C, \\ +\infty & \text{otherwise.} \end{cases} \qquad (2.2)$$

The projection $\mathsf{P}_C y$ of $y \in \mathbb{R}^n$ onto C is the unique minimizer of

$$z \mapsto \imath_C(z) + \frac{1}{2}\|z - y\|^2. \qquad (2.3)$$

We have thus $\operatorname{prox}_{\imath_C} = \mathsf{P}_C$ where P_C denotes the projection onto C.

The relative interior of a convex set C will be denoted by $\operatorname{ri} C$. We also need to introduce the conjugate of a convex function:

DEFINITION 2.3. *Let φ be a convex function from $\mathbb{R}^n$ to $]-\infty, +\infty]$. Its* conjugate function *is the convex function φ^* defined as*

$$(\forall y \in \mathbb{R}^n) \qquad \varphi^*(y) = \sup_{z \in \mathbb{R}^n} (y^\top z - \varphi(z)).$$

Note that, if $\varphi \in \Gamma_0(\mathbb{R}^n)$, then $\varphi^{**} = \varphi$ and $\varphi^* \in \Gamma_0(\mathbb{R}^n)$. The conjugate function of $\imath_C$ where C is a nonempty closed convex subset of $\mathbb{R}^n$ is the *support function* σ_C of C which is expressed as

$$\sigma_C \colon \mathbb{R}^n \to]-\infty, +\infty] \colon y \mapsto \sup_{z \in C} y^\top z. \qquad (2.4)$$

The *infimal convolution* of two convex functions φ and ψ from $\mathbb{R}^n$ to $]-\infty, +\infty]$ is the convex function defined as

$$(\forall y \in \mathbb{R}^n) \quad (\varphi \square \psi)(y) = \inf_{z \in \mathbb{R}^m} \varphi(z) + \psi(y - z). \qquad (2.5)$$

We have then $(\varphi \square \psi)^* = \varphi^* + \psi^*$.

Let $\varphi \in \Gamma_0(\mathbb{R}^n)$. The *subdifferential* of φ at $x \in \mathbb{R}^n$ is denoted by $\partial\varphi(x)$. For every $(x, y) \in (\mathbb{R}^n)^2$, the following equivalence holds:

$$y \in \partial\varphi(x) \Leftrightarrow x \in \partial\varphi^*(y). \qquad (2.6)$$

2.2. Alternative formulation. In order to address Problem (1.2) where H is not necessarily injective, we consider the following optimization problem:

$$\underset{x\in\mathbb{R}^n}{\text{minimize}}\ \ \sigma_C(Dx) + \frac{1}{2}(Hx - f)^\top \Lambda^{-1}(Hx - f) + \frac{\eta}{2}\|Kx\|^2 \tag{2.7}$$

where $\eta \in]0, +\infty[$ and $K \in \mathbb{R}^{n\times n}$ is the projection matrix onto the nullspace of H, specifically, $K = I - H^\dagger H$ where $H^\dagger$ is the pseudo-inverse of H. When H is injective (rank $H = n$), the last term vanishes and (2.7) is equivalent to (1.2). The term $x \mapsto \eta\|Kx\|^2/2$ thus aims at introducing an additional regularization when H is not injective, so that the objective function remains strictly convex. The following holds:

PROPOSITION 2.4. *Assume that C is a closed convex set such that $0 \in \text{ri}\,C$. Then, Problem (2.7) admits a unique solution. The dual Fenchel-Rockafellar form of the problem is*

$$\underset{F\in\mathbb{R}^m}{\text{minimize}}\ \ \varphi(F) + \imath_C(F), \tag{2.8}$$

where $\varphi\colon F \mapsto \frac{1}{2}F^\top D\Gamma D^\top F - F^\top D\Gamma H^\top \Lambda^{-1}f$ and $\Gamma = (H^\top \Lambda^{-1}H + \eta K)^{-1}$. The optimal solution $\widehat{x}$ to the primal problem (2.7) is deduced from any optimal solution $\widehat{F}$ to the dual problem by the relation

$$\widehat{x} = \Gamma\left(H^\top \Lambda^{-1}f - D^\top \widehat{F}\right). \tag{2.9}$$

Proof. As the support function σ_C is lower semi-continuous, proper and convex, it possesses an affine minorant. In addition, $x \mapsto \frac{1}{2}(Hx - f)^\top \Lambda^{-1}(Hx - f) + \frac{\eta}{2}\|Kx\|^2$ is a strictly convex quadratic function. We deduce that

$$x \mapsto \sigma_C(Dx) + \frac{1}{2}(Hx - f)^\top \Lambda^{-1}(Hx - f) + \frac{\eta}{2}\|Kx\|^2 \tag{2.10}$$

is a function in $\Gamma_0(\mathbb{R}^n)$ which is coercive (i.e. its limit is $+\infty$ as $\|x\| \to +\infty$). This secures the existence of a solution to Problem (2.7). The uniqueness of the minimizer follows from the strict convexity of the objective function.

According to Fenchel-Rockafellar's duality theorem, we have that if $h \in \Gamma_0(\mathbb{R}^m)$ and $\psi \in \Gamma_0(\mathbb{R}^n)$ are such that

$$\text{ri}(\text{dom}\,h) \cap D\big(\text{ri}(\text{dom}\,\psi)\big) \neq \varnothing, \tag{2.11}$$

then

$$\underset{x\in\mathbb{R}^n}{\inf}\ (h(Dx) + \psi(x)) = -\underset{F\in\mathbb{R}^m}{\min}\ (\psi^*(-D^\top F) + h^*(F)). \tag{2.12}$$

Let us set $h = \sigma_C$ and

$$(\forall x \in \mathbb{R}^n) \qquad \psi(x) = \frac{1}{2}(Hx - f)^\top \Lambda^{-1}(Hx - f) + \frac{\eta}{2}\|Kx\|^2. \tag{2.13}$$

As dom $\psi = \mathbb{R}^n$ and $0 \in \text{ri}(C) \cap D(\mathbb{R}^n)$, Condition (2.11) holds. The conjugate of h is $h^* = \imath_C$ and the conjugate of ψ is

$$(\forall a \in \mathbb{R}^n) \qquad \psi^*(a) = \underset{x\in\mathbb{R}^n}{\sup}\ (a^\top x - \psi(x)). \tag{2.14}$$

Note that

$$(\forall x \in \mathbb{R}^n) \qquad \psi(x) = \frac{1}{2}x^\top \Gamma^{-1} x - f^\top \Lambda^{-1} H x + \frac{1}{2}f^\top \Lambda^{-1} f \qquad (2.15)$$

where $\Gamma = (H^\top \Lambda^{-1} H + \eta K^\top K)^{-1} = (H^\top \Lambda^{-1} H + \eta K)^{-1}$ (since K is a projection operator). Function $x \mapsto a^\top x - \psi(x)$ is strictly concave and its maximizer $\widetilde{x}$ is such that $\nabla \psi(\widetilde{x}) = a$. So,

$$\widetilde{x} = \Gamma(a + H^\top \Lambda^{-1} f) \qquad (2.16)$$

and

$$\psi^*(a) = a^\top \widetilde{x} - \psi(\widetilde{x}). \qquad (2.17)$$

Eq. (2.15) then yields the following expression of the conjugate:

$$\psi^*(a) = \frac{1}{2}(a + H^\top \Lambda^{-1} f)^\top \Gamma(a + H^\top \Lambda^{-1} f) - \frac{1}{2}f^\top \Lambda^{-1} f. \qquad (2.18)$$

Using (2.12), we deduce that the dual form of the primal problem is given by (2.8). Let $\widehat{F}$ be an optimal solution to the dual problem. According to Fermat's rule,

$$0 \in \partial(\varphi + \imath_C)(\widehat{F}). \qquad (2.19)$$

Since $\operatorname{dom} \varphi = \mathbb{R}^m$, this is equivalent to

$$
\begin{aligned}
& 0 \in \nabla\varphi(\widehat{F}) + \partial\imath_C(\widehat{F}) \\
\Leftrightarrow \quad & D\Gamma(H^\top \Lambda^{-1} f - D^\top \widehat{F}) \in \partial\imath_C(\widehat{F}) \\
\Leftrightarrow \quad & \widehat{F} \in \partial\imath_C^*\big(D\Gamma(H^\top \Lambda^{-1} f - D^\top \widehat{F})\big) \\
\Rightarrow \quad & D^\top \widehat{F} \in D^\top \partial\sigma_C(D\widehat{x}) \\
\Leftrightarrow \quad & H^\top \Lambda^{-1} f - \Gamma^{-1}\widehat{x} \in D^\top \partial\sigma_C(D\widehat{x}) \\
\Leftrightarrow \quad & 0 \in \nabla\psi(\widehat{x}) + D^\top \partial h(D\widehat{x}) = \partial(\psi + h \circ D)(\widehat{x}). \qquad (2.20)
\end{aligned}
$$

This shows that $\widehat{x}$ is the optimal solution to Problem (2.7).
□

In order to efficiently optimize the DCTV problem, the convex set C in (1.4) can generally be decomposed as an intersection of closed convex subsets $(C_r)_{1 \le r \le s}$ of $\mathbb{R}^m$, the projections onto which take closed forms. More specifically,

$$C = \bigcap_{r=1}^{s} C_r$$

$$(\forall r \in \{1, \dots, s\}) \qquad C_r = \{F \in \mathbb{R}^m \mid (\forall i \in S_r) \, \|\theta^{(i)} \cdot F\|_\alpha \le g_i\}. \qquad (2.21)$$

where $(S_r)_{1 \le r \le s}$ is a partition of $\{1, \dots, n\}$. In order to numerically solve (2.8), as $C = \bigcap_{r=1}^{s} C_r$, the indicator function $\imath_C$ can be decomposed into the sum of the indicator functions of the convex sets $(C_r)_{1 \le r \le s}$. Hence, the dual problem is equivalent to

$$\underset{F \in \mathbb{R}^m}{\text{minimize}} \ \sum_{r=1}^{s} \imath_{C_r}(F) + \varphi(F). \qquad (2.22)$$

2.3. Parallel proximal algorithm (PPXA). The above sum of $(s+1)$ convex functions can be efficiently optimized by resorting to the Parallel ProXimal Algorithm (PPXA) proposed in [21]. The choice of this particular algorithm is motivated by its flexibility in parallelizing the computation of the involved proximity operators. PPXA was generalized in [53], where links with augmented Lagrangian methods [35] were established.

Note that the dual forward-backward algorithm proposed in [54] is not directly applicable to Problem (2.8) since the projection onto C is not explicit in general. Dykstra's algorithm [55] to compute the projection onto an intersection of convex sets would not be applicable either, as the matrix $D\Gamma D^\top$ is usually singular.

Algorithm 1: Parallel ProXimal Algorithm solving (2.22)

Fix $\gamma > 0$ and $\nu \in]0, 2[$. Set $k = 0$.
Choose $y_1^{(0)} = y_2^{(0)} = \ldots = y_{s+1}^{(0)} \in \mathbb{R}^m$ and $F^{(0)} \in \mathbb{R}^m$.
repeat
    (Computation of the proximity operators in parallel)
    for $r = 1, \ldots, s+1$ **do in parallel**

$$\pi_r^{(k)} = \begin{cases} \mathsf{P}_{C_r}(y_r^{(k)}) & \text{if } r \leq s \\ (\gamma D\Gamma D^\top + I)^{-1}(\gamma D\Gamma H^\top \Lambda^{-1} f + y_{s+1}^{(k)}) & \text{otherwise.} \end{cases} \quad (2.23)$$

    (Averaging step)
    $z^{(k)} = \frac{2}{s+1}(\pi_1^{(k)} + \cdots + \pi_{s+1}^{(k)}) - F^{(k)}$
    (Updating variables in parallel)
    for $r = 1, \ldots, s+1$ **do in parallel**
        $y_r^{(k+1)} = y_r^{(k)} + \nu(z^{(k)} - \pi_r^{(k)})$
    (Updating the dual variable)
    $F^{(k+1)} = F^{(k)} + \frac{\nu}{2}(z^{(k)} - F^{(k)})$
    $k \leftarrow k + 1$
until *convergence*

We have then the following convergence result:

PROPOSITION 2.5. *Let $C = \bigcap_{r=1}^s C_r$ where, for every $r \in \{1, \ldots, s\}$, C_r is a closed convex set of $\mathbb{R}^m$ such that $0 \in \operatorname{ri} C_r$. Then, any sequence $(F^{(k)})_{k \in \mathbb{N}}$ generated by Algorithm 1 is such that*

$$x^{(k)} = \Gamma\left(H^\top \Lambda^{-1} f - D^\top F^{(k)}\right) \to \widehat{x} \quad (2.24)$$

where $\widehat{x}$ is the solution to Problem (2.7).

Proof. We have : $0 \in \bigcap_{r=1}^s \operatorname{ri} C_r = \operatorname{ri}\left(\bigcap_{r=1}^s C_r\right) = \operatorname{ri} C$. According to Proposition 2.4, there thus exists a unique solution $\widehat{x}$ to Problem (2.7). In addition, since $\operatorname{dom} \varphi = \mathbb{R}^n$,

$$0 \in \left(\bigcap_{r=1}^s \operatorname{ri} C_r\right) \cap \operatorname{dom} \varphi.$$

According to [21, Theorem 3.4&Proposition 3.6], the convergence of a sequence $(F^{(k)})_{k \in \mathbb{N}}$ generated by this algorithm to a solution $\widehat{F}$ of (2.8) is guaranteed. Using now Relation (2.9) allows us to deduce (2.24). □

Note that the robustness of PPXA w.r.t. summable numerical errors is guaranteed [21]. A potential numerical problem could however arise in (2.24), when Γ is ill-conditioned. Although this behavior was not observed in our experiments, it could be avoided by redefining K as the projection matrix onto the subspace spanned by the eigenvectors of $H^\top \Lambda^{-1} H$ corresponding to eigenvalues which are close to zero.

2.4. Choosing the decomposition of convex set C. As shown in Algorithm 1, the optimization in (2.22) requires computing projections onto each set C_r with $r \in \{1, \ldots, s\}$ that may be computed in parallel to achieve speed gains (see Definition 2.2). The convex sets as described in (2.21) obviously satisfy the assumptions made in Proposition 2.5. The main difficulty for applying PPXA is to find a partition $(S_r)_{1 \le r \le s}$ for which the projections $(\mathsf{P}_{C_r})_{1 \le r \le s}$ can be computed easily. To ensure this property, the partition can be defined in such a way that

$$(\forall r \in \{1, \ldots, s\})(\forall (i, i') \in S_r^2) \qquad i \ne i' \Rightarrow \#(\theta^{(i)} \cdot \theta^{(i')}) = 0, \qquad (2.25)$$

where $\#(a)$ denotes the number of nonzero components in a vector a. Then, it is easy to see that the projection onto C_r with $r \in \{1, \ldots, s\}$ of a vector F can be decomposed as a finite number of projections of subvectors of dimensions $\big(\#(\theta^{(i)})\big)_{i \in S_r}$ extracted from F onto weighted ℓ_α balls.

For a 4-connected lattice, when $(\forall i \in \{1, \ldots, n\})$ $\theta^{(i)}$ is the i-th line of matrix $|A^\top|$ (case of Example 1.2), a decomposition of C can be performed by setting $s = 2$. The partition $(S_r)_{1 \le r \le 2}$ corresponds to two spatial disjoint sets each with a checkered pattern (also called *red black checkerboard* pattern). The projection onto C_r with $r \in \{1, 2\}$ for $\alpha = 2$ (resp. $\alpha = 1$ or $\alpha = +\infty$) reduces to simple projections onto hyperspheres (resp. hypercubes [56]). For example, the decomposition on a small lattice when $\alpha = 2$ is illustrated by Fig. 2.1.
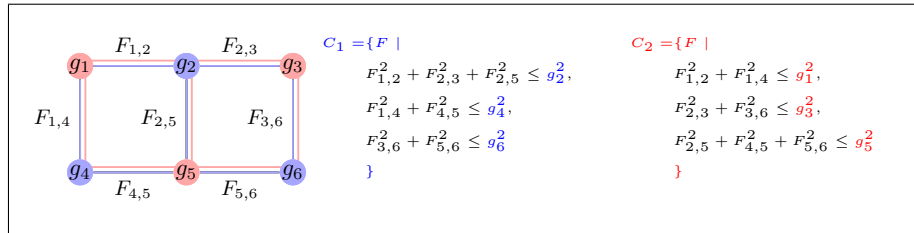

FIG. 2.1. *Decomposition of $C = C_1 \cap C_2$*

Note also that, at each iteration $k \in \mathbb{N}$, the computation of the proximity operator $\pi_{s+1}^{(k)}$ (computed in the first step of Algorithm 1, Eq. (2.23)) requires a matrix inversion. In cases when H and Λ are (or can be well approximated by) circulant-block circulant matrices, and the graph is regular, this step can be performed efficiently [57]. For this purpose, the inverse of matrix $\gamma D \Gamma D^\top + I$ is expressed by using the matrix inversion lemma, also known as Sherman-Morrison-Woodbury matrix identity. The resulting linear system then involves a circulant-block circulant matrix and is solvable efficiently using a Fast Fourier Transform.

Decomposing arbitrary graphs

In arbitrary graphs, the graph decomposition in a checkerboard pattern (Fig. 2.1) is no longer valid. So, we use a graph coloring algorithm for the decomposition. Coloring a graph with a minimum number of colors for the nodes is an NP-complete problem.

However, there exist heuristic algorithms that find a non-optimal, but feasible solution in polynomial time. In our experiments, we employed the DSAT (Degree SATuration) algorithm [58], which is guaranteed to color triangulated graphs in linear time. An illustration is provided in Fig. 2.2. Experimentally, this greedy algorithm typically decomposes the convex set C in ten sets or fewer.
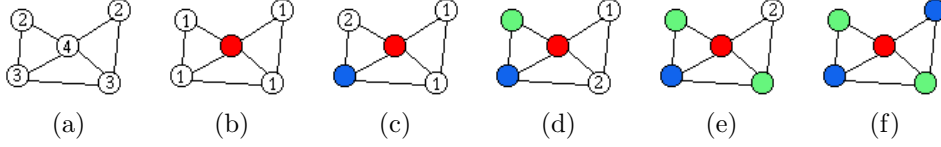

(a)   (b)   (c)   (d)   (e)   (f)

FIG. 2.2. *Coloring a graph using the DSAT algorithm [58]. (a) Sort nodes by decreasing degree. Initial labeling DSAT equal to node degrees. (b) Choose a node with maximum DSAT. Assign color number 1. Update DSAT: For each non-colored node v, if at least one neighbor of v is colored, DSAT(v) equals the number of different colors used in the neighborhood of v. (c) Choose a non-colored node of maximum DSAT. In case of conflict, choose a node of maximum degree. Assign the smallest possible color. Update DSAT. (d,e,f) Repeat the procedure in (c) until all nodes are colored.*

2.5. M+SFBF algorithm. Starting from Problem (2.7), we deduce an alternative equivalent formulation of the energy to minimize in (2.26). The formulation we obtain shall allow us to employ another parallel proximal algorithm, the Monotone + Skew Forward Backward Forward (M+SFBF) algorithm [59], which differs from PPXA. Interestingly, when H is the identity matrix and Λ is diagonal, M+SFBF does not require us to solve numerically a linear system at each iteration of the algorithm. M+SFBF belongs to the rich class of primal-dual proximal methods which also include the methods by Chen and Teboulle [60], Chambolle and Pock [37] Esser *et al.* [61], and Combettes and Pesquet [62]. We first introduce the equivalent formulation of our problem (2.7) in Proposition 2.6, and then present the M+SFBF algorithm to solve it.

PROPOSITION 2.6. *Let $C = \bigcap_{r=1}^{s} C_r$ where, for every $r \in \{1, \ldots, s\}$, C_r is a closed convex set of $\mathbb{R}^m$ such that $0 \in \operatorname{ri} C_r$. Then, the problem*

$$\underset{\substack{a_1 \in \mathbb{R}^m, \ldots, a_s \in \mathbb{R}^m, x \in \mathbb{R}^n \\ a_1 + \cdots + a_s = Dx}}{\text{minimize}} \quad \sum_{r=1}^{s} \sigma_{C_r}(a_r) + \frac{1}{2}(Hx - f)^\top \Lambda^{-1}(Hx - f) + \frac{\eta}{2}\|Kx\|^2 \quad (2.26)$$

has a solution. In addition, there exists $(\widehat{a}_1, \ldots, \widehat{a}_s) \in (\mathbb{R}^m)^s$ such that $(\widehat{a}_1, \ldots, \widehat{a}_s, \widehat{x})$ is a solution to (2.26) if and only if $\widehat{x}$ is the solution to (2.7).

Proof. Noticing that

$$\sigma_C^* = \imath_C = \sum_{r=1}^{s} \imath_{C_r} = \sum_{r=1}^{s} \sigma_{C_r}^* = (\sigma_{C_1} \square \cdots \square \sigma_{C_s})^*, \quad (2.27)$$

where $\square$ denotes the infimal convolution (see (2.5)), we deduce that

$$\sigma_C = (\sigma_C^*)^* = \sigma_{C_1} \square \cdots \square \sigma_{C_s}. \quad (2.28)$$

We have then

$$(\forall a \in \mathbb{R}^m) \quad \sigma_C(a) = \underset{\substack{a_1 \in \mathbb{R}^m, \ldots a_s \in \mathbb{R}^m \\ a_1 + \cdots + a_s = a}}{\inf} \sigma_{C_1}(a_1) + \cdots + \sigma_{C_s}(a_s). \quad (2.29)$$

In addition, since $\bigcap_{r=1}^{s} \operatorname{ri} C_r \neq \varnothing$, the above infimum is attained (see [63, Theorem 16.4]). From this, we deduce that (2.7) can be reformulated as in (2.26). The existence of a solution to (2.26) results from Proposition 2.4. $\square$

We see that the proposed approach amounts to decomposing the gradient Dx of a candidate solution x in a sum of components $(a_r)_{1 \leq r \leq s}$ onto which a simpler regularization is performed. In particular, as in the previous section, we choose the decomposition of C in such a way that each C_r is the product of convex sets in a lower-dimensional space, i.e. $\times_{i \in S_r} C_r^i$ (for example, C_r^i is a hypersphere of $\mathbb{R}^4$). In this case, $\sigma_{C_r^i}$ with $i \in S_r$ is a "local" sparsity measure, in the sense that the proximity operator of each $\sigma_{C_r^i}$ is a proximal thresholder [64]:

$$\operatorname{prox}_{\sigma_{C_r^i}} a_r^i = 0 \Leftrightarrow a_r^i \in C_r^i. \tag{2.30}$$

This interpretation provides more insight in the choice for C.

Another positive point of the formulation in (2.26) is that it allows us to propose an alternative parallel algorithm to solve the problem. More precisely, we propose to employ Algorithm 2 which is derived from the M+SFBF algorithm in [59].

We have then the following convergence result:

PROPOSITION 2.7. *Under the same assumption as in Proposition 2.6, any sequence* $(a_1^{(k)}, \ldots, a_s^{(k)}, x^{(k)})_{k \in \mathbb{N}}$ *generated by Algorithm 2 converges to a solution to Problem* (2.26).

Proof. Problem (2.26) takes the form of the dual problem in [59, Proposition 4.4]

$$\underset{\substack{a_1 \in \mathbb{R}^m, \ldots, a_s \in \mathbb{R}^m a_{s+1} \in \mathbb{R}^n \\ (s+1)^{-1} \sum_{r=1}^{s+1} L_r^\top a_r = 0}}{\operatorname{minimize}} \sum_{r=1}^{s+1} \varphi_r^*(a_r) \tag{2.31}$$

where we have set $a_{s+1} = x$, $L_1 = \ldots = L_s = (s+1)I$, $L_{s+1} = -(s+1)D^\top$, $(\forall r \in \{1, \ldots, s\})$ $\varphi_r = \iota_{C_r}$, $\varphi_{s+1} = \psi^*$. In a finite dimensional space, the qualification condition in [59, Remark 4.5] is satisfied if

$$(\exists a \in \mathbb{R}^m)(\forall r \in \{1, \ldots, s+1\}) \qquad L_r a \in \operatorname{ri}(\operatorname{dom} \varphi_r). \tag{2.32}$$

The latter condition holds since $0 \in \cap_{r=1}^{s} \operatorname{ri} C_r$ and $\operatorname{dom} \psi = \mathbb{R}^n$. The convergence of the sequence $(a_1^{(k)}, \ldots, a_s^{(k)}, x^{(k)})_{k \in \mathbb{N}}$ to a solution to (2.31) then follows from [59, Proposition 4.4]. $\square$

One additional advantage of this primal-dual algorithm is that the convergence of $(a_1^{(k)}, \ldots, a_s^{(k)}, x^{(k)})_{k \in \mathbb{N}}$ to a solution to Problem (2.26) is guaranteed even in the limit case when $\eta = 0$.

3. Results. Our validation intends to demonstrate the improvements of DCTV over state-of-the-art approaches in image denoising in terms of signal-to-noise ratio and contrast preservation, while keeping low computation times and suitability to a variety of problems. To the best of the our knowledge, [19] represents the state-of-the-art in variational image restoration. Our experiments are organized in three parts.

  (i). First we evaluate the DCTV denoising performance in comparison with the state-of-the-art. The first experiment compares TV, weighted TV, and DCTV results using different norms for the convex set that constrains the flow variable. A second experiment confirms the SNR improvements of DCTV over

---

Algorithm 2: The M+S Forward-Backward-Forward algorithm solving (2.26)

Fix $\varepsilon \in\ ]0, \big((s+1)\beta + 1\big)^{-1}[$ where $\beta = \max\{1, \|D\|\}$. Set $k = 0$.

Choose $(F_r^{(0)})_{1 \le r \le s+1}$ and $(a_r^{(0)})_{1 \le r \le s}$ in $\mathbb{R}^m$ and $x^{(0)} \in \mathbb{R}^n$.

repeat

$\quad F^{(k)} = \frac{1}{s+1} \sum_{r=1}^{s+1} F_r^{(k)}$

$\quad \gamma^{(k)} \in [\varepsilon(s+1), (1-\varepsilon)/\beta]$

$\quad$ Updating steps in parallel

$\quad$ **for** $r = 1, \dots, s+1$ **do in parallel**

$\qquad G_r^{(k)} = \begin{cases} F_r^{(k)} - \gamma^{(k)} a_r^{(k)} & \text{if } r \le s \\ F_{s+1}^{(k)} + \gamma^{(k)} D x^{(k)} & \text{otherwise.} \end{cases}$

$\qquad b_r^{(k)} = \begin{cases} a_r^{(k)} + \gamma^{(k)} F_r^{(k)} & \text{if } r \le s \\ x^{(k)} - \gamma^{(k)} D^\top F_{s+1}^{(k)} & \text{otherwise.} \end{cases}$

$\quad$ Averaging step

$\quad \Pi^{(k)} = \frac{1}{s+1} \sum_{r=1}^{s+1} G_r^{(k)}$

$\quad$ **for** $r = 1, \dots, s+1$ **do in parallel**

$\qquad$ Computation of proximity operators in parallel

$\qquad \pi_r^{(k)} =$
$\qquad \begin{cases} b_r^{(k)} - \gamma^{(k)} \mathsf{P}_{C_r}(\gamma^{(k)-1} b_r^{(k)}) & \text{if } r \le s \\ \big(\gamma^{(k)}(H^\top \Lambda^{-1} H + \eta K) + I\big)^{-1}(\gamma^{(k)} H^\top \Lambda^{-1} f + b_{s+1}^{(k)}) & \text{otherwise.} \end{cases}$

$\qquad$ Updating steps in parallel

$\qquad G_r^{(k)'} = \begin{cases} \Pi^{(k)} - \gamma^{(k)} \pi_r^{(k)} & \text{if } r \le s \\ \Pi^{(k)} + \gamma^{(k)} D \pi_{s+1}^{(k)} & \text{otherwise.} \end{cases}$

$\qquad b_r^{(k)'} = \begin{cases} \pi_r^{(k)} + \gamma^{(k)} \Pi^{(k)} & \text{if } r \le s \\ \pi_{s+1}^{(k)} - \gamma^{(k)} D^\top \Pi^{(k)} & \text{otherwise.} \end{cases}$

$\qquad F_r^{(k+1)} = F_r^{(k)} - G_r^{(k)} + G_r^{(k)'}$

$\qquad \begin{cases} a_r^{(k+1)} = a_r^{(k)} - b_r^{(k)} + b_r^{(k)'} & \text{if } r \le s \\ x^{(k+1)} = x^{(k)} - b_{s+1}^{(k)} + b_{s+1}^{(k)'} & \text{otherwise.} \end{cases}$

$\quad k \leftarrow k + 1$

until *convergence*

---

weighted TV for the recovery of several images corrupted with different noise variances. We repeat the second experiment in non-local graphs to gain additional SNR improvements.

(ii). Secondly, we evaluate computation times for the differents algorithms we employed.

(iii). We finally show that the scope of DCTV is not limited to restoration problems and it also leads to improved results in tasks such as image fusion and mesh filtering.

3.1. Image denoising performance.

3.1.1. Local denoising. In our first application, we demonstrate the performance of DCTV with respect to weighted TV of (**??**) for image denoising. In Section 3.1.1, the considered graph connects each node corresponding to an image value

| (a) Original image | (b) Noisy image
SNR = 19.1 dB | (c) TV
SNR =29.2 dB | (d) Weighted TV
SNR = 29.7 dB |

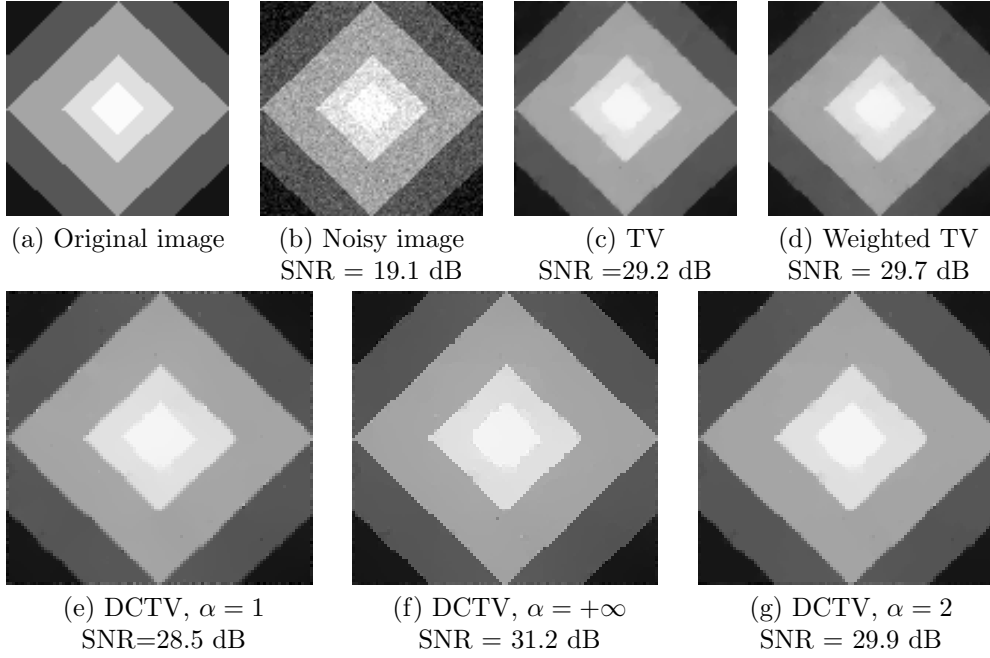| (e) DCTV, $\alpha = 1$
SNR=28.5 dB | (f) DCTV, $\alpha = +\infty$
SNR = 31.2 dB | (g) DCTV, $\alpha = 2$
SNR = 29.9 dB |

FIG. 3.1. *Denoising an image for constant levels of additive Gaussian noise. (a) Original image, (b) Noisy image, (c) TV [26] $\lambda = 0.06$, (d) Weighted TV on a 4-connected graph, see (1.8) , weights given by the gradients of the image ($\lambda = 0.06$, $\epsilon = 0.2$) , (e) PPXA with norm ℓ_1 ($\lambda = 0.8$, $\epsilon = 0.35$, projections onto an ℓ_1 hypercube performed using the software of [56]), (f) PPXA with norm ℓ_∞ ($\lambda = 0.4$, $\epsilon = 0.3$), (g) PPXA with norm ℓ_2 ($\lambda = 0.33$, $\epsilon = 0.16$).*

to its neighbors in a 4-connected local neighborhood. We recall that, for image denoising, the DCTV problem we are considering is expressed as

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \left(\sup_{\substack{F\in\mathbb{R}^m \\ (\forall i\in\{1,\dots,n\}) \, \|\theta^{(i)}\cdot F\|_\alpha \leq g_i}} F^\top Dx + \frac{1}{2\lambda}\|x - f\|^2 \right), \tag{3.1}$$

where $\theta^{(i)} \in \mathbb{R}^m$ is an indicator vector of the edges connected to node i in the graph, and D is the weighted incidence matrix defined just after (1.3).

Concerning the choice of the node weight g_i at the i-th node, a simple strategy consists of considering a monotonically decreasing function of the data gradient. More specifically, given positive reals ϵ and χ, we suggest using g defined as

$$(\forall i \in \{1,\dots,n\}) \qquad g_i = \exp(-\chi\|\nabla\widetilde{x}_i\|_2) + \epsilon, \tag{3.2}$$

where $\widetilde{x}$ is some reference data defined on a graph, for instance noisy measurements f or image data which corresponds to some rough estimate of the original image $\bar{x}$. The Euclidean norm of its discrete gradient $\nabla\widetilde{x}_i$ at node i can be computed as

$$\|\nabla\widetilde{x}_i\|_2 = \begin{cases} (\sum_{j\in N_i}(\widetilde{x}_i - \widetilde{x}_j)^2)^{\frac{1}{2}} & \text{if } i \in J \\ 0 & \text{otherwise.} \end{cases} \tag{3.3}$$

In the absence of a contour, g_i takes large values, so are the components of F corresponding to nonzero components of $\theta^{(i)}$, preventing large local variations of x in the

minimization (1.2). Conversely, in the presence of a contour, $g_i \simeq \epsilon$, and the components of F corresponding to nonzero components of $\theta^{(i)}$ are small, thus allowing large local variations of x. An illustration is given in Fig. 1.2(a).

In Fig. 3.1, we compare denoising results on a synthetic image, using different TV-based formulations for a denoising task where the noise is assumed to be zero-mean Gaussian and white. The general problem (2.7) is solved using $\eta = 0$, $H = I$, $\Lambda = \lambda I$ with $\lambda \in]0, +\infty[$, thus being equivalent to the problem recalled in (3.1). The node weights are set according to (3.2) with $\chi = 0.04$ and edge weights are set to one. The stopping criterion for the algorithms is $\|x^{(k)} - x^{(k-1)}\| < 5.10^{-3}\|x^{(k)}\|$. The results obtained with weighted TV [10] (see (1.8)) look slightly better than using the classical, unweighted TV model [26]. Performing the DCTV optimization with PPXA using the convex set defined in (1.11) ($\alpha = 2$) further improves the results. The choice of the ℓ_∞ norm allows us to achieve better results than all the other variational approaches we tried on this synthetic image. Contrary to the ℓ_2 norm that has a smoothing effect, the ℓ_∞ norm allows us to retrieve sharper contours. However, this effect is no longer present in non-synthetic images, where the ℓ_2 norm is the best choice.

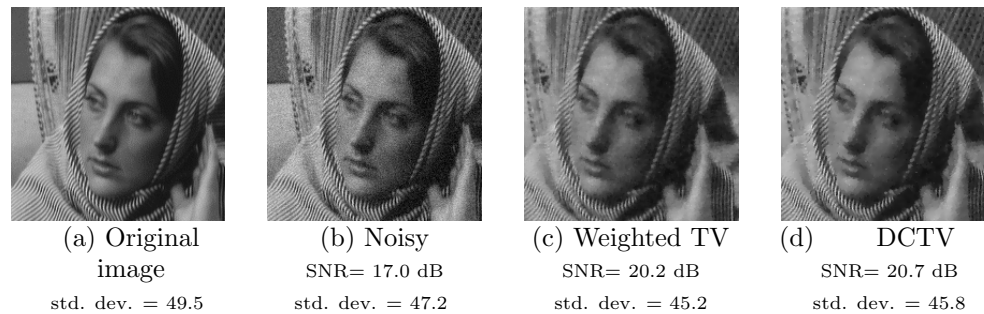| (a) Original image	(b) Noisy	(c) Weighted TV	(d) DCTV
	SNR= 17.0 dB	SNR= 20.2 dB	SNR= 20.7 dB
std. dev. = 49.5	std. dev. = 47.2	std. dev. = 45.2	std. dev. = 45.8

FIG. 3.2. *Denoising the 'barbara' image (zoom) corrupted with Gaussian noise with variance* $\sigma^2 = 15$.

In a second experiment, we compare quantitatively DCTV – using the Euclidean norm ($\alpha = 2$) – with weighted TV. We used four standard test images that we corrupted with zero-mean white Gaussian noise with variance σ^2. The value of the image fidelity parameter λ was set according to a grid search depending on the variance of the noise, and ϵ was set to λ. Signal-to-Noise Ratio (SNR) is used as the performance measure in our quantitative evaluation. Table 3.1 reports SNR values for DCTV and weighted TV results obtained on each image corrupted with different values of noise variance. Result examples are shown in Fig. 3.2. Those experiments show that DCTV leads to improved results when the variance of the noise is lower than 50. Visually, DCTV results are sharper and feature better contrast than the weighted TV results. The better contrast preservation can be quantified by computing the standard deviation of the pixel values, as shown in Fig. 3.4.

3.1.2. Nonlocal denoising. Rather than using locally connected graphs, non-local strategies [18, 19] can be employed which have been shown to achieve denoising improvements. A nonlocal strategy may naturally be employed in the DCTV framework. The weights between non-neighbor nodes are computed following the main idea of [18, 19]: for each pixel i of the image, edges are added between i and the nodes corresponding to most similar patches, as depicted in Fig. 1.2(b). We build edge

σ^2	5	10	15	20	25	50
SNR values obtained by optimizing weighted TV						
house	32.5	28.8	27.0	25.7	25.0	21.6
man	27.4	23.8	22.0	20.7	19.6	17.2
lena	31.5	27.7	25.9	25.0	24.2	21.2
barbara	27.1	22.8	20.2	18.5	17.3	15.2
mean	29.9	26.0	24.0	22.7	21.8	**18.8**
SNR values obtained by optimizing DCTV						
house	32.6	29.2	27.3	25.7	25.1	21.6
man	27.9	24.5	22.4	20.9	19.8	17.0
lena	31.9	28.5	26.5	25.1	24.3	21.1
barbara	27.1	23.0	20.7	19.0	17.7	15.1
mean	**30.1**	**26.6**	**24.5**	**22.9**	**22.0**	18.7

σ^2	5	10	15	20	25	50
SNR values obtained by optimizing **NonLocal** DCTV						
The same SNR values are obtained by optimizing NLTV						
house	33.7	30.3	28.5	26.9	25.9	21.4
man	28.7	24.3	21.6	19.7	18.6	14.9
lena	31.7	28.9	26.7	25.0	23.9	20.1
barbara	28.4	24.8	22.6	20.8	19.5	15.5
mean	30.6	27.1	24.8	23.1	22.0	18.0

TABLE 3.1

Quantitative denoising experiment on standard images corrupted with additive Gaussian noise with variance σ^2. The regularization parameter λ was chosen empirically to give the best results for each method.

weights as proposed in [19],

$$w_{i,j} = \exp\left(-\frac{1}{h^2}\sum_z G_a(z)(\widetilde{x}_{i+z} - \widetilde{x}_{j+z})^2\right), \tag{3.4}$$

where $\sum_z G_a(z)(\widetilde{x}_{i+z} - \widetilde{x}_{j+z})^2$ represents the distance between patches of center i and j, G_a is a Gaussian function with standard deviation a, and h is a real number which acts as a scale parameter.

As a result of the graph construction, the Laplacian related to the matrix to be inverted (at each iteration of PPXA) of this nonlocal graph may be no longer circulant block circulant, resulting in loss of speed efficiency. We propose to employ the M+SFBF algorithm (Algorithm 2) in this case.

We provide some quantitative – in Table 3.1– and qualitative – in Fig. 3.3 – results. We observe similar results obtained with our dual model in comparison with the state-of-the-art weighted TV optimized using the split-Bregman algorithm [19]. We note that our model offers stronger convergence guaranties than the split-Bregman approach implemented in [19]. Instead of solving a Poisson problem exactly, [19] computes an approximation for speed issues, that requires the choice of a regularization constant and of a number of inner iterations to converge. We also give quantitative measurements in terms of contrast preservation in Fig. 3.4, showing that the local DCTV approach always improves contrast over the weighted TV in local or non-local

(a) Original image   (b) Noisy SNR=14.47dB

(c) Nonlocal TV SNR=20.76dB   (d) Nonlocal DCTV SNR=20.78dB

Original   Noisy   NonLocal TV   NonLocal DCTV

FIG. 3.3. *Local and nonlocal denoising (Gaussian noise of variance $\sigma^2 = 20$). The nonlocal weights are computed using [19]. Regularization parameters used to obtain the best results: (e) $\lambda = 0.048$ (f) $\lambda = 0.093$.*

graphs.

3.2. Computation times. In terms of computation time, DCTV using PPXA is competitive with the most efficient weighted TV algorithm. Denoising the 512×512 Lena image corrupted with Gaussian noise ($\sigma^2 = 15$) requires 0.4 seconds for split-Bregman, versus 0.7 seconds for PPXA to converge on an Intel Xeon 2.5GHz 8-core system. Note that as explained in [19], the energy optimized by the split-Bregman method only corresponds to an approximation of the weighted TV energy.
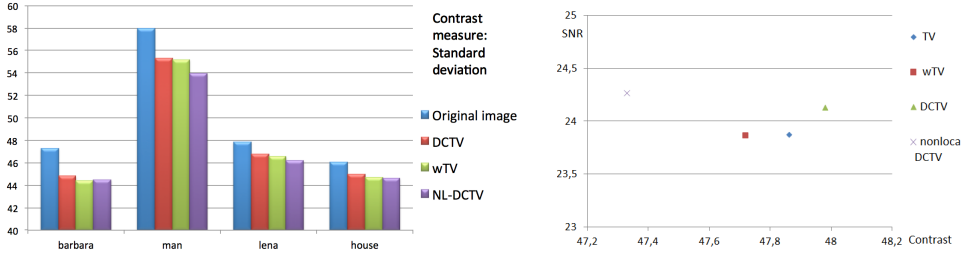
FIG. 3.4. *Left hand side: Standard deviation of each test image compared with the standard deviation of the denoising results. We computed the mean of the standard deviation over the six different results given by DCTV and weighted TV for various noise variances ($\sigma^2 = 5, 10, 15, 20, 25, 50$). Right hand side: mean over the experiments (4 images corrupted with 6 different noise intensity) for the different denoising strategies.*

3.2.1. Comparison of applicability of M+SFBF and PPXA. The proximity operator associated with the data fidelity term in PPXA requires solving a linear system. In the case of an image recovery task using a local, regular graph (for example a 4-connected grid) as in the first part of Section 3.1, the special configuration of the matrix in the system may be taken into account. As this resulting matrix is circulant block-circulant, the system is efficiently solved using a Fast Fourier Transform. However, in the case of an arbitrary graph, for example a nonlocal graph as in Section 3.1.2, the matrix of the linear system is no longer circulant, so the system can instead be solved using a conjugate gradient method, for instance.
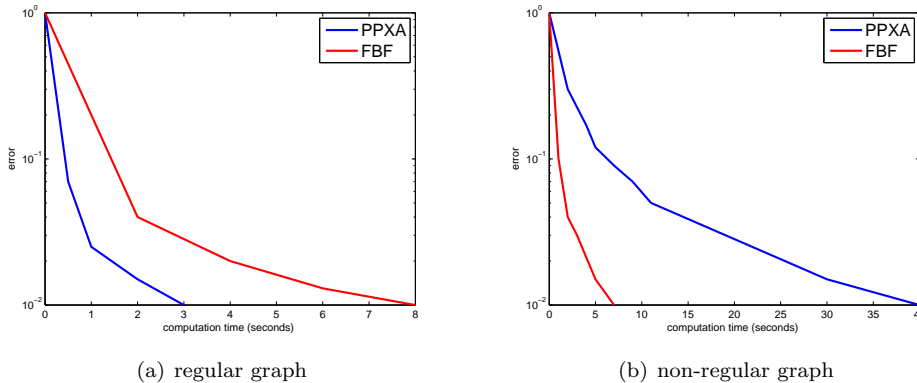

(a) regular graph　　　　　　　　(b) non-regular graph

FIG. 3.5. *Comparison of convergence speed of the PPXA (blue) and M+SFBF (red) algorithms for denoising an image.(a) using a regular, 4-connected lattice, (b) using a non-regular graph. In regular graphs, PPXA is the fastest algorithm thanks to the use of FFT. However, another linear solver has to be employed in PPXA in non-regular graphs. As a result, the M+SFBF algorithm is faster in this case.*

When using the M+SFBF algorithm, the proximity operator of the function enforcing the data fidelity does not require us to solve a linear system provided that H is the identity matrix. Instead, four matrix multiplications by the weighted incidence matrix are performed per iteration. The speed of convergence using M+SFBF and PPXA are compared in Fig. 3.5, for denoising an image using a local, and a nonlocal graph. In the figure, only computation times are presented because the required num-

ber of iterations to reach convergence is about the same for either algorithms. In a local graph, the matrix multiplications in M+SFBF make the algorithm slower than PPXA. Conversely, in a nonlocal graph, where the linear system of PPXA is solved using a conjugate gradient, M+SFBF converges much faster.

3.2.2. Fast image deconvolution. We now give some quantitative comparison results for joint denoising and deblurring tasks.
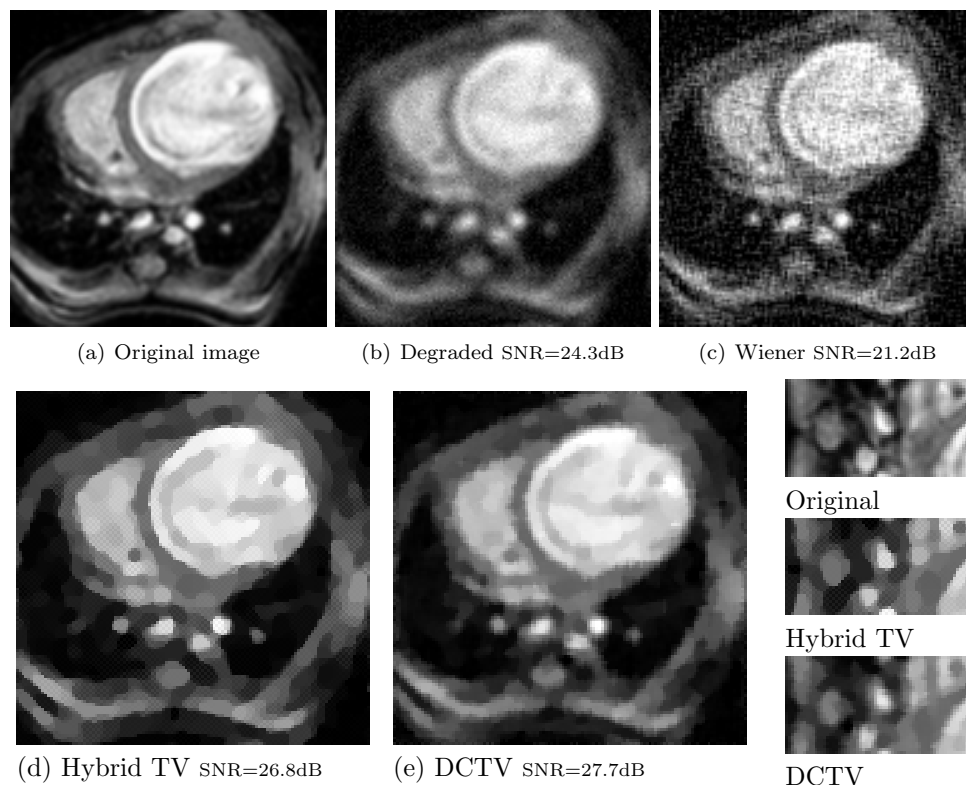

(a) Original image   (b) Degraded SNR=24.3dB   (c) Wiener SNR=21.2dB

(d) Hybrid TV SNR=26.8dB   (e) DCTV SNR=27.7dB   DCTV

Original

Hybrid TV

FIG. 3.6. *Denoising and deblurring an MRI image corrupted with synthetic Gaussian 5×5 blur and Gaussian noise ($\sigma^2 = 10$). We observe in the hybrid TV result the presence of checkerboard artifacts due to the use of discrete filters for approximating the gradient. DCTV reduces the staircase effect of TV while preserving more details. Parameter used: Hybrid TV [21] with regularization parameters $\alpha = 0$ and $\beta = 0.025$, and DCTV with $\lambda = 0.005$ and $\eta = 0.04$.*

We compare the DCTV results to Wiener based deconvolution using the Matlab function ``deconvwrn''. The comparison also includes the hybrid TV/wavelet regularization method of Combettes and Pesquet [21]. We report in Table 3.2 the SNR values for restoration of images corrupted with additive zero-mean white Gaussian noise with variances $\sigma^2 = 5$, and $\sigma^2 = 10$, and convolved with uniform blur kernels of sizes 5×5 and 7×7. We observe that DCTV unsurprisingly outperforms the standard Wiener filter. More significantly, DCTV is competitive with the recent method in [21] both quantitatively in terms of SNR, and qualitatively. Its results are free of the checkerboard artifacts observed for the Hybrid TV method in Fig. 3.6. Furthermore, the results were obtained twice as fast, for the same number of iterations, using a Matlab implementations for both methods.

3.3. Generality of the framework: beyond image restoration.

(σ^2, b)	(5,5)	(5,7)	(10,5)	(10,7)
SNR values obtained using **Wiener** deconvolution				
house	19.2	19.8	18.8	19.2
lena	18.2	18.9	18.1	18.8
barbara	14.7	14.5	14.2	14.0
man	15.0	15.6	15.2	15.9
mean	**16.8**	**17.2**	**16.5**	**17.2**
SNR values obtained using **Hybrid TV** [21]				
house	26.2	25.3	24.9	23.8
lena	24.9	23.8	23.9	22.9
barbara	16.4	15.7	15.8	15.4
man	21.6	20.6	20.6	19.6
mean	**22.3**	**21.4**	**21.3**	**20.4**
SNR values obtained by optimizing **DCTV**				
house	26.3	24.7	24.6	23.3
lena	25.5	23.9	24.0	22.8
barbara	16.9	15.8	16.0	15.5
man	22.1	20.6	20.8	19.6
mean	**22.7**	**21.3**	**21.4**	**20.3**

TABLE 3.2

Quantitative deconvolution experiment on standard images corrupted with additive white Gaussian noise with variance σ^2, and a Gaussian blur of kernel size $b \times b$. The DCTV weights have been set according to (3.2), where the estimate of the image $\bar{x}$ has been precomputed using DCTV with the corrupted image f. This trick allows us to obtain an SNR improvement of about 0.5 dB, while computation times are still twice as fast than the hybrid TV method [21].

3.3.1. Image fusion. We show here how to apply our Dual Constrained TV approach to image fusion. From N degraded images $f_1, \ldots, f_N$, the problem is to find a restored image $\hat{x}$. The degradation model is the following:

$$\begin{bmatrix} f_1 \\ \vdots \\ f_N \end{bmatrix} = \begin{bmatrix} H_1 \\ \vdots \\ H_N \end{bmatrix} \bar{x} + \begin{bmatrix} b_1 \\ \vdots \\ b_N \end{bmatrix}, \tag{3.5}$$

where $H_1, ..., H_N$ are degradation matrices (e.g. generating blur), and $b_1, ..., b_N$ represent zero-mean white Gaussian noises, possibly with different variances, assumed to be uncorrelated. The problem can be formulated as in (1.2) by setting

$$f = \begin{bmatrix} f_1 \\ \vdots \\ f_N \end{bmatrix}, \quad H = \begin{bmatrix} H_1 \\ \vdots \\ H_N \end{bmatrix}, \quad \Lambda = \begin{bmatrix} \lambda_1 I & 0 & \ldots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \lambda_N I \end{bmatrix} \tag{3.6}$$

with $(\lambda_1, \ldots, \lambda_N) \in]0, +\infty[^N$. We present in Fig. 3.7 an example of fusion between a blurry image and a noisy image in a 4-connected lattice ($N = 2$ and $H_2 = I$).

On a rectangular grid, where the convex C is decomposed as described in Fig. 2.1, the parallel proximal algorithm given in Algorithm 1 can be employed efficiently for this task.

(a) Original image   (b) Noisy image SNR=17.3dB   (c) Blurry image SNR=23.9dB

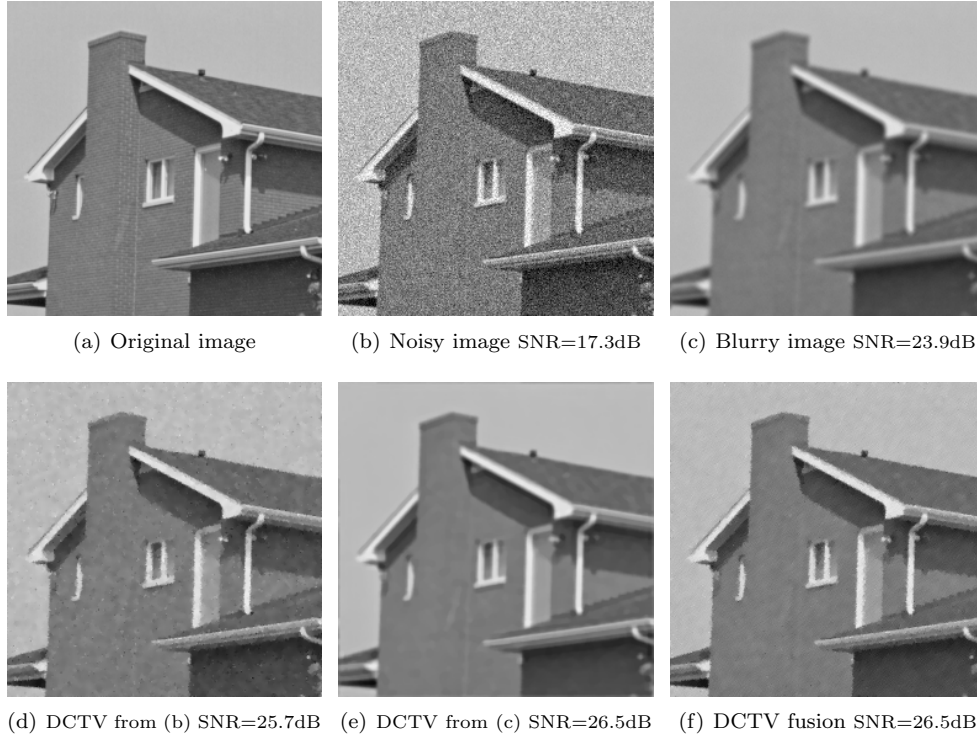(d) DCTV from (b) SNR=25.7dB   (e) DCTV from (c) SNR=26.5dB   (f) DCTV fusion SNR=26.5dB

FIG. 3.7. *Image fusion from an image corrupted with Gaussian noise with variance $\sigma_2^2 = 20$ (b) and a blurry image (c) (uniform blur kernel of size 5×5, with additive Gaussian noise with variance $\sigma_1^2 = 1$). (d) DCTV result obtained using only the noisy image. (e) DCTV result obtained using only the blurry image. (f) DCTV fusion result with $\lambda_1 = 0.16$ and $\lambda_2 = 0.21$. SNR values are identical for (e) and (f), however the fusion result is visibly sharper.*

3.3.2. 3D mesh filtering. Since the dual-constrained TV-based formulation is defined on arbitrary graphs, x is not limited to represent only image pixel values. In Fig. 3.8, we present an example of mesh denoising, where $\bar{x}$ is a vector composed of the spatial coordinates $\bar{x}_X$, $\bar{x}_Y$ and $\bar{x}_Z$ of the mesh nodes. In this experiment, we added a randomly oriented noise vector with zero-mean white Gaussian magnitudes to the original node coordinates of a mesh. This results in noisy mesh nodes of coordinates f_X, f_Y and f_Z. The degradation model is the following:

$$f = \begin{bmatrix} f_X \\ f_Y \\ f_Z \end{bmatrix} = \begin{bmatrix} \bar{x}_X \\ \bar{x}_Y \\ \bar{x}_Z \end{bmatrix} + \sigma^2 \begin{bmatrix} b_X \\ b_Y \\ b_Z \end{bmatrix}, \tag{3.7}$$

where b_X, b_Y, b_Z represent uncorrelated vectors of additive noises with unit magnitude variance.

For this application, we used the M+SFBF algorithm, which is more efficient than PPXA as the graph is not regular. This application shows that the DCTV framework is also well suited for regularizing this type of data. We compare our denoising result to a typical mesh smoothing technique, and show that DCTV outperforms Laplacian smoothing.
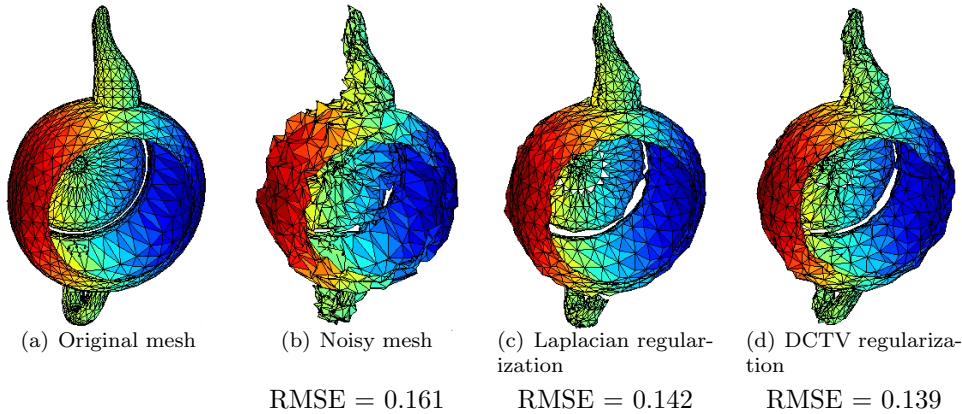
(a) Original mesh	(b) Noisy mesh	(c) Laplacian regular- ization	(d) DCTV regulariza- tion
	RMSE = 0.161	RMSE = 0.142	RMSE = 0.139

FIG. 3.8. *Example of mesh denoising using DCTV on the spatial coordinates of the nodes. We use the Root Square Mean Error to measure an average distance to the original mesh. (c) Laplacian smoothing (d) The M+SFBF algorithm was used to optimize DCTV with $\lambda = 0.25$. The Root Mean Square Error measures a distance between the result and the original mesh.*

3.3.3. Biologically sampled image filtering. This final example demonstrates the ability of DCTV to efficiently filter data on arbitrary graphs. In some situations, images are acquired from nonuniform samples. Specifically, most biological systems are known to acquire light nonuniformly. Following the work of [65], in Fig. 3.9, we used the Graph Analysis Toolbox [66] that contains implementations of several filtering techniques in arbitrary graphs. From a graph representing the spatial resolution of the Bottlenosed dolphin visual system, the image (c) represents an image resampled with the dolphin's mesh, corresponding to the input f of our method. It is then straightforward to run the M+SFBF algorithm, using the incidence matrix of the graph represented in (b).

Results obtained with DCTV are compared with results obtained using the low-pass filtering method of Taubin *et al.* [67]. We note that DCTV improves contrast preservation and thin objects rendering compared with Taubin's method.

4. Conclusion. In this work we have extended existing TV models by generalizing the constraint on the projection variable of the dual TV formulation. This new approach shows improved results compared with the weighted TV approach in image restoration applications. The advantage of our DCTV formulation over unconstrained formulations is that it allows us to take into account both edges and node weights, as well as more general constraints on the flow variable. For local image denoising, the induced bottleneck that arises in the nodes in the presence of a contour helps in blocking the diffusion present in classical TV approaches. As shown in our experiments, this behaviour results in better contrast preservation in addition to SNR improvements. Secondly, our dual formulation allows us to employ fast optimization strategies without requiring approximations. Finally, the versatility of DCTV allows us to exploit extra information, and may be used in different contexts.

More generally, the proposed proximal algorithms make it possible to efficiently solve convex minimization problems involving the support function of an intersection of convex sets as a penalty term. It is also worth emphasizing that this approach can be applied to any graph data structures, in particular those frequently employed in 3D modeling.

(a) Original
image

(b) Bottlenosed
dolphin structure

(c) Sampled
image

(d) Noisy sampled
SNR = 22.1 dB

(e) Taubin filtered
result [67] SNR = 19.4 dB

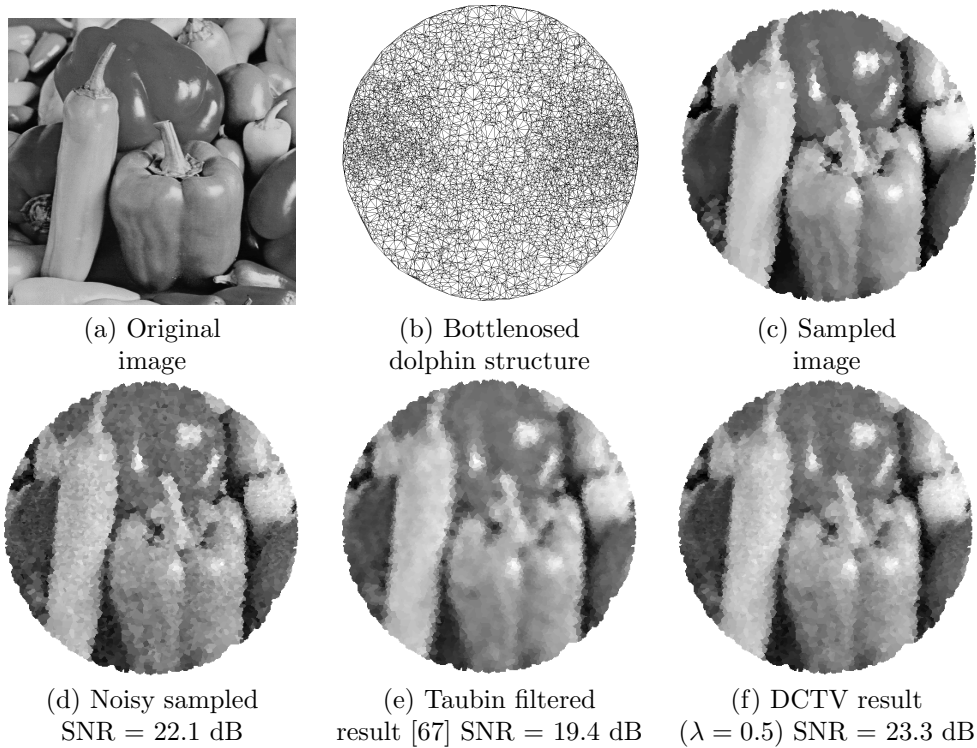(f) DCTV result
($\lambda = 0.5$) SNR = 23.3 dB

FIG. 3.9. *Filtering image data on a biologically sampled image [66]. Noise with variance $\sigma^2 = 10$ was added to the resampled values of the image (c) to produce (d).*

Future work will address choices of convex sets that are appropriate to the type of data present in graphs that need to be regularized. Source code is available at `http://sourceforge.net/projects/dctv`.

REFERENCES

[1] Couprie, C., Talbot, H., Pesquet, J.C., Najman, L., Grady, L.J.: Dual constrained TV-based regularization. In: Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP'11, Prague, Czech Republic (2011) 945–948

[2] Osher, S.J., Esedoglu, S.: Decomposition of images by the anisotropic rudin-osher-fatemi model. Comm. Pure Appl. Math **57** (2003) 1609–1626

[3] Bredies, K., Kunisch, K., Pock, T.: Total generalized variation. SIAM J. Imaging Sci. **3** (2010) 492–526

[4] Bredies, K., Kunisch, K., Valkonen, T.: Properties of l^1 - TGV^2 : The one-dimensional case. (2011) Submitted for publication.

[5] Lenzen, F., Becker, F., Lellmann, J., Petra, S., Schnörr, C.: Variational image denoising with adaptive constraint sets. In Bruckstein, A.M., ter Haar Romeny, B.M., Bronstein, A.M., Bronstein, M.M., eds.: SSVM. Volume 6667 of Lecture Notes in Computer Science., Springer (2011) 206–217

[6] Lenzen, F., Becker, F., Lellmann, J., Petra, S., Schnorr, C.: A class of quasi-variational inequalities for adaptive image denoising and decomposition. Computational Optimization and Applications (2012)

[7] Hidane, M., Lézoray, O., Elmoataz, A.: Nonlinear multilayered representation of graph-signals. J. Math. Imaging Vision **45** (2013) 114–137

[8] Hirani, A.N.: Discrete exterior calculus. PhD thesis, Cal. Tech (2003)

[9] Grady, L.J., Polimeni, J.R.: Discrete Calculus: Applied Analysis on Graphs for Computational

Science. Springer-Verlag, New York (2010)

[10] Elmoataz, A., Lézoray, O., Bougleux, S.: Nonlocal discrete regularization on weighted graphs: A framework for image and manifold processing. IEEE Trans. Image Process. **17** (2008) 1047–1060

[11] Grady, L., Alvino, C.: The piecewise smooth Mumford-Shah functional on an arbitrary graph. IEEE Trans. Image Process. **18** (2009) 2547–2561

[12] Couprie, C., Grady, L., Talbot, H., Najman, L.: Combinatorial Continuous Maximum Flow. SIAM J. Imaging Sci. **4** (2011) 905–930

[13] Lézoray, O., Grady, L., eds.: Image Processing and Analysis with Graphs: Theory and Practice. Digital Imaging and Computer Vision. CRC Press / Taylor and Francis (2012)

[14] Gilboa, G., Osher, S.: Nonlocal linear image regularization and supervised segmentation. Multiscale Model. Simul. **6** (2007) 595–630

[15] Bougleux, S., Elmoataz, A., Melkemi, M.: Discrete regularization on weighted graphs for image and mesh filtering. In: Proceedings of International Conference on Scale Space and Variational Methods in Computer Vision. SSVM '07, Ischia, Italy (2007) 128–139

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

[17] Coupé, P., Hellier, P., Prima, S., Kervrann, C., Barillot, C.: 3D wavelet subbands mixing for image denoising. Journal of Biomedical Imaging **2008** (2008) 1–11

[18] Buades, A., Coll, B., Morel, J.M.: A non-local algorithm for image denoising. In: Proceedings of IEEE Computer Vision and Pattern Recognition. Volume 2 of CVPR'05., San Diego, CA, USA (2005) 60–65

[19] Bresson, X.: A short note for nonlocal TV minimization (2009) `http://www.cs.cityu.edu.hk/~xbresson/ucla/papers/a%20short%20note%20for%20nonlocal%20TV%20minimization.pdf`.

[20] Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60** (1992) 259–268

[21] Combettes, P.L., Pesquet, J.C.: A proximal decomposition method for solving convex variational inverse problems. Inverse problems **24** (2008)

[22] Raj, A., Singh, G., Zabih, R., Kressler, B., Wang, Y., Schuff, N., Weiner, M.: Bayesian parallel imaging with edge-preserving priors. Magnetic Resonance in Medicine **57** (2007) 8–21

[23] Bagnato, L., Frossard, P., Vandergheynst, P.: Optical flow and depth from motion for omnidirectional images using a TV-ℓ_1 variational framework on graphs. In: Proceedings of the 16th IEEE International Conference on Image Processing. ICIP'09, Cairo, Egypt (2009) 1453–1456

[24] Pock, T., Schoenemann, T., Graber, G., Bischof, H., Cremers, D.: A convex formulation of continuous multi-label problems. In: Proceedings of the 10th European Conference on Computer Vision: Part III. ECCV '08, Berlin, Heidelberg, Springer-Verlag (2008) 792–805

[25] Carter, J.: Dual Methods for Total Variation-based Image Restoration. PhD thesis, UCLA, Los Angeles, CA, USA (2001)

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

[27] Aujol, J.F.: Some first-order algorithms for total variation based image restoration. J. Math. Imaging Vision **34** (2009) 307–327

[28] Chambolle, A., Darbon, J.: On total variation minimization and surface evolution using parametric maximum flows. International Journal of Computer Vision **84** (2009) 288–307

[29] Chan, T.F., Osher, S., Shen, J.: The digital TV filter and nonlinear denoising. IEEE Trans. Image Process. **10** (2001) 231–241

[30] Gilboa, G., Darbon, J., Osher, S., Chan, T.: Nonlocal convex functionals for image regularization (2006) UCLA CAM Report 06-57.

[31] Goldstein, T., Osher, S.: The split Bregman method for ℓ_1-regularized problems. SIAM J. Imaging Sci. **2** (2009) 323–343

[32] Zhang, X., Burger, M., Bresson, X., Osher, S.: Bregmanized nonlocal regularization for deconvolution and sparse reconstruction. SIAM J. Imaging Sci. **3** (2010) 253–276

[33] Setzer, S., Steidl, G., Teuber, T.: Deblurring Poissonian images by split Bregman techniques. J. Vis. Commun. Image Represent. **21** (2010) 193–199

[34] Fortin, M., Glowinski, R., eds.: Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems. North-Holland, Amsterdam (1983)

[35] Afonso, M., Bioucas Dias, J., Figueiredo, M.: An augmented Lagrangian approach to the constrained optimization formulation of imaging inverse problems. IEEE Trans. Image Process. **20** (2011) 681–695

[36] Glowinski, R., Tallec, P.: Augmented Lagrangian and operator-splitting methods in nonlinear

mechanics. Studies in Applied Mathematics, SIAM (1989)

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

[38] Chambolle, A., Lions, P.L.: Image recovery via total variation minimization and related problems. Numer. Math. **76** (1997) 167–188

[39] Jalalzai, K., Chambolle, A.: Enhancement of blurred and noisy images based on an original variant of the total variation. In: Proceedings of International Conference on Scale Space and Variational Methods in Computer Vision. SSVM '09, Voss, Norway (2009) 368–376

[40] Pustelnik, N., Chaux, C., Pesquet, J.C.: Parallel proximal algorithm for image restoration using hybrid regularization. IEEE Trans. Image Process. **20** (2011) 2450–2462

[41] Osher, S., Burger, M., Goldfarb, D., Xu, J., Yin, W.: An iterated regularization method for total variation-based image restoration. Model. Simul. **4** (2005) 460–489

[42] Darbon, J., Sigelle, M.: Image restoration with discrete constrained total variation part I: Fast and exact optimization. J. Math. Imaging Vision **26** (2006) 261–276 10.1007/s10851-006-8803-0.

[43] Ciril, I., Darbon, J.: Image denoising with a constrained discrete total variation scale space. In: Proceedings of Discrete Geometry for Computer Imagery, Nancy, France, Springer (2011) 465–476

[44] Combettes, P.L., Pesquet, J.C.: Proximal splitting methods in signal processing. In Bauschke, H.H., Burachik, R., Combettes, P.L., Elser, V., Luke, D.R., Wolkowicz, H., eds.: Fixed-Point Algorithms for Inverse Problems in Science and Engineering. Springer-Verlag, New York (2010) 185–212

[45] Jenatton, R., Mairal, J., Obozinski, G., Bach, F.: Proximal methods for hierarchical sparse coding. Journal of Machine Learning Research **12** (2011) 2297–2334

[46] Combettes, P.L., Pesquet, J.C.: Image restoration subject to a total variation constraint. IEEE Trans. Image Process. **13** (2004) 1213–1222

[47] Malgouyres, F.: Minimizing the total variation under a general convex constraint for image restoration. IEEE Trans. Image Process. **11** (2002) 1450–1456

[48] Meyer, Y.: Oscillating Patterns in Image Processing and Nonlinear Evolution Equations: The Fifteenth Dean Jacqueline B. Lewis Memorial Lectures. American Mathematical Society, Boston, MA, USA (2001)

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

[50] Arrow, K., Hurwicz, L., Uzawa, H.: Studies in linear and non-linear programming. Stanford University Press (1958)

[51] Bramble, J., Pasciak, J., Vassilev, A.: Analysis of the inexact uzawa algorithm for saddle point problems. SIAM J. Numer. Anal. **34** (1997) 3

[52] Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, New York (2011)

[53] Pesquet, J.C., Pustelnik, N.: A parallel inertial proximal optimization method (2012)

[54] Combettes, P.L., Dũng, D., Vũ, B.C.: Dualization of signal recovery problems. arXiv:0907.0436v4 (2009)

[55] Boyle, J.P., Dykstra, R.L.: A method for finding projections onto the intersection of convex sets in hilbert spaces. Advances in order restricted statistical inference (Iowa City, Iowa, 1985), Lecture Notes in Statist. **37** (1986) 2847

[56] Van den Berg, E., Friedlander, M.P.: Probing the Pareto frontier for basis pursuit solutions. SIAM J. Sci. Comput. **31** (2008) 890–912

[57] De Mazancourt, T., Gerlic, D.: The inverse of a block-circulant matrix. IEEE Trans. on Antennas and Propagation **31** (1983) 808–810

[58] Brélaz, D.: New methods to color the vertices of a graph. Communications of the ACM **22** (1979) 251–256

[59] Briceño Arias, L., Combettes, P.L.: A monotone+skew splitting model for composite monotone inclusions in duality. arXiv:1011.5517 (29 Nov 2010)

[60] Chen, G., Teboulle, M.: A proximal-based decomposition method for convex minimization problems. Math. Program. **64** (1994) 81–101

[61] Esser, E., Zhang, X., Chan, T.F.: A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. SIAM J. Imaging Sci. **3** (2010) 1015–1046

[62] Combettes, P.L., Pesquet, J.C.: Primal-dual splitting algorithm for solving inclusions with mixtures of composite, lipschitzian, and parallel-sum type monotone operators. Set-Valued and Variational Analysis **20** (2012) 307–330

[63] Rockafellar, R.T.: Convex Analysis (Princeton Mathematical Series). Princeton University

Press (1996)

[64] Combettes, P.L., Pesquet, J.C.: Proximal thresholding algorithm for minimization over orthonormal bases. SIAM J. Optim. **18** (2007) 1351–1376

[65] Grady, L.: Space-Variant Computer Vision: A Graph-Theoretic Approach. PhD thesis, Boston University, Boston, MA (2004)

[66] Grady, L., Schwartz, E.L.: The Graph Analysis Toolbox: Image processing on arbitrary graphs. Technical Report CAS/CNS-TR-03-021, Department of Cognitive and Neural Systems, Boston University, Boston, MA (2003)

[67] Taubin, G., Zhang, T., Golub, G.H.: Optimal surface smoothing as filter design. In: Proceedings of the 4th European Conference on Computer Vision - Volume I. ECCV '96, London, UK, Springer-Verlag (1996) 283–292