Total Cyclic Variation and Generalizations

Daniel Cremers · Evgeny Strekalovskiy

Abstract We introduce a general framework for regularization of signals with values in a cyclic structure, such as angles, phases or hue values. These include the total cyclic variation TV_{S^1}, as well as cyclic versions of quadratic regularization, Huber-TV and Mumford-Shah regularity. The key idea is to introduce a convex relaxation of the original non-convex optimization problem. The method handles the periodicity of values in a simple way, is invariant to cyclical shifts and has a number of other useful properties such as lower-semicontinuity. The framework allows general, possibly non-convex data terms. Experimental results are superior to those obtained without special care about wrapping interval end points. Moreover, we propose an equivalent formulation of the total cyclic variation which can be minimized with the same time and memory efficiency as the standard total variation. We show that discretized versions of these regularizers amount to NP-hard optimization problems. Nevertheless, the proposed framework provides optimal or near-optimal solutions in most practical applications.

D. Cremers · E. Strekalovskiy (✉)
Technical University Munich, Boltzmannstrasse 3,
85748 Garching, Germany
e-mail: evgeny.strekalovskiy@in.tum.de

D. Cremers
e-mail: cremers@tum.de

1 Introduction

1.1 Total Variation and Cyclic Structures

Total variation (TV) has been widely recognized as an effective and efficient means to regularize variational problems. The main advantage of TV is that it is able to preserve sharp discontinuities in the data while still providing a smoothing effect. Secondly, TV is convex and thus relatively easy to handle in minimization problems. Thirdly, due to the coarea formula the total variation applied to the characteristic function of a set coincides with a measure of its boundary making it well suited for various geometric optimization problems. TV has been applied to a great variety of image analysis problems, such as denoising [25], segmentation [15, 17], superresolution [16], inpainting [3], multi-view reconstruction [14] and optical flow [6, 26]. All these examples assume that the range of the function u, which is to be regularized, is a space of linear structure, e.g. $\mathbb{R}^n$ for some $n \geq 1$.

In many interesting image and signal processing applications, however, the range of signals is the unit circle S^1, or equivalently the space of planar orientations. Here, the function values representing e.g. angles, phases or hue values are periodic in nature. Important application areas where cyclical data arises naturally include interference microscopy, synthetic aperture radar (SAR) satellite imagery [8], or the analysis of imagery from time of flight cameras. In SAR image processing, one aims at reconstructing the object of interest, e.g. the ground elevation from the difference of *phases* between two phase image measurements. Similarly, time-of-flight cameras capture distance measurements of the scene, but only for distances up to the wavelength of the employed light. Distances above this limit are cyclically wrapped into the first basic interval. In all these applications, since the measurements are typically rather noisy, denoising of the cyclical signals is an integral part.

A straightforward attempt to realize a notion of total variation for functions with cyclical range is to simply represent the circle S^1 as $[0, 1)$ and then to apply the usual TV seminorm. However, this approach leads to artifacts due to neglecting the value wrapping at the boundary of $[0, 1)$. Jumps from 0 to 1, or to values near 1, will be penalized with a substantial contribution to the energy whereas there is actually no jump at all with respect to the S^1 structure. This generally leads to oversmoothing, important image structures may get lost when applied to segmentation, and artificial structures may appear. Moreover, TV is dependent on the specific representation of S^1 as $[0, 1)$, i.e. the penalization is different depending on which point of S^1 is taken to be 0. Specifically, cyclic shifts of input data will not result in the corresponding cyclic shift of the output.

1.2 Prior Work

The notion of total variation for functions $u : \Omega \to \mathcal{Y}$ defined on manifold domains Ω and with values in general manifolds $\mathcal{Y}$ has been studied by Giaquinta and Mucci [11, 12] using the theory of cartesian currents. They rigorously define the space $BV(\Omega; \mathcal{Y})$ and provide a number of theoretical insights, proving several properties of functions with bounded total variation such as lower-semicontinuity and structure theorems. This extends the previous work [9] by Giaquinta et al. which handles specifically the case $\mathcal{Y} = S^1$ and, among others, shows existence of minimizers of certain energies in the space of functions with bounded total cyclic variation. Although these papers provide a notion of total cyclic variation TV_{S^1}, they differ from our approach in several fundamental ways.

First, the above approach uses an *embedding* of the manifold S^1 into $\mathbb{R}^2$ to define $BV(\Omega; S^1)$ and TV_{S^1}. This means that one works entirely with functions $\widehat{u} : \Omega \to \mathbb{R}^2$ satisfying $|u(x)| = 1$ for a.e. $x \in \Omega$. The theory is formulated entirely in terms of cartesian currents, which are specific continuous linear maps on differential forms and represent generalized graphs of functions of bounded variation. It is not clear how to transfer the theoretical properties of TV_{S^1} to the corresponding statements in the setting of functions $u : \Omega \to S^1$, and even whether these properties still hold. In contrast, in the present paper we work directly with functions $u : \Omega \to S^1$, respectively $u : \Omega \to \mathbb{R}$. We present proofs of the theoretical properties for this setting and refer to the related results in [9, 11, 12] for cartesian currents where appropriate. Most of our results do not have equivalent formulations in [9, 11, 12], since they only make sense when working directly with $u : \Omega \to \mathbb{R}$, i.e. without the embedding. For instance, this is the case for our main theoretical contribution, the novel convex representation of TV_{S^1} in Theorem 1. We believe that our derivations may be more accessible to the computer vision community because they

do not rely on advanced mathematical concepts like k-forms and k-currents.

Second, and most importantly, for the approach of Giaquinta et al. it is not clear how to derive an algorithm to practically compute the minimizers of energies with cyclic regularizers. Neither of these works have performed any numerical validations or experiments. In contrast, our method leads to a convex formulation of the energy minimization problem and can be easily employed to compute the solutions numerically.

Finally, our approach also extends to more general regularizers such as Huber-TV, and even to non-convex regularizers such as the piecewise smooth Mumford-Shah model for regularizing signals with cyclic values.

To the best of our knowledge, the derivation, analysis and numerical implementation of regularizers for cyclic signals (with a direct definition on these signals without embedding the values into $\mathbb{R}^2$) has not been addressed systematically in the literature. Moreover, it turns out that this is by no means straightforward.

1.3 Contributions

In this paper we introduce and mathematically characterize regularizers for functions with values in a cyclical range. Specifically we make the following contributions:

- We introduce a novel kind of total variation, TV_{S^1}, for functions $u : \Omega \to S^1$ with cyclic or periodic values. The regularizer resembles closely the usual total variation and provides an elegant solution to handle the wrap-around of values. We show that TV_{S^1} can be defined for all measurable functions $u : \Omega \to \mathbb{R}$ by means of a dual representation.
- From the theoretical viewpoint we show that TV_{S^1} has a number interesting and desired mathematical properties such as invariance to cyclic shifts and isometries as well as lower-semicontinuity.
- Beyond the total cyclic variation TV_{S^1}, our framework also allows more general regularizers for S^1-valued functions. This includes regularizers such as Huber-TV and the quadratic penalizer as well as more advanced ones like truncated linear and the full Mumford-Shah regularizer. We show how to incorporate the S^1 structure into these regularizers, so that changes of the representation of the data will have no effect on the regularization.
- We prove the existence of minimizers for general TV_{S^1} regularized functionals. This result requires an especially tailored argument since functions u with $TV_{S^1}(u) < \infty$ are not necessarily elements of $BV(\Omega)$. We show that TV_{S^1} optimization is in general NP-hard and also propose an adaptive binarization method to obtain solutions closer to the global optimum.

– Furthermore, we give an equivalent reformulation of the constraint sets which enables one to minimize TV_{S^1} regularized energies with the same time and memory efficiency as the standard total variation. Finally, several experiments illustrate the applications of the S^1 framework. Comparing to usual regularizers for functions with linear values like TV, the proposed framework gives rise to visually superior and more natural results.

A preliminary version of this work was presented at a conference [28]. In comparison to [28] the present work has a number of novelties:

– We eliminate the range constraint $u : \Omega \to [0, 1)$ in the definition of TV_{S^1} and work entirely with functions $u : \Omega \to \mathbb{R}$. This gives rise to more intuitive formulations and reveals novel properties of TV_{S^1}. For instance, the cyclic shift property is reduced to a more fundamental representation invariance property.
– We extend the S^1 framework to more general regularizers, such as Huber-TV and Mumford-Shah.
– The mathematical properties of TV_{S^1} are extended, for instance we prove the existence of minimizers for TV_{S^1} regularized problems.
– We provide new experiments—phase denoising and cyclical smoothing—further demonstrating the advantages of cyclic regularizers.

The paper is organized as follows. First, in Sect. 2 we introduce the notion of the total cyclic variation TV_{S^1} for functions exhibiting certain regularity, for instance $u \in SBV$. Motivated by the non-convexity of TV_{S^1}, we prove in Sect. 3 a convex formulation in terms of the graph function 1_u of u based on the functional lifting theory. In Sect. 4 we show a number of mathematical properties of TV_{S^1}. Subsequently, Sect. 5 applies the framework to general optimization problems where an energy involving a data term and TV_{S^1} as the regularizer is to be minimized. Section 6 introduces more general regularizers for S^1-valued functions. Finally, we present experimental results in Sect. 7 and give a conclusion in Sect. 8.
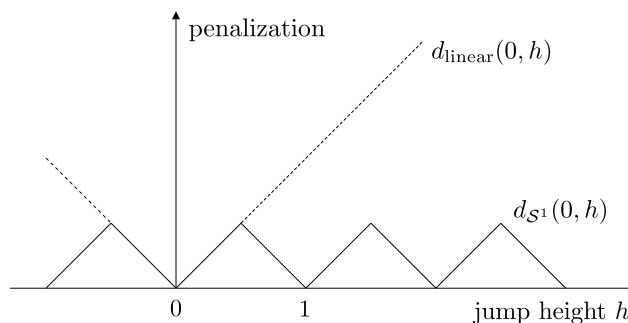
2 Definition of TV_{S^1}

In this section we give a definition of the total cyclic variation TV_{S^1}. This first definition can be seen as an intuitive and natural way of introducing TV_{S^1}, and can be formulated only for functions exhibiting certain regularity. We will extend this definition in later sections.

2.1 Preliminaries

Let $\Omega \subset \mathbb{R}^m$ be a bounded open set, $m \geq 1$. Let

$$S^1 = \left\{ x \in \mathbb{R}^2 \,\middle|\, |x| = 1 \right\} \tag{1}$$

Fig. 1 In contrast to the linear distance (*dashed*), the cyclic distance d_{S^1} (*solid*) accounts for the periodicity of S^1

be the unit circle. Since there is a one-to-one correspondence between S^1 and the unit interval $\Gamma := [0, 1)$, any function $u : \Omega \to S^1$ can be conveniently represented as a function $u : \Omega \to \Gamma$. This representation proves to be both convenient and useful for theoretical treatment of TV_{S^1}. More generally, we give a definition of TV_{S^1} for functions $u : \Omega \to \mathbb{R}$, without restricting the range of u to Γ. This means that at each point $x \in \Omega$ one is free to choose different representatives in $\mathbb{R}$ modulo 1 of the value $u(x)$.

As one easily checks, the distance function on S^1 transfers to the distance function d_{S^1} on $\mathbb{R}$ given by $d_{S^1}(a, b) = \min(|a - b|, 1 - |a - b|)$ for $a, b \in \Gamma$. For general $a, b \in \mathbb{R}$, we can easily show the representation in the lemma below. For fixed b this gives a "saw-tooth" shaped curve in a (Fig. 1).

Lemma 1 *Let $x, y \in \mathbb{R}$. Then*

$$d_{S^1}(x, y) = \min_{k \in \mathbb{Z}} |y - x - k|. \tag{2}$$

Proof Observe that $d_{S^1}(x, y) = d_{S^1}(0, y - x)$ and the right hand side of (2) are 1-periodic in $y - x$. Since both coincide with $\min(y - x, 1 - (y - x))$ for $y - x \in \Gamma$ the equation follows. $\qquad\square$

For functions $u \in L^1(\Omega)$ the *total variation* of u is defined by

$$TV(u) = \int_\Omega |Du| := \sup_{|\varphi| \leq 1} \int_\Omega u \operatorname{div} \varphi \, dx \tag{3}$$

where the supremum is taken over all $\varphi \in C_c^\infty(\Omega; \mathbb{R}^m)$ with $|\varphi(x)| \leq 1$ for all x. The space of *functions of bounded variation*, i.e. where $TV(u) < \infty$, is denoted by $BV(\Omega)$.

If $u \in BV(\Omega)$, we denote by S_u its "essential" jump set. This is where one cannot unambiguously define $u(x)$, but rather two values $u^+(x) > u^-(x)$ and a normal vector $\nu_u(x)$, indicating the direction of the jump and pointing from the u^- to the u^+ side.
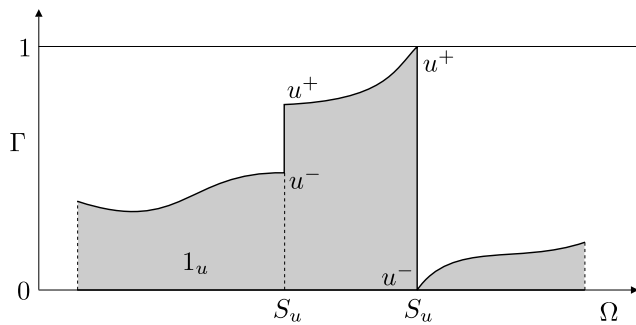
Fig. 2 Graph of a function $u \in SBV(\Omega)$. The jump set S_u consists here of two points, the second jump being just a wrap-around and not an actual jump in S^1 sense. The function 1_u is defined as 1 in the shaded area under the graph and 0 otherwise

One says that $u \in SBV(\Omega)$, the *special functions of bounded variation*, if the distributional gradient can be decomposed as $Du = \nabla u \, dx + (u^+ - u^-)\nu_u d\mathcal{H}^{m-1}|S_u$ into a smooth part and a jump part. The former is the absolutely continuous part of Du, the corresponding density function $\nabla u : \Omega \to \mathbb{R}^m$ is called the approximate gradient of u [2, Definition 3.70]. The latter is the jump height $u^+ - u^-$ multiplied with the Hausdorff $(m-1)$-dimensional measure restricted to the jump set S_u (Fig. 2). A general function $u \in BV(\Omega)$ may in addition have a "Cantor" part, an uncountable accumulation of infinitesimal jumps. We refer to [2] for a comprehensive introduction to functions of bounded variation.

2.2 Definition of TV_{S^1} for SBV Functions

We first define TV_{S^1} for functions $u \in SBV$, as this space allows to handle the smooth part and the jump part of the gradient in an explicit way. A more general definition for all measurable functions $u : \Omega \to \mathbb{R}$ will be given in Sect. 3.3 based on a dual formulation.

Definition 1 (TV_{S^1} for SBV Functions) For $u \in SBV(\Omega)$ we define the *total cyclic variation* as

$$TV_{S^1}(u) = \int_{\Omega \setminus S_u} |\nabla u| \, dx + \int_{S_u} d_{S^1}(u^-, u^+) d\mathcal{H}^{m-1}. \quad (4)$$

A straightforward generalization is the *weighted* TV_{S^1} where the contribution to TV_{S^1} at each point $x \in \Omega$ is weighted by a function $g : \Omega \to \mathbb{R}_{>0}$:

$$TV_{S^1}^g(u) = \int_{\Omega \setminus S_u} g \, |\nabla u| \, dx$$
$$+ \int_{S_u} g \, d_{S^1}(u^-, u^+) d\mathcal{H}^{m-1}. \quad (5)$$

This can be used to enable spatially adaptive regularizations. Writing $|u^- - u^+|$ instead of $d_{S^1}(u^-, u^+)$ in (4) would give

just the usual TV seminorm. Thus, in the region where u is smooth TV_{S^1} behaves just like TV, whereas at jumps it explicitly handles the S^1 structure by taking the cyclic distance d_{S^1} instead of the linear one.

Note that TV_{S^1} is not convex on the vector space $SBV(\Omega)$ because of the nonconvex distance d_{S^1}:

Proposition 1 TV_{S^1} *is not convex on* $SBV(\Omega)$.

Proof For any $A \subset \Omega$ with a rectifiable nonzero perimeter, $\text{Per}_\Omega(A) := TV(\chi_A) > 0$, define $u_1, u_2 \in SBV(\Omega)$ by $u_1 := 0$, $u_2 := \chi_A$. Then $TV_{S^1}(u_1) = TV_{S^1}(u_2) = 0$ but $TV_{S^1}(\frac{u_1+u_2}{2}) = \frac{1}{2}\text{Per}_\Omega(A) > 0$. □

In the next section, we will introduce a convex representation for efficiently minimizing functionals involving TV_{S^1}. As a consequence, this will also allow us to extend the definition of TV_{S^1} to *all* measurable functions $u : \Omega \to \mathbb{R}$.

3 Convex Representation of TV_{S^1}

The convexification of TV_{S^1} is based on the general framework of *functional lifting*.

3.1 Functional Lifting

In a series of papers [1, 20–23] it was shown that for functionals of the form

$$E(u) = \int_{\Omega \setminus S_u} h\big(x, u(x), \nabla u(x)\big) dx$$
$$+ \int_{S_u} d\big(x, u^-(x), u^+(x)\big) d\mathcal{H}^{m-1}(x) \quad (6)$$

a tight convex relaxation can be found using the method of cartesian currents [10] and that this relaxation can be employed to minimize these functionals. The idea is to restate $E(u)$ in terms of the graph function 1_u of u, see Figs. 2 and 3:

$$1_u : \Omega \times \mathbb{R} \to \mathbb{R}, \quad 1_u(x, t) = \begin{cases} 1 & \text{if } t < u(x), \\ 0 & \text{else.} \end{cases} \quad (7)$$

Although the functional $E(u)$ is in general highly nonconvex, it turns out that in case where $h : \Omega \times \mathbb{R} \times \mathbb{R}^m \to \mathbb{R}$ is convex in its third argument (i.e. in the gradient ∇u) it can be recast as a convex function of 1_u [1, Lemma 3.7]:

Lemma 2 *For $v \in BV(\Omega \times \mathbb{R})$ define*

$$\mathcal{E}(v) = \sup_\varphi \int_{\Omega \times \mathbb{R}} \varphi D v \quad (8)$$
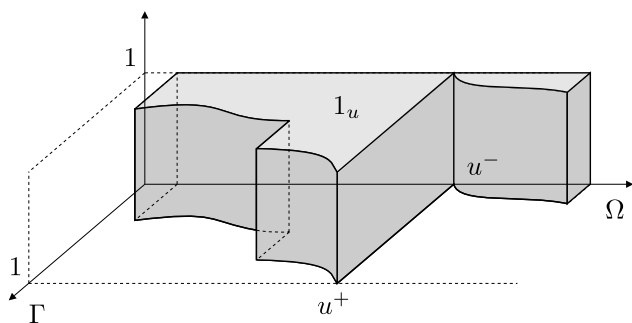
Fig. 3 The graph function $1_u : \Omega \times \Gamma \rightarrow \{0, 1\}$ of a function $u \in SBV(\Omega)$. The function 1_u has the value 1 under the graph of u, and 0 otherwise

where the supremum is taken over all vector fields $\varphi = (\varphi^x, \varphi^t) : \Omega \times \mathbb{R} \mapsto \mathbb{R}^m \times \mathbb{R}$ satisfying

$$\varphi^t(x, t) \geq h^*(x, t, \varphi^x(x, t)) \quad and$$
$$\left| \int_t^{t'} \varphi^x(x, s) \, ds \right| \leq d(x, t, t') \tag{9}$$

for all $x \in \Omega$ and $t, t' \in \mathbb{R}$. Then, for all $u \in SBV(\Omega)$,

$$\mathcal{E}(1_u) \leq E(u). \tag{10}$$

Moreover, equality holds in (10) for a given u if and only if there is a vector field $\varphi \in K$ with certain properties, see [1, Lemma 3.7]. Here, $h^*(x, t, q) := \sup_p p \cdot q - h(x, t, p)$ is the Legendre-Fenchel conjugate of $h(x, t, p)$ with respect to p. For a detailed introduction to convex analysis we refer to [24]. Dv denotes the distributional gradient of v.

The main advantage of formulation (10) is that $\mathcal{E}$ in (8) is convex in $v = 1_u$. This can be used to efficiently find minimizers of E as was done e.g. in [21], see also Sect. 5. In our case $E = TV_{S^1}$, we have $h(x, t, p) = |p|$ and $d(x, t, t') = d_{S^1}(t, t')$, comparing (4) with (6). Application of the above Lemma 2 leads to the following convex representation of TV_{S^1}.

Proposition 2 For $u \in SBV(\Omega)$,

$$TV_{S^1}(u) = \sup_{\varphi^x \in K} \int_{\Omega \times \mathbb{R}} \varphi^x \cdot \nabla_x 1_u \tag{11}$$

with the constraint set

$$K = \left\{ \varphi^x : \Omega \times \mathbb{R} \mapsto \mathbb{R}^m \,\middle|\, |\varphi^x(x, t)| \leq 1, \right.$$
$$\left| \int_t^{t'} \varphi^x(x, s) \, ds \right| \leq d_{S^1}(t, t')$$
$$\left. \forall x \in \Omega, \ t, t' \in \mathbb{R} \right\}. \tag{12}$$

We write $\nabla_x 1_u$ for the distributional gradient $D_x 1_u$ with respect to the spatial variable x, and later also $\operatorname{div}_x \varphi^x$ for the spatial divergence of φ^x and $\operatorname{div} \varphi = \operatorname{div}_x \varphi^x + \partial_t \varphi^t$ for the spatio-temporal divergence of $\varphi = (\varphi^x, \varphi^t)$. Since equality (11) is fundamental for our discussion, for the convenience of the reader we present here a direct proof without referring to Lemma 2. This also has the advantage that one sees how the constraints (12) on φ naturally arise. Lemma 2 will be used later when we deal with more general regularizers for S^1-valued functions.

Proof of Proposition 2 Let $u \in SBV(\Omega)$. For a general vector field $\varphi : \Omega \times \mathbb{R} \rightarrow \mathbb{R}^m \times \mathbb{R}$, $\varphi = (\varphi^x, \varphi^t)$ with $\varphi^t \equiv 0$, Lemma 2.8 of [1] states that

$$\int_{\Omega \times \mathbb{R}} \varphi^x \nabla_x 1_u = \int_{\Omega \setminus S_u} \varphi^x(x, u) \nabla u \, dx$$
$$+ \int_{S_u} \left(\int_{u^-}^{u^+} \varphi^x(x, s) \, ds \right) \cdot \nu_u d\mathcal{H}^{m-1}(x)$$

where u, $u^\pm$, ∇u and ν_u are all evaluated at x. Here, ν_u is the normal vector indicating the jump direction of u as described in Sect. 2.1. Comparing the corresponding parts, this expression is less than or equal to

$$TV_{S^1}(u) = \int_{\Omega \setminus S_u} |\nabla u| \, dx + \int_{S_u} d_{S^1}(u^-, u^+) d\mathcal{H}^{m-1}(x)$$

if

$$\varphi^x(x, u) \nabla u \leq |\nabla u|,$$
$$\left(\int_{u^-}^{u^+} \varphi^x(x, s) \, ds \right) \cdot \nu_u \leq d_{S^1}(u^-, u^+) \tag{13}$$

for a.e. $x \in \Omega \setminus S_u$, respectively for $\mathcal{H}^{m-1}$-a.e. $x \in S_u$. Demanding these constraints for all possible vectors $p \in \mathbb{R}^m$ instead of ∇u and all t, t' instead of u and $u^\pm$, we arrive at the constraints

$$|\varphi^x(x, t)| \leq 1, \qquad \left| \int_t^{t'} \varphi^x(x, s) \, ds \right| \leq d_{S^1}(t, t') \tag{14}$$

for all $x \in \Omega$, $t, t' \in \mathbb{R}$, which are exactly the ones in (12). Inequalities (13) are certainly fulfilled given (14). This proves (11) at least with $\geq$ instead of equality.

To show equality, we must construct a vector field φ^x satisfying (14) such that equalities hold in (13). We first define $\varphi^x(x, t)$ for $t \in \Gamma = [0, 1)$ only, and then extend this definition to all $t \in \mathbb{R}$ by 1-periodicity. For $x \in \Omega \setminus S_u$ set

$$\varphi^x(x, t) := \frac{\nabla u}{|\nabla u|} \left(1 - 2((t - u) \bmod 1) \right)$$

if $\nabla u \neq 0$, and $\varphi^x(x, t) := 0$ otherwise. Note that the approximate gradient $\nabla u : \Omega \rightarrow \mathbb{R}^m$ is by definition always a

function and not only a measure (Sect. 2.1). For $x \in S_u$, let $h := (u^+ - u^-) \bmod 1$, so that $h \in \Gamma$, and set

$$\varphi^x(x, t) := \nu_u(x) d_{S^1}(0, h)$$
$$\times \left(\frac{\chi_{[0,h]}(t - u^-)}{h} - \frac{\chi_{[0,1]\setminus[0,h]}(t - u^-)}{1 - h} \right)$$

if $h > 0$, and $\varphi^x(x, t) := 0$ otherwise. Let us first verify the equalities in (13). The first one for $x \in \Omega \setminus S_u$ is immediate. For the second one for $x \in S_u$, observe that $\int_{u^-}^{u^-+1} \varphi^x(x, s) \, ds = 0$ and thus $\int_{u^-+k}^{u^-+k+1} \varphi^x(x, s) \, ds = 0$ for any $k \in \mathbb{Z}$ due to 1-periodicity of φ^x. Therefore, the integral in (13) reduces to

$$\int_{u^-}^{u^+} \varphi^x(x, s) \, ds = \int_{u^-}^{u^-+h} \varphi^x(x, s) \, ds$$

with h as defined above. This can be easily evaluated to $d_{S^1}(0, h) = d_{S^1}(u^-, u^+)$ times $\nu_u(x)$, giving the second equality in (13).

Next, we must verify the constraints (14). The first one is obviously fulfilled for $x \in \Omega \setminus S_u$, and also for $x \in S_u$ due to $d_{S^1}(0, h) \leq h$ and $d_{S^1}(0, h) \leq 1 - h$. The second constraint can be verified directly for all t, t', but is most easily seen to hold using the equivalent formulation in Proposition 3 below: The integral of φ^x over the interval $[u, u+1]$ for $x \in \Omega \setminus S_u$, respectively over $[u^-, u^- + 1]$ for $x \in S_u$ is zero. Since φ is 1-periodic by construction, also the integrals over Γ are zero by Lemma 3 below. Thus, φ^x satisfies the conditions of Proposition 3, so that also the second constraint of (14) is fulfilled. $\qquad\square$

3.2 From Quadratic to Linear Complexity

The integral constraint in the definition (12) of K, using this formulation, must be satisfied for all pairs $t < t'$. In practice, the space of cyclical values Γ needs to be discretized at n levels $\frac{0}{n}, \ldots, \frac{n-1}{n}$ for some $n \geq 1$. In the discretization, this means that the number of constraints in (12) grows quadratically with the number of levels. Enforcing them quickly becomes unfeasible with respect to both runtime and memory consumption even for a moderate number of levels. Surprisingly, for the special metric $d = d_{S^1}$ there is an equivalent formulation of this constraint which leads to a much more efficient implementation.

One of the main contributions of this paper is to show that one can *efficiently* minimize functionals involving the TV_{S^1} regularizer. This is based on the following proposition which states that one can replace the integral constraint by a much simpler equivalent formulation:

Proposition 3 (Constraint Equivalence)

$$\left| \int_t^{t'} \varphi^x(x, s) \, ds \right| \leq d_{S^1}(t, t') \quad \text{for all } t, t' \in \mathbb{R} \qquad (15)$$

$\Longleftrightarrow$

$$\varphi^x(x, \cdot) \text{ 1-periodic}, \qquad (16)$$

$$|\varphi^x(x, t)| \leq 1 \quad \text{for a.e. } t \in \Gamma, \qquad (17)$$

$$\int_\Gamma \varphi^x(x, s) \, ds = 0. \qquad (18)$$

Proof Direction "$\Rightarrow$": From (15) we have

$$\left| \int_t^{t'} \varphi^x(x, s) \, ds \right| \leq d_{S^1}(t, t') \leq t' - t.$$

Dividing both sides by $t' - t$ for $t < t'$ and letting $t, t' \to t_0$ for some $t_0 \in \Gamma$ we obtain (17). Furthermore, setting $t = t_0$ and $t' = t_0 + 1$ for a fixed t_0 yields

$$\left| \int_{t_0}^{t_0+1} \varphi^x(x, s) \, ds \right| \leq d_{S^1}(t_0, t_0 + 1) = 0$$

and therefore $\int_{t_0}^{t_0+1} \varphi^x(x, s) \, ds = 0$. Specifically, $t_0 = 0$ yields (18). Defining $f(t) := \int_0^t \varphi^x(x, s) \, ds$, this means that $f(t + 1) = f(t)$ for all t, i.e. f is 1-periodic. Since $\varphi^x(x, t) = f'(t)$, (16) follows.

Direction "$\Leftarrow$": Let $t, t' \in \mathbb{R}$. Since $\varphi^x(x, \cdot)$ is 1-periodic, from (18) we easily get that also

$$\int_t^{t+k} \varphi^x(s) \, ds = 0$$

for all integers k, using Lemma 3 below. Also, by 1-periodicity (17) holds more general for all $t \in \mathbb{R}$. Therefore,

$$\left| \int_t^{t'} \varphi^x(s) \, ds \right| = \left| \int_t^{t'} \varphi^x(s) \, ds - \int_t^{t+k} \varphi^x(s) \, ds \right|$$
$$= \left| \int_{t+k}^{t'} \varphi^x(s) \, ds \right| \leq \int_{t+k}^{t'} |\varphi^x(s)| \, ds$$
$$\leq \int_{t+k}^{t'} 1 \, ds = |t' - t - k|$$

for k with $t + k \leq t'$ and in the same way also for k with $t + k \geq t'$. Combining these estimates for every k, we arrive at (15) using the representation of d_{S^1} in Lemma 1. $\qquad\square$

Effectively, this reduces the memory and time complexity of enforcing the constraint from quadratic to just linear. This way, from the implementation side the regularization of general data terms with TV_{S^1} becomes just as inexpensive as the usual TV regularization [22], where the number of constraints also grows linearly with the number of levels. In the proof, we have used the following simple fact about integrals of periodic functions:

Lemma 3 *Let* $p : \mathbb{R} \to \mathbb{R}$ *be* 1-*periodic. Then, for any* $a, b \in \mathbb{R}$,

$$\int_a^{a+1} p(s)\, ds = \int_b^{b+1} p(s)\, ds. \tag{19}$$

Proof By the periodicity of p we obtain for all $a \in \mathbb{R}$

$$\begin{aligned}
\int_a^{a+1} p(s)\, ds &= \int_a^1 p(s)\, ds + \int_1^{a+1} p(s)\, ds \\
&= \int_a^1 p(s)\, ds + \int_0^a p(s+1)\, ds \\
&= \int_a^1 p(s)\, ds + \int_0^a p(s)\, ds = \int_0^1 p(s)\, ds,
\end{aligned}$$

from which the statement follows. $\qquad\square$

3.3 Dual Representation of TV_{S^1} and Definition for All Measurable Functions

In the representation (11), the vector fields φ^x over which the supremum is taken are not required to be smooth or continuous. However, having only smooth φ can be useful for applications. In fact, one can show that one can restrict the supremum to vector fields φ^x which are smooth in x and t, and have compact support with respect to the x variable (compact support in the t variable is not required). Since the proof of this is quite technical, we only give the main lines here. Basically, if u is smooth outside the jump set S_u, i.e. $u \in C^\infty(\Omega \setminus S_u)$, then a smooth vector field φ^x with (14) and equalities in (13) can be constructed, so that the supremum is attained for this φ^x. For general $u \in SBV$, using partitions of unity one can construct smooth vector fields φ^x such that the integrals of $\varphi^x \nabla_x 1_u$ are arbitrarily close to $TV_{S^1}(u)$.

Starting from representation 2 of TV_{S^1} and considering only smooth φ^x we can employ integration by parts to obtain the following useful dual formulation of TV_{S^1} which does not require to know the jump set S_u explicitly:

Theorem 1 (Dual Representation) *For every* $u \in SBV(\Omega)$ *it holds*

$$TV_{S^1}(u) = \sup_{\varphi^x \in K} \int_\Omega \int_0^{u(x)} \operatorname{div}_x \varphi^x(x, s)\, ds\, dx \tag{20}$$

with

$$\begin{aligned}
K = \Big\{ \varphi^x \in C_c^\infty\big(\Omega; C^\infty(\mathbb{R}; \mathbb{R}^m)\big) \Big| \\
\varphi^x(x, \cdot)\ 1\text{-periodic},\ \big|\varphi^x(x, t)\big| \le 1, \\
\int_\Gamma \varphi^x(x, s)\, ds = 0\ \forall x \in \Omega,\ t \in \Gamma \Big\}. \tag{21}
\end{aligned}$$

Proof Follows directly from representation 2 by integration by parts and utilizing the efficient constraint reformulation in Proposition 3 (and substituting φ^x by $-\varphi^x$). $\qquad\square$

The lower bound 0 of the inner integral in (20) can be replaced by any other constant, leaving the overall integral value unchanged. For the weighted variant $TV_{S^1}^g$ the constraint $|\varphi^x(x, t)| \le 1$ must be replaced by $|\varphi^x(x, t)| \le g(x)$.

Note that the expression (20) is well defined for *all* measurable $u : \Omega \to \mathbb{R}$. Indeed, since $\varphi^x(x, \cdot)$ is 1-periodic with $\int_\Gamma \varphi^x(x, s)\, ds = 0$, for each $x \in \Omega$ we can replace the upper integration bound $u(x)$ in (20) by a value from $\Gamma = [0, 1)$. Hence, the overall integral is finite and well defined for each $\varphi \in K$.

Definition 2 (TV_{S^1} for Measurable Functions) We define $TV_{S^1}(u)$ for every measurable $u : \Omega \to \mathbb{R}$ by (20).

The representation in Theorem 1 is amazingly similar to the dual representation (3) of the usual total variation for scalar valued function. In fact, the integral in (20) reduces directly to (3) assuming that $\varphi(x, t)$ only depends on x. Therefore, our definition of TV_{S^1} can be seen as a natural notion of total variation for S^1-valued functions.

We remark that the dual formulation of TV_{S^1} requires the introduction of an additional dimension to the problem corresponding to the range space of u. On one hand this means additional memory requirements (after discretization). But on the other hand this allows one to easily couple and minimize TV_{S^1} with arbitrary, possibly nonconvex data terms, as will be done in Sect. 5.

4 Properties of TV_{S^1}

We will now investigate some mathematical properties of TV_{S^1}. First of all, TV_{S^1} has the desired property that, whenever u attains two different values $a, b \in \mathbb{R}$, the value change is penalized by the d_{S^1} distance of a and b multiplied with the length of the interface:

Proposition 4 (Consistency with d_{S^1}) *With fixed* $a, b \in \mathbb{R}$ *and* $A \subset \Omega$, *let* $u = a\chi_A + b\chi_{\bar{A}}$. *Then*

$$TV_{S^1}(u) = d_{S^1}(a, b)\operatorname{Per}_\Omega(A) \tag{22}$$

where $\operatorname{Per}_\Omega(A) = TV(\chi_A)$ *is the perimeter of A in Ω.*

Proof By Theorem 1,

$$\begin{aligned}
TV_{S^1}(u) = \sup_{\varphi \in K} \int_\Omega \chi_A(x) \int_0^a \operatorname{div}_x \varphi(x, s)\, ds\, dx \\
+ \int_\Omega \big(1 - \chi_A(x)\big) \int_0^b \operatorname{div}_x \varphi(x, s)\, ds\, dx
\end{aligned}$$

$$= \sup_{\varphi \in K} \int_{\Omega} \chi_A(x) \operatorname{div}_x \left(\int_b^a \varphi(x, s) \, ds \right) dx$$

$$+ \int_0^b \int_{\Omega} \operatorname{div}_x \varphi(x, s) \, dx \, ds. \tag{23}$$

The last integral is zero since φ is zero on the boundary. For each $\varphi \in K$, the function $p(x) := \int_b^a \varphi(x, s) \, ds$ in the first integral belongs to $C_c^{\infty}(\Omega; \mathbb{R}^m)$ and it holds $|p(x)| \leq d_{S^1}(a, b)$ using Proposition 3. Conversely, any $p \in C_c^{\infty}(\Omega; \mathbb{R}^m)$ bounded like this can be represented by a $\varphi \in K$, e.g. set $\varphi(x, t) := p(x) w_{a,b}(t)$ for any 1-periodic function $w \in C^{\infty}(\mathbb{R})$ satisfying $\int_b^a w(s) \, ds = d_{S^1}(a, b)$ and $|w_{a,b}(t)| \leq 1$. This means that in (23) we can instead take the supremum over all $p \in C_c^{\infty}(\Omega; \mathbb{R}^m)$ with $|p| \leq d_{S^1}(a, b)$. This results in the definition of just the usual total variation and we obtain

$$TV_{S^1}(u) = \sup_{|p| \leq d_{S^1}(a,b)} \int_{\Omega} \chi_A(x) \operatorname{div}_x p(x) \, dx$$

$$= d_{S^1}(a, b) TV(\chi_A) = d_{S^1}(a, b) \operatorname{Per}_{\Omega}(A). \qquad \square$$

Thus, TV_{S^1} as defined above is indeed the total *cyclic* variation of u. As a corollary we obtain that $TV_{S^1}(u)$ is zero for any constant $u : \Omega \to \mathbb{R}$.

Since the values in S^1 are periodic in nature one expects that $TV_{S^1}(u)$ does not change if the values of u are cyclically shifted everywhere by the same amount. This is indeed the case as stated in the following proposition. This means that TV_{S^1} is independent of a concrete representation of S^1 as the unit interval $\Gamma = [0, 1)$, respectively as the line $\mathbb{R}$, i.e. it does not matter which point of S^1 is taken to be 0. This property is not shared by TV regularization (Fig. 10). For $\alpha \in \mathbb{R}$ we define the shift operator $T_{\alpha} : \Gamma \to \Gamma$ by $T_{\alpha}(t) = t + \alpha \mod 1$.

Proposition 5 (Cyclic Shift Invariance) *For every* $u : \Omega \to \Gamma$ *and* $\alpha \in \mathbb{R}$ *it holds*

$$TV_{S^1}(T_{\alpha} \circ u) = TV_{S^1}(u). \tag{24}$$

Note that the cyclic shift T_{α} is a usual shift followed by a change of representation modulo 1. Since TV_{S^1} is clearly invariant under usual shifts, i.e. $TV_{S^1}(u + \alpha) = TV_{S^1}(u)$ which follows immediately from the general definition (20), the statement follows from the more fundamental Proposition 6, stated below.

Proposition 6 (Representation Invariance) *For a* $u : \Omega \to \mathbb{R}$ *and an integer function* $k : \Omega \to \mathbb{Z}$ *define* $u_k(x) := u(x) + k(x)$. *Then*

$$TV_{S^1}(u_k) = TV_{S^1}(u). \tag{25}$$

Proof By Theorem 1,

$$TV_{S^1}(u_k) = \sup_{\varphi^x \in K} \int_{\Omega} \int_0^{u_k(x)} \operatorname{div}_x \varphi^x(x, s) \, ds \, dx.$$

Writing $a(s) := \operatorname{div}_x \varphi^x(x, s)$ for brevity, the inner integral is

$$\int_0^{u_k(x)} a(s) \, ds = \int_0^{u(x)+k(x)} a(s) \, ds$$

$$= \int_{k(x)}^{u(x)+k(x)} a(s) \, ds + \int_0^{k(x)} a(s) \, ds$$

$$= \int_0^{u(x)} a(s + k(x)) \, ds + \int_0^{k(x)} a(s) \, ds.$$

Since $\varphi^x(x, \cdot)$ is 1-periodic with $\int_0^1 \varphi^x(x, s) \, ds = 0$ and the operator div_x is linear, also $(\operatorname{div}_x \varphi^x)(x, \cdot)$ is 1-periodic with $\int_0^1 \operatorname{div}_x \varphi^x(x, s) \, ds = 0$. Therefore, the above equality gives

$$\int_{\Omega} \int_0^{u_k(x)} \operatorname{div}_x \varphi^x \, ds \, dx = \int_{\Omega} \int_0^{u(x)} \operatorname{div}_x \varphi^x \, ds \, dx.$$

Taking the supremum over $\varphi^x \in K$ on both sides yields the required identity. $\qquad \square$

This means that TV_{S^1} is independent of the $\mathbb{R}$-representation of the S^1-value $u(x)$ at each point $x \in \Omega$. We can safely choose any other representation $u(x) + k(x)$ at every $x \in \Omega$ without altering the value of $TV_{S^1}(u)$. On the other hand, the usual TV of course does depend on the representation.

A useful geometrical property of TV_{S^1} is that it is invariant with respect to rigid transformations of the domain Ω:

Proposition 7 (Rigid Transformation Invariance) *Let* $\Omega' \subset \mathbb{R}^m$ *and* $A : \Omega' \to \Omega$ *be an isometry. Then for any* $u : \Omega \to \mathbb{R}$,

$$TV_{S^1}(u \circ A) = TV_{S^1}(u). \tag{26}$$

Proof Follows from representation (20) by the integral transformation formula and the invariance of the set K in (21) with respect to linear isometries. $\qquad \square$

In Sect. 5 we will address the existence of minimizers for general TV_{S^1} regularized problems. The proof relies on the following important lower-semicontinuity property of TV_{S^1}.

Proposition 8 (Lower-Semicontinuity) TV_{S^1} *is lower-semicontinuous with respect to pointwise a.e. convergence in the* d_{S^1} *metric, i.e. it holds*

$$TV_{S^1}(u) \leq \liminf_{n \to \infty} TV_{S^1}(u_n) \tag{27}$$

for all sequences $u_n : \Omega \to \mathbb{R}$, $n \geq 1$, *and functions* $u : \Omega \to \mathbb{R}$ *with* $d_{S^1}(u_n(x), u(x)) \to 0$ *for a.e.* $x \in \Omega$.

Proof Let $u_n, u : \Omega \to \mathbb{R}$ with $u_n(x) \to u(x)$ pointwise a.e. in Ω in the d_{S^1} metric, i.e. $d_{S^1}(u_n(x), u(x)) \to 0$ for a.e. $x \in \Omega$. For every $\varphi \in K$ and $v : \Omega \to \mathbb{R}$ define

$$F_\varphi(v) := \int_\Omega \int_0^{v(x)} \operatorname{div}_x \varphi^x(x, s) \, ds \, dx. \tag{28}$$

By representation in Theorem 1 it holds

$$T V_{S^1}(v) = \sup_{\varphi \in K} F_\varphi(v) \quad \text{for all } v : \Omega \to \mathbb{R}, \tag{29}$$

and for instance

$$F_\varphi(v) \leq T V_{S^1}(v) \quad \text{for all } v : \Omega \to \mathbb{R} \text{ and } \varphi \in K. \tag{30}$$

For each fixed $\varphi \in K$ and $n \geq 1$ we have the estimate

$$\begin{aligned}
\left| F_\varphi(u_n) - F_\varphi(u) \right| &= \left| \int_\Omega \int_{u(x)}^{u_n(x)} \operatorname{div}_x \varphi^x(x, s) \, ds \, dx \right| \\
&= \left| \int_\Omega \int_{u(x)}^{u_n(x)-k(x)} \operatorname{div}_x \varphi^x(x, s) \, ds \, dx \right| \\
&\leq C_\varphi \int_\Omega \left| u_n(x) - u(x) - k(x) \right| dx
\end{aligned}$$

with $C_\varphi := \| \operatorname{div}_x \varphi^x \|_{L^\infty(\Omega \times \Gamma)}$ for every $k : \Omega \to \mathbb{Z}$. We have used $\int_\Gamma \operatorname{div}_x \varphi^x(x, s) \, ds = 0$ and the 1-periodicity of $\operatorname{div}_x \varphi(x, \cdot)$ together with Lemma 3 for the second equality. For each $x \in \Omega$ we can choose a $k(x) \in \mathbb{Z}$ such that $|u_n(x) - u(x) - k(x)| = d_{S^1}(u_n(x), u(x))$ due to Lemma 1. Choosing this specific function $k : \Omega \to \mathbb{Z}$ in the above inequality, we obtain for all $\varphi \in K$ and $n \geq 1$

$$\left| F_\varphi(u_n) - F_\varphi(u) \right| \leq C_\varphi \int_\Omega d_{S^1}\big(u_n(x), u(x)\big) \, dx. \tag{31}$$

By assumption, the integrand on the right hand side tends to 0 pointwise a.e. in Ω and is obviously bounded by 1, with $1 \in L^1(\Omega)$. Thus, Lebesgue's convergence theorem yields that the right hand side of (31) tends to zero for $n \to \infty$, so that

$$F_\varphi(u) = \lim_{n \to \infty} F_\varphi(u_n). \tag{32}$$

Therefore, by (30) we also have

$$F_\varphi(u) = \liminf_{n \to \infty} F_\varphi(u_n) \leq \liminf_{n \to \infty} T V_{S^1}(u_n). \tag{33}$$

Taking the supremum over all φ in (33), and using (29), we finally get

$$T V_{S^1}(u) = \sup_{\varphi \in K} F_\varphi(u) \leq \liminf_{n \to \infty} T V_{S^1}(u_n),$$

i.e. the lower-semicontinuity of $T V_{S^1}$. $\qquad \square$

Giaquinta et al. proved several weak lower-semicontinuity results in [11, Sect. 6] for the $\mathbb{R}^2$-embedding formulation of $T V_{S^1}$ using cartesian currents.

5 Regularization with $T V_{S^1}$

The introduction of the regularizer $T V_{S^1}$ is of course motivated by optimization problems, namely the ones where the S^1 structure must be taken into account.

5.1 Optimization Problem and Existence of Minimizers

We consider an arbitrary, possibly nonconvex pointwise data term $\int_\Omega \varrho(x, u(x)) \, dx$ and apply $T V_{S^1}$ regularization to obtain the overall optimization problem

$$\inf_{u:\Omega \to \mathbb{R}} E(u), \quad E(u) = \int_\Omega \varrho\big(x, u(x)\big) \, dx + \lambda T V_{S^1}(u). \tag{34}$$

The weight $\lambda > 0$ allows to choose the influence of the regularizer. Since we work with functions $u : \Omega \to S^1$, represented by $u : \Omega \to \mathbb{R}$, it is natural to assume that ϱ is representation invariant, i.e. that $\varrho(x, \cdot)$ is 1-periodic. To make sure that the problem (34) has a minimizer we also assume that $\varrho(x, \cdot)$ is lower-semicontinuous with respect to d_{S^1} convergence, i.e.

$$\varrho(x, t) \leq \liminf_{n \to \infty} \varrho(x, t_n) \quad \text{whenever } d_{S^1}(t_n, t) \to 0 \tag{35}$$

for each fixed $x \in \Omega$ and $t \in \mathbb{R}$. To ensure lower semicontinuity of the data term energy, we further assume a lower bound $\varrho(x, t) \geq g(x)$ for all $x \in \Omega$, $t \in \mathbb{R}$ with some $g \in L^1(\Omega)$. For example, this is automatically satisfied if the data term is nonnegative, as will be the case in all of our considered applications.

First of all, the next theorem ensures the existence of minimizers of (34). The proof of this result requires an argument especially tailored for the S^1 case, since functions u with $T V_{S^1}(u) < \infty$ need not necessarily lie in $BV(\Omega)$. As a simple example consider any function $u : \Omega \to \{0, 1\}$ on $\Omega = (0, 1)$ which jumps infinitely often between 0 and 1, so that $T V_{S^1}(u) = 0$ and $T V(u) = \infty$. Note that the theorem establishes existence of functions $u : \Omega \to \mathbb{R}$ instead of $u : \Omega \to [0, 1)$, in accordance with the extended Definition 2 of $T V_{S^1}$.

Theorem 2 (Solvability) *Problem* (34) *admits a minimizer* $u : \Omega \to \mathbb{R}$ *(with* $T V_{S^1}(u) < \infty$*).*

Proof Consider a minimizing sequence $(u_n)_n$, so that

$$\int_\Omega \varrho\big(x, u_n(x)\big) \, dx + \lambda T V_{S^1}(u_n) \overset{n \to \infty}{\longrightarrow} E^*. \tag{36}$$

with E^* being the optimal infimum value in (34). First, this gives the uniform bound $TV_{S^1}(u_n) \leq C$ for all n with some $C > 0$. Consider the embedding $e : \mathbb{R} \to \mathbb{R}^2$ of the circle S^1, respectively line $\mathbb{R}$ into $\mathbb{R}^2$ as in Lemma 4 below. By this lemma the functions $e \circ u_n : \Omega \to \mathbb{R}^2$ are elements of $BV(\Omega; \mathbb{R}^2)$ with $TV(e \circ u_n) \leq TV_{S^1}(u_n) \leq C$ for all n. Thus, the sequence $(e \circ u_n)_n$ is uniformly bounded in $BV(\Omega; \mathbb{R}^2)$ (the L^1 norm is trivially bounded since $|e \circ u_n(x)| = \frac{1}{2\pi}$ for all x and n). By the weak* compactness of $BV(\Omega; \mathbb{R}^2)$ there exists a subsequence, again denoted by $e \circ u_n$, weak* converging to some $v : \Omega \to \mathbb{R}^2$. For instance, $e \circ u_n \to v$ in $L^1(\Omega; \mathbb{R}^2)$.

Therefore, also $e \circ u_n \to v$ pointwise a.e. in Ω for a subsequence. Since $|e \circ u_n(x)| = \frac{1}{2\pi}$ for all x and n, this gives that also $|v(x)| = \frac{1}{2\pi}$ a.e. in Ω. This allows us to define a function $u : \Omega \to \mathbb{R}$ with $e \circ u(x) = v(x)$ a.e. in Ω. The convergence $(e \circ u_n)(x) \to v(x) = (e \circ u)(x)$ directly gives $u_n(x) \to u(x)$ a.e. in the d_{S^1} metric, i.e. $d_{S^1}(u_n(x), u(x)) \to 0$. By the lower-semicontinuity property 8 of TV_{S^1} it follows that

$$TV_{S^1}(u) \leq \liminf_{n \to \infty} TV_{S^1}(u_n).$$

Also, by the lower-semicontinuity and the lower bound of the data term ϱ we can apply Fatou's lemma to get

$$\int_\Omega \varrho(x, u(x)) \, dx \leq \liminf_{n \to \infty} \int_\Omega \varrho(x, u_n(x)) \, dx.$$

Combining these two estimates we get

$$E(u) = \int_\Omega \varrho(x, u(x)) \, dx + \lambda TV_{S^1}(u)$$

$$\leq \liminf_{n \to \infty} \int_\Omega \varrho(x, u_n(x)) \, dx + \lambda TV_{S^1}(u_n) = E^*,$$

i.e. $u : \Omega \to \mathbb{R}$ minimizes E and $TV_{S^1}(u) < \infty$. □

Existence of Dirichlet minimizers for the corresponding formulation of energy (34) in the setting of cartesian currents was proven in [9, Sect. 5, Theorem 1].

In the proof we have used the following embedding result.

Lemma 4 *Define an embedding map of the circle S^1, respectively the line $\mathbb{R}$, into $\mathbb{R}^2$ by*

$$e : \mathbb{R} \to \mathbb{R}^2, \quad e(t) := \frac{1}{2\pi} \begin{pmatrix} \cos 2\pi t \\ \sin 2\pi t \end{pmatrix}. \tag{37}$$

Then each $u : \Omega \to \mathbb{R}$ with $TV_{S^1}(u) < \infty$ defines a function $e \circ u : \Omega \to \mathbb{R}^2$ with $e \circ u \in BV(\Omega; \mathbb{R}^2)$ and the following inequality holds:

$$TV(e \circ u) \leq TV_{S^1}(u). \tag{38}$$

Proof Recall that the total variation for vectorial functions $v \in L^1(\Omega; \mathbb{R}^2)$ is defined similarly as in the scalar case (3) by

$$TV(v) := \sup_{|\varphi| \leq 1} \int_\Omega v_1(x) \operatorname{div}_x \varphi_1(x) + v_2(x) \operatorname{div}_x \varphi_2(x) \, dx$$

where the supremum is taken over all $\varphi \in C_c^\infty(\Omega; \mathbb{R}^m)^2$ with $\sqrt{\varphi_1(x)^2 + \varphi_2(x)^2} \leq 1$ for all x. Consider specifically $v := e \circ u$ and an arbitrary fixed $\varphi \in C_c^\infty(\Omega; \mathbb{R}^m)^2$. Define a $p : \Omega \times \mathbb{R} \to \mathbb{R}^m$ by

$$p(x, t) := -\varphi_1(x) \sin 2\pi t + \varphi_2(x) \cos 2\pi t$$

for all $x \in \Omega$ and $t \in \mathbb{R}$. Then one easily checks that $p \in K$ with the set K in (21). Furthermore, by simple integration

$$\int_\Omega \int_0^{u(x)} \operatorname{div}_x p(x, s) \, ds \, dx$$

$$= \int_\Omega \int_0^{u(x)} \left(-\varphi_1(x) \sin 2\pi s + \varphi_2(x) \cos 2\pi s \right) ds \, dx$$

$$= \int_\Omega \frac{\cos 2\pi u(x) - 1}{2\pi} \operatorname{div}_x \varphi_1(x)$$

$$+ \frac{\sin 2\pi u(x)}{2\pi} \operatorname{div}_x \varphi_2(x) \, dx$$

$$= \int_\Omega \left((e \circ u)_1(x) - \frac{1}{2\pi} \right) \operatorname{div}_x \varphi_1(x)$$

$$+ (e \circ u)_2(x) \operatorname{div}_x \varphi_2(x) \, dx.$$

The constant $\frac{1}{2\pi}$ in this expression can be eliminated, since φ_1 is zero on the boundary of Ω. We obtain

$$\int_\Omega (e \circ u)_1(x) \operatorname{div}_x \varphi_1(x) + (e \circ u)_2(x) \operatorname{div}_x \varphi_2(x) \, dx$$

$$= \int_\Omega \int_0^{u(x)} \operatorname{div}_x p(x, s) \, ds \, dx.$$

By the representation (20) of TV_{S^1}, the right hand side is bounded by $TV_{S^1}(u)$ independently of p and thus of φ. Taking the supremum over all φ we get the required inequality (38). □

5.2 Lifted Optimization Problem and Relaxation

The question is now how to efficiently solve the problem (34). For the implementation, it is convenient to work with specific *unique* representatives of the cyclic values $u(x)$ for each x. We choose $\Gamma = [0, 1)$ as the basic interval, i.e. consider $u : \Omega \to \Gamma$ as the solution candidates.

First, using representation (20) the problem (34) can be rewritten as the saddle-point problem

$$\inf_u \sup_{\varphi \in K} \int_{\Omega \times \Gamma} -1_u \operatorname{div} \varphi \, ds \, dx \tag{39}$$

with the set K given by

$$K = \left\{ \varphi = (\varphi^x, \varphi^t) \in C_c^\infty\big(\Omega; C^\infty(\Gamma; \mathbb{R}^m \times \mathbb{R})\big) \,\Big|\right.$$

$$\varphi^t(x, t) \geq -\varrho(x, t), \left|\varphi^x(x, t)\right| \leq \lambda,$$

$$\left. \int_\Gamma \varphi^x(x, s)\, ds = 0 \, \forall x \in \Omega, \, t \in \Gamma \right\}. \qquad (40)$$

The dual variables $\varphi = (\varphi^x, \varphi^t)$ in the set K now also have a "vertical" component φ^t in addition to φ^x to handle the data term ϱ. The constraint set (40) can be derived exactly as in Sect. 3, using the equivalent constraint reformulation of Proposition 3. Alternatively, one could derive the constraints from the general Lemma 2. As opposed to (21) the 1-periodicity is not required in (40) because we consider φ only on the basic set $\Gamma = [0, 1)$. Note that we have incorporated the weight λ of TV_{S^1} in (34) into the set K. For the weighted variant $TV_{S^1}^g$, the second constraint in K must be replaced by $|\varphi^x(x, t)| \leq \lambda g(x)$.

The idea to find a solution u of (39) is to optimize directly in terms of graph functions 1_u instead of u. Note that although the expression (39) is convex in 1_u, the set of solutions $\{1_u \mid u : \Omega \to \Gamma\}$ is not convex. The reason is that the graph functions $1_u : \Omega \times \Gamma \to \{0, 1\}$ have binary range. Therefore, to overcome this difficulty we set $v = 1_u$ and relax the problem (39) to

$$\inf_{v \in C} \mathcal{E}(v), \quad \mathcal{E}(v) = \sup_{\varphi \in K} \int_{\Omega \times \Gamma} -v \operatorname{div} \varphi \, ds \, dx \qquad (41)$$

with the convex set

$$C = \left\{ v : \Omega \times \mathbb{R} \to [0, 1] \,\big|\, v(x, \cdot) \text{ non-increasing,} \right.$$

$$v(x, t) = 1 \, \forall t \leq 0,$$

$$\left. v(x, t) = 0 \, \forall t \geq 1, \, \forall x \in \Omega \right\}. \qquad (42)$$

This set is the convex hull of the valid binary graph functions. In other words, we extend the set of possible solutions 1_u to also allow relaxed "graph functions" with possible intermediate values in $[0, 1]$.

Note that the condition in C on the monotonicity of $v(x, \cdot)$ is actually not necessary. In fact, since φ^t can be chosen arbitrarily large in (40), the supremum in (41) will be finite only if $\partial_t v \leq 0$, i.e. if $v(x, \cdot)$ is non-increasing.

5.3 Efficient Primal-Dual Minimization

We discretize the two-dimensional image domain Ω into a $\{0, \dots, W - 1\} \times \{0, \dots, H - 1\}$ pixel grid, with $W \geq 1$ pixels horizontally and $H \geq 1$ pixels vertically. Dimensions other than $m = 2$ can be handled in the same way. The discretized image domain is again denoted by Ω. We also

discretize the range space $\Gamma = [0, 1)$ of u into $n \geq 1$ levels $\frac{0}{n}, \dots, \frac{n-1}{n}$. This is needed since the relaxed problem (41) is defined on the space $\Omega \times \Gamma$. The discretized variables v and φ, and the data term ϱ are represented by their node values $v(x, \frac{i}{n}) = v_i(x) \in \mathbb{R}$, $\varphi^x(x, \frac{i}{n}) = n\varphi_i^x(x) \in \mathbb{R}^m$, $\varphi^t(x, \frac{i}{n}) = \varphi_i^t(x) \in \mathbb{R}$ and $\varrho(x, \frac{i}{n}) = \varrho_i(x) \in \mathbb{R}$ at the pixels $x \in \Omega$ for every $0 \leq i < n$. The factor n for φ^x is chosen to simplify the resulting discretized energy.

The discretized version of (41) becomes

$$\inf_{v \in C^d} \sup_{\varphi \in K^d} \sum_{x \in \Omega} \sum_{i=0}^{n-1} \varphi_i^x(x) \nabla v_i(x) + \varphi_i^t(x) \partial_t^+ v_i(x) \qquad (43)$$

with

$$C^d = \left\{ (v_i)_{0 \leq i < n} : \Omega \to [0, 1] \,\big|\, v_0(x) = 1 \, \forall x \in \Omega \right\} \qquad (44)$$

and

$$K^d = \left\{ (\varphi_i^x, \varphi_i^t)_{0 \leq i < n} : \Omega \to \mathbb{R}^m \times \mathbb{R} \,\Big|\right.$$

$$\varphi_i^t(x) \geq -\varrho_i(x), \, \left|\varphi_i^x(x)\right| \leq \frac{\lambda}{n},$$

$$\left. \sum_{j=0}^{n-1} \varphi_j^x(x) = 0 \, \forall 0 \leq i < n, x \in \Omega \right\}. \qquad (45)$$

The discretization C^d of the set C in (42) does not include the monotonicity constraint since it is already implied through the supremum over φ^t as described at end of the last section. The slightly altered constraint $|\varphi^x| \leq \frac{\lambda}{n}$ in (45) instead of $|\varphi^x| \leq \lambda$ in (40) arises as a consequence of discretizing the integral $\int_\Gamma$ and the derivative ∂_t.

We use forward differences with Neumann boundary conditions for the spatial gradient ∇. For example, $\partial_{x_1} f_{x_1, x_2} = f_{x_1+1, x_2} - f_{x_1, x_2}$ if $x_1 < W - 1$ and 0 if $x_1 = W - 1$ and define the divergence by $\operatorname{div} := -\nabla^T$, the negative adjoint operator, in order for a discrete form of partial integration to hold. For the t-derivative ∂_t^+ we use forward differences with zero boundary condition: $\partial_t^+ f_t = f_{t+1} - f_t$ if $t < n - 1$ and $-f_t$ if $t = n - 1$. This is done to implicitly have $v_n(x) = 0$. The negative adjoint t-derivative $\partial_t^- = -(\partial_t^+)^T$ is then given by backward differences: $\partial_t^- f_t = f_t - f_{t-1}$ if $t > 0$ and f_t if $t = 0$.

The optimization problem (43) is a classical saddle-point problem. We solve it by the general algorithm of [21] which is especially devised for this kind of problems. This is a fast primal-dual algorithm which consists essentially in a gradient descent in v and a gradient ascent in φ, with an orthogonal reprojection onto the sets C^d and K^d. Since K^d includes the non-local constraint $\sum_{j=0}^{n-1} \varphi_j^x = 0$ there is no simple formula for the projection. Therefore, instead of a

direct projection we implement this sum constraint by Lagrange multipliers. For this, we add the term

$$\inf_{q:\Omega \to \mathbb{R}^m} \sum_{x \in \Omega} -q(x) \cdot \sum_{j=0}^{n-1} \varphi_j^x(x) \tag{46}$$

to the energy (43). The overall optimization problem becomes

$$\inf_{v \in C^d, q} \sup_{\varphi \in K_0^d} \sum_{x \in \Omega} \sum_{i=0}^{n-1} \varphi_i^x(x) \nabla v_i(x) + \varphi_i^t(x) \partial_t^+ v_i(x)$$
$$- \sum_{x \in \Omega} q(x) \sum_{j=0}^{n-1} \varphi_j^x(x) \tag{47}$$

with no constraints on the Lagrange multiplier $q : \Omega \to \mathbb{R}^m$, the set C^d as in (44) and the set K_0^d, which is K^d in (45) without the integral constraint:

$$K_0^d = \left\{ \varphi \,\middle|\, \varphi_i^t(x) \geq -\varrho_i(x), \ \big|\varphi_i^x(x)\big| \leq \frac{\lambda}{n} \ \forall i, x \right\}. \tag{48}$$

The equations for minimization of (47) can now be obtained in a straightforward way from the general algorithm in [21] and are listed in Fig. 4. In the algorithm, we denote by π_M the euclidean projection onto M. Beside $\tau = \sigma = 1$ one is also free to choose any other time steps $\tau, \sigma > 0$ with $\tau \sigma \leq 1$. We use [19] to set the time steps automatically.

Note that due to Proposition 3 we can substantially reduce the amount of memory required to implement the TV_{S^1} regularization by an order of magnitude. For illustration, for a 256×256 image and $n = 64$ levels discretization for the range set Γ, the original set (12) having quadratically many constraints requires 1232 MB memory overall, while the set (21) with linearly many constraints requires only 97 MB. In fact, for (21) only one constraints needs to be dualized as an additional energy term (46), introducing new variables $q(x)$, $x \in \Omega$.

5.4 Optimality of Solutions

In practice, the computed relaxed graph function solution $v : \Omega \times \Gamma \to [0, 1]$ may actually not be binary at some points $x \in \Omega$, so that $v(x, \cdot)$ is some decreasing function different from a step function. For the optimization problem with the usual TV regularizer instead of TV_{S^1} in (34) it is known [22] that a simple thresholding of the relaxed solution v e.g. at level $\frac{1}{2}$,

$$u(x) := \sup_{t \in \Gamma: \, v(x,t) \geq \frac{1}{2}} t, \tag{49}$$

produces an optimal solution u of (34). In our TV_{S^1} case there is a striking difference: The discretized problem is NP-hard!

Initialize:

$v_i^0 = 0$ for $i \geq 1$, $v_0^0 = 1$, $\bar{v}^0 = v^0$, $q^0 = 0$, $\varphi^0 = 0$,

$\tau = \sigma = 1$.

Iterate:

Dual ascent in φ: for $0 \leq i < n$, pointwise in $x \in \Omega$:

$$\varphi_i^{x,k+1} = \pi_{|\cdot| \leq \frac{\lambda}{n}} \left(\varphi_i^{x,k} + \frac{\sigma}{3} \big(\nabla \bar{v}_i^k - \bar{q}^k \big) \right),$$

$$\varphi_i^{t,k+1} = \max \left(\varphi_i^{t,k} + \frac{\sigma}{2} \partial_t^+ \bar{v}_i^k, \ -\varrho_i \right).$$

Primal descent in v, q: for $1 \leq i < n$, pointwise in $x \in \Omega$:

$$v_i^{k+1} = \pi_{[0,1]} \left(v_i^k - \frac{\tau}{6} \big(\text{div}_x\, \varphi_i^{x,k+1} + \partial_t^- \varphi_i^{t,k+1} \big) \right),$$

$$q^{k+1} = q^k + \frac{\tau}{n} \sum_{j=0}^{n-1} \varphi_j^{x,k+1}.$$

Extrapolation step: for $1 \leq i < n$, pointwise in $x \in \Omega$:

$$\bar{v}_i^{k+1} = 2v_i^{k+1} - v_i^k,$$

$$\bar{q}^{k+1} = 2q^{k+1} - q^k.$$

Fig. 4 Primal-dual algorithm for the TV_{S^1} regularized optimization problem (34), with the discrete formulation (47)

Proposition 9 (Discrete TV_{S^1} is NP-Hard) *For a discrete range Γ with $n \geq 3$ levels, where $n = 3z$ for some $z \in \mathbb{Z}$, the optimization problem (34) is NP-hard.*

Proof The range Γ is discretized into n equally spaced angles $\alpha_i = \frac{i}{n}$, $i = 0, \ldots, n-1$. Since $n = 3z$ with some $z \in \mathbb{Z}$, the set of angles contains the subset $\mathcal{M} = \{0, \frac{1}{3}, \frac{2}{3}\}$. Choosing the data term $\varrho_i(x) = \infty$ for every i with $\alpha_i \notin \mathcal{M}$ ensures that any candidate solution with finite energy will be a function $u : \Omega \to \mathcal{M}$, i.e. only these three angles will be present. As the distance d_{S^1} is the same for any pair of these three angles, by Proposition 4, TV_{S^1} then is nothing but a Potts regularizer on three labels. The latter problem is known to be NP hard [5] as the multiway cut problem [13] can be reduced to it. $\square$

Of course, the discretization step is necessary in order to practically compute a solution. In light of the above result, in general we cannot expect exact solutions to (34). Nevertheless, we still can use the thresholding formula (49) to obtain an *approximate* solution. Furthermore, an energy bound holds which gives an a posteriori energy estimate of

the quality of the obtained solution:

$$\left|E(u) - E(u^*)\right| \le \left|\mathcal{E}(1_u) - \mathcal{E}(v)\right|. \tag{50}$$

Here u^* is the unknown optimal solution to (34), v the relaxed solution of (41) and u an approximate solution to (34) obtained by some binarization method from v, e.g. by (49). The bound follows from the simple observation that

$$\mathcal{E}(v) \le \mathcal{E}(1_{u^*}) = E(u^*) \le E(u) = \mathcal{E}(1_u), \tag{51}$$

where the inequalities hold because v is optimal for $\mathcal{E}$ and u^* is optimal for E. In all our experiments the deviation from the optimal solution was less than 5 % energy-wise. This is an overestimation of the true deviation obtained with regard to formula (50) by comparing $\mathcal{E}(1_u)$ to $\mathcal{E}(v)$.

5.5 Primal Energy Computation

The primal energy $\mathcal{E}(v)$ for each candidate solution $v \in C^d$ can be computed by evaluating the supremum in (43) or, equivalently, the supremum in (47) and the infimum over the Lagrange multiplier q:

$$\mathcal{E}(v) = \sum_{x \in \Omega} \sum_{i=0}^{n-1} -\varrho_i(x) \partial_t^+ v_i(x)$$

$$+ \frac{\lambda}{n} \sum_{x \in \Omega} \min_{q \in \mathbb{R}^m} \sum_{i=0}^{n-1} \left|\nabla v_i(x) - q\right|. \tag{52}$$
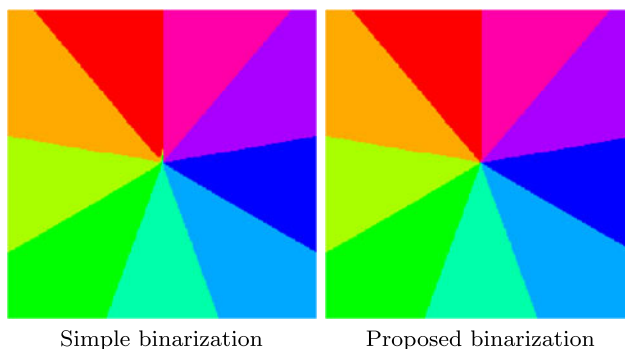
A value $q \in \mathbb{R}^m$ minimizing

$$\|z\| := \min_{q \in \mathbb{R}^m} \sum_{i=0}^{n-1} |z_i - q| \tag{53}$$

for a collection of vectors $z = (z_0, \ldots, z_{n-1}) \in (\mathbb{R}^m)^n$ is known as the "geometric median" of z. It can be computed e.g. using the iterative algorithm in [18].

5.6 Improved Binarization Method

The binary solution u obtained by the simple thresholding method (49) may produce undesirable artifacts in some cases as seen in the left image of Fig. 5. We propose an improved binarization method, which chooses the "best" binarized solution with respect to the actually used regularizer TV_{S^1}. For a fixed point $x \in \Omega$ let $v(x) = (v_0(x), \ldots, v_{n-1}(x))$ be the vector $v(x, \cdot)$ computed by the algorithm. We compare $v(x)$ with all possible n candidate *binary* graph functions $f^i = (f_0^i, \ldots, f_{n-1}^i), 0 \le i < n$, with $f_j^i := \chi_{j \le i}$. For all $x \in \Omega$ we assign

$$u(x) := \frac{i(x)}{n}, \quad \text{where } i(x) := \operatorname*{arg\,min}_{0 \le i < n} \|v(x) - f^i\| \tag{54}$$

Simple binarization   Proposed binarization

Fig. 5 Improved binarization of the relaxed solution, here for the inpainting problem from Sect. 7.3. While simple binarization by thresholding the relaxed graph function (*left*) may lead to artifacts (*center*), the proposed adaptive method (*right*) picks a binary value closest with respect to a distance derived from the regularizer TV_{S^1} itself

with the seminorm $\|\cdot\|$ in (53). Since here we only deal with *scalar* values $z_j := v_j(x) - f_j^i$, the "geometric median" q realizing the minimum in (53) is easily seen to be the usual median of the numbers $z_0, \ldots, z_{n-1}$. It can thus be computed by simply sorting these numbers. In practice, we observed that this binarization removes the artifacts of the simple method (49), see Fig. 5.

6 General Regularizers for S^1-Valued Functions

Recall that our formulation of the total cyclic variation TV_{S^1} is, in essence, based on the functional lifting framework of Sect. 3.1, choosing special penalizations for the smooth and the jump part of u. We have explicitly accounted for the S^1 structure by setting the jump penalization to zero if u jumps by 1 or some other integer value. Following this idea, we can also devise convex representations for more general functionals for S^1-valued functions.

Consider a general functional of the form (10). In order for it to be well defined for functions $u : \Omega \to S^1$ with cyclical range S^1, which are represented by functions $u : \Omega \to \mathbb{R}$, the functions h and d must naturally satisfy the following compatibility conditions:

- $h(x, \cdot, p)$ 1-periodic $\forall x \in \Omega, \quad p \in \mathbb{R}^m$, (55)

- $d(x, t, t') = 0$ if $t' - t \in \mathbb{Z} \ \forall x \in \Omega, \ t, t' \in \mathbb{R}$. (56)

This is because we want the values $u(x) + k$ to actually represent one and the same point on S^1 independently of $k \in \mathbb{Z}$. In other words, we demand $E(u)$ to be representation invariant in the sense of Proposition 6.

By functional lifting, a convex representation of E in terms of 1_u is given by (8) with the constraints (9) on φ. Just as in the proof of Proposition 3, we can show that (56) implies that $\varphi(x, \cdot)$ is necessarily 1-periodic and that $\int_\Gamma \varphi^x(x, s) \, ds = 0$.

The supremum over φ may not be attained for general E. Then, functional lifting yields a convex *relaxation* of E in terms of 1_u. We also assume that we have smooth vector fields, so that we can employ integration by parts. Summarizing, a convex relaxation of $E(u)$ for $u \in SBV(\Omega)$ is given by

$$E_{\text{conv}}(u) := \sum_{\varphi \in K} \int_\Omega -1_u \operatorname{div} \varphi \, ds \, dx$$

$$= \sup_{\varphi \in K} \int_\Omega \int_0^{u(x)} - \operatorname{div} \varphi(x, s) \, ds \, dx \qquad (57)$$

with the constraint set

$$K = \left\{ \varphi = (\varphi^x, \varphi^t) \in C_c^\infty\big(\Omega; C^\infty(\mathbb{R}; \mathbb{R}^m \times \mathbb{R})\big) \,\Big|\, \right.$$

$$\varphi(x, \cdot) \text{ 1-periodic,}$$

$$\varphi^t(x, t) \geq h^*\big(x, t, \varphi^x(x, t)\big),$$

$$\left| \int_t^{t'} \varphi^x(x, s) \, ds \right| \leq d(x, t, t'),$$

$$\left. \int_\Gamma \varphi^x(x, s) \, ds = 0 \,\, \forall x \in \Omega,\,\, t, t' \in \Gamma \right\}. \qquad (58)$$

Note that the inequality constraints are required only for $t, t' \in \Gamma = [0, 1)$. Assuming the compatibility conditions (55) and (56) we can prove a similar representation invariance property of the convex relaxation E_{conv} as in Proposition 6. This means that the convex relaxation E_{conv} of E can indeed be viewed as a functional defined for functions $u : \Omega \to S^1$, independently of their representation as $u : \Omega \to \mathbb{R}$.

Proposition 10 (Representation Invariance) *Let $u : \Omega \to \mathbb{R}$ and let $k : \Omega \to \mathbb{Z}$ be some integer function. Define $u_k(x) := u(x) + k(x)$. Then $E_{\text{conv}}(u_k) = E_{\text{conv}}(u)$.*

Proof By (57),

$$E_{\text{conv}}(u) = \sup_{\varphi \in K} \int_\Omega \int_0^{u_k(x)} \big(\operatorname{div}_x \varphi^x + \partial_t \varphi^t\big) \, ds \, dx$$

$$= \sup_{\varphi \in K} \int_\Omega \int_0^{u(x)+k(x)} \big(\operatorname{div}_x \varphi^x + \partial_t \varphi^t\big) \, ds \, dx.$$

Just as in the proof of Proposition 6, the integral over $\operatorname{div}_x \varphi^x$ can be written as

$$\int_\Omega \int_0^{u(x)+k(x)} \operatorname{div}_x \varphi^x \, ds \, dx = \int_\Omega \int_0^{u(x)} \operatorname{div}_x \varphi^x \, ds \, dx.$$

On the other hand, for the inner integral over $\partial_t \varphi^t$ we can write

$$\int_0^{u(x)+k(x)} \partial_t \varphi^t \, ds = \varphi^t\big(u(x) + k(x)\big) - \varphi^t(0)$$

$$= \varphi^t\big(u(x)\big) - \varphi^t(0) = \int_0^{u(x)} \partial_t \varphi^t \, ds$$

since $\varphi^t(x, \cdot)$ is 1-periodic. Together, this combines to

$$\int_\Omega \int_0^{u_k(x)} \big(\operatorname{div}_x \varphi^x + \partial_t \varphi^t\big) \, ds \, dx$$

$$= \int_\Omega \int_0^{u(x)} \big(\operatorname{div}_x \varphi^x + \partial_t \varphi^t\big) \, ds \, dx.$$

Taking the supremum over $\varphi \in K$ on both sides we obtain the required identity. $\square$

In the next sections we give some notable examples of regularizers R for S^1-valued functions, which have already been successfully applied in the case of functions with linear range. In each case, we give the constraints on the duals φ for the convex relaxation of the overall energy $\int_\Omega \varrho(x, u(x)) \, dx + R(u)$.

6.1 Huber S^1-Regularization

Regularization with the usual TV is known to produce piecewise constant solutions, an effect known as staircasing. Experiments show that this is also the case for TV_{S^1}. A simple remedy is to use quadratic penalization for small values of ∇u and linear penalization otherwise. For this, one penalizes the gradients ∇u (in the region $\Omega \setminus S_u$ where u is smooth) by the Huber function

$$h_\varepsilon(p) := \begin{cases} \frac{|p|^2}{2\varepsilon} & \text{if } |p| \leq \varepsilon, \\ |p| - \frac{\varepsilon}{2} & \text{else} \end{cases} \qquad (59)$$

for some small $\varepsilon > 0$, which smooths out the kink at the origin. The regularizer is then

$$R(u) = \lambda \int_{\Omega \setminus S_u} h_\varepsilon(\nabla u) \, dx + \lambda \int_{S_u} d_{S^1}\big(u^-, u^+\big) d\mathcal{H}^{m-1}. \qquad (60)$$

The corresponding convex set K for linear structures was computed in [22]. Adding the vanishing integral constraint, the constraints (58) become

$$\varphi^t(x, t) \geq -\varrho(x, t) + \frac{\varepsilon}{2\lambda}\big|\varphi^x(x, t)\big|^2,$$

$$\big|\varphi^x(x, t)\big| \leq \lambda, \qquad (61)$$

$$\int_\Gamma \varphi^x(x, s) \, ds = 0.$$

6.2 Quadratic S^1-Regularization

In the case of vector space valued functions, application of the quadratic regularizer produces a smoothing of the image, which is roughly equivalent to smoothing with a Gaussian kernel. Our framework enables one to define such a smoothing operation also for S^1-valued functions in a natural way. We consider the regularizer

$$R(u) = \alpha \int_\Omega |\nabla u|^2 \, dx + \int_{S_u} \delta_{d_{S^1}(u^-, u^+)=0} d\mathcal{H}^{m-1} \quad (62)$$

with parameter $\alpha > 0$ where the indicator function δ_t is defined as zero if $t = 0$ and as ∞ otherwise. It can be seen as the limiting case of the Huber regularization (60), defining λ depending on ε by $\lambda_\varepsilon := \alpha \cdot 2\varepsilon$ and letting $\varepsilon \to \infty$. Then $\lambda_\varepsilon h_\varepsilon(\nabla u) \to \alpha|\nabla u|^2$ in the smooth part, and in the jump part the jumps are penalized more and more, so that in the limit they become prohibited unless $d_{S^1}(u^-, u^+) = 0$, i.e. if u^- and u^+ actually represent one and the same value in S^1.

The relaxation is derived from (61) by means of this limiting process giving the constraints

$$\varphi^t(x, t) \geq -\varrho(x, t) + \frac{1}{4\alpha} |\varphi^x(x, t)|^2,$$
$$\int_\Gamma \varphi^x(x, s) \, ds = 0. \quad (63)$$

6.3 Truncated Linear S^1-Regularization

If two values u_1, u_2 are "too" distinct, it is often useful to penalize just the fact that the values are different, by a constant c independent of u_1, u_2. For small differences one can still use e.g. the d_{S^1} distance as in the case of TV_{S^1}. This leads to the truncated linear penalizer

$$R(u) = \lambda \int_{\Omega \setminus S_u} |\nabla u| \, dx$$
$$+ \int_{S_u} \min(\nu, \lambda d_{S^1}(u^-, u^+)) d\mathcal{H}^{m-1} \quad (64)$$

with parameters $\lambda, \nu > 0$. A relaxation is given by (41) with the constraints

$$\varphi^t(x, t) \geq -\varrho(x, t),$$
$$|\varphi^x(x, t)| \leq \lambda,$$
$$\left| \int_t^{t'} \varphi^x(x, s) \, ds \right| \leq \nu, \quad (65)$$
$$\int_\Gamma \varphi^x(x, s) \, ds = 0.$$

In order for the truncation in the jump part to have any effect, the truncation parameter ν must of course be chosen smaller than $\frac{1}{2}\lambda$, where $\frac{1}{2}$ is the maximal value achievable by d_{S^1}.

6.4 Mumford-Shah S^1-Regularization

We can even obtain the full Mumford-Shah regularizer for cyclic structures:

$$R(u) = \alpha \int_{\Omega \setminus S_u} |\nabla u|^2 \, dx + \nu \int_{S_u} \chi_{d_{S^1}(u^-, u^+) \neq 0} d\mathcal{H}^{m-1} \quad (66)$$

with parameters $\alpha, \nu > 0$. The last integral over the jump part penalizes the overall length of the interface S_u, but it considers only the interface points where $d_{S^1}(u^-, u^+) \neq 0$, i.e. if there is really a jump in the S^1 sense. Minimization of the Mumford-Shah functional produces piecewise smooth approximations of input signals $f : \Omega \to \mathbb{R}$, choosing for example $\varrho(x, u(x)) := d_{S^1}(u(x), f(x))^2$.

For the case of functions with linear range, in [21] Pock et al. introduced an algorithm for the minimization of this functional based on the method of functional lifting. A relaxation for functions $u : \Omega \to S^1$ is given by (41) with the constraints

$$\varphi^t(x, t) \geq -\varrho(x, t) + \frac{1}{4\alpha} |\varphi^x(x, t)|^2,$$
$$\left| \int_t^{t'} \varphi^x(x, s) \, ds \right| \leq \nu, \quad (67)$$
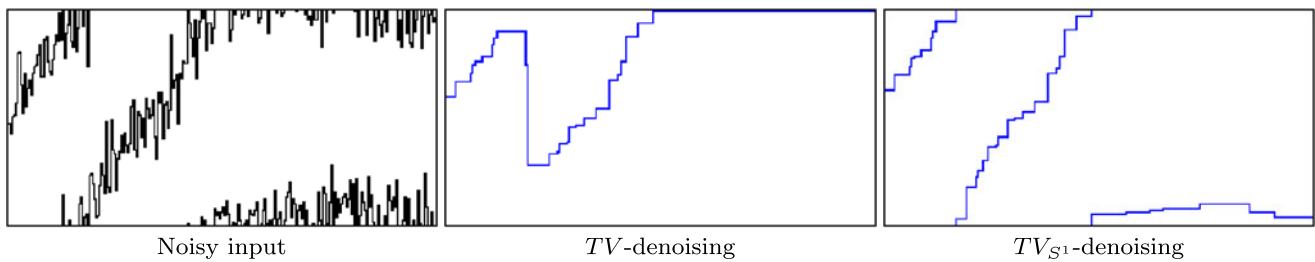$$\int_\Gamma \varphi^x(x, s) \, ds = 0.$$

6.5 Implementation

The quadratic and Huber-TV penalization can be implemented in exactly the same way as in the case of TV_{S^1}. The only difference is the projection of dual variables φ onto the corresponding constraint sets K in (58). For the quadratic penalizer one has to project onto a parabola, and for Huber-TV onto a sideways truncated parabola, see [22] for more details.

The situation is different for the truncated linear and the Mumford-Shah penalizer since there are quadratically many non-local constraints $|\int_t^{t'} \varphi^x(x, s) ds| \leq \nu$, respectively in the discrete setting

$$\left| \sum_{i < z \leq j} \varphi_z^x(x) \right| \leq \nu \quad \forall 0 \leq i < j < n. \quad (68)$$

Since each constraint involves many φ's at once, we use a different scheme than in Sect. 5.3. First, we introduce auxiliary variables $p_i : \Omega \to \mathbb{R}^m$ for $0 \leq i < n$ through $\partial_t^- p_i = \varphi_i^x$ and add the corresponding Lagrange multiplier terms

$$\inf_{\mu_i : \Omega \to \mathbb{R}^m} \sum_{x \in \Omega} \sum_{i=0}^{n-1} \mu_i (\partial_t^- p_i - \sigma_i^x) \quad (69)$$

| Noisy input | TV-denoising | TV_{S^1}-denoising |

Fig. 6 Denoising of angle values. TV_{S^1} (*right*) reconstructs the monotonous and the constant signal part correctly despite the strong noise (Gaussian 10 %). TV (*center*) falsely penalizes wrap-arounds driving the values towards the middle to reduce jump height. The reconstruction lacks monotonicity and is falsely equal to 1 in the constant part to avoid the jump from 1 to 0.05 in the middle

to the energy (43) to enforce these equalities. Second, by this the constraints on φ^x translate to the simpler ones $|p_j - p_i| \leq \nu$ for $0 \leq i < j < n$, $x \in \Omega$. We now use the observation that

$$\delta_{|d| \leq \nu} = \sup_{a \in \mathbb{R}^m} a\, d - \nu |a| \quad \forall d \in \mathbb{R}^m \tag{70}$$

to enforce them (the indicator function $\delta_{|d| \leq \nu}$ is 0 if the constraint $|d| \leq \nu$ is fulfilled and ∞ otherwise), adding the terms

$$\inf_{a_{ij} : \Omega \to \mathbb{R}^m} \sum_{x \in \Omega} \sum_{0 \leq i < j < n} -a_{ij}(p_j - p_i) + \nu |a_{ij}| \tag{71}$$

to the energy. Finally, in terms of p the S^1-constraint $\sum_{0 \leq j < n} \varphi_j(x) = 0$ translates into simply $p_{n-1}(x) = 0$.

We use the extension [7] of the primal-dual algorithm [21], since the proximal operators are not simply projections this time. As there are far more primal than dual variables due to quadratically many a_{ij}'s in (71), we suggest to use the version of the algorithm in [7] where the "bar"-copies are introduced for the duals rather than for the primals.

7 Experiments

In this section we present several experiments demonstrating the use of TV_{S^1} and other cyclic regularizers for various imaging problems. We evaluate the TV_{S^1} results by comparing them with TV regularization, since no regularizers for S^1-valued functions have yet been investigated and TV_{S^1} is supposed to behave like TV in regions where u is smooth.

With a parallel CUDA implementation on NVIDIA GTX 480 a typical runtime for 256×256 images using 64 levels for Γ is about 10 seconds for TV_{S^1}. TV [22] requires about 5 seconds. For illustration, the TV_{S^1} runtime without using our efficient formulation in Proposition 3 is about 12 minutes for 64 levels, and still about 2 minutes for 32 levels. The latter are also the runtimes for the more advanced regularizers such as truncated linear and Mumford-Shah from Sect. 6, since we use the same implementation scheme in

Sect. 6.5 for the non-local integral constraints in (58) and in (15). We used $n = 32$ levels for Γ in all our experiments.

7.1 The TV_{S^1}-L^2 Model

The analogon of the famous ROF denoising model of Rudin, Osher and Fatemi [25] for the cyclic case is

$$\inf_u \int_\Omega d_{S^1}\big(u(x), f(x)\big)^2 dx + \lambda TV_{S^1}(u). \tag{72}$$

We apply this to a one dimensional example in Fig. 6. The signal f consisting of a monotonous and a constant part has been degraded by adding 10 % Gaussian noise producing numerous wrap-arounds. The results show that only TV_{S^1} is able to reconstruct the signal. Thus cyclic wrap-arounds are handled correctly by TV_{S^1}. In contrast, for TV there is no choice of λ leading to a reconstruction without heavy fluctuations or large displacements. This example clearly demonstrates the advantage of TV_{S^1} over noncyclic regularizers.

7.2 Cyclical Smoothing

A frequently encountered task in image processing is to produce a slightly smoothed version of a noisy signal. For functions with a linear range this can be easily accomplished by the standard convolution with a smoothing kernel. However, this simple approach is not applicable for functions with cyclic values. The reason is that the smoothing also occurs across jumps of length 1 which do not constitute real jumps in the S^1 sense but only a change of representation.

In contrast, as can be seen in Fig. 7 the proposed quadratic S^1-penalizer from Sect. 6.2 indeed provides a smoothing effect of cyclical signals while preserving jumps of length 1.
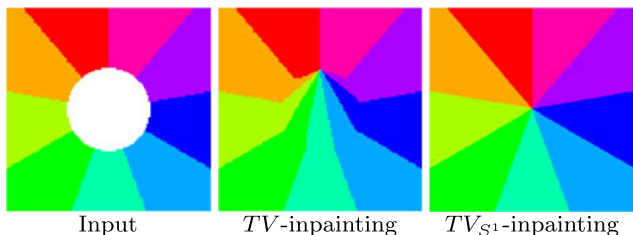
7.3 Inpainting

Figure 8 shows an inpainting experiment for the periodic hue values of the HSV color space. Values $0/9, \ldots, 8/9$ are circularly arranged starting from the top left red region with

Noisy input   Linear smoothing S^1-smoothing
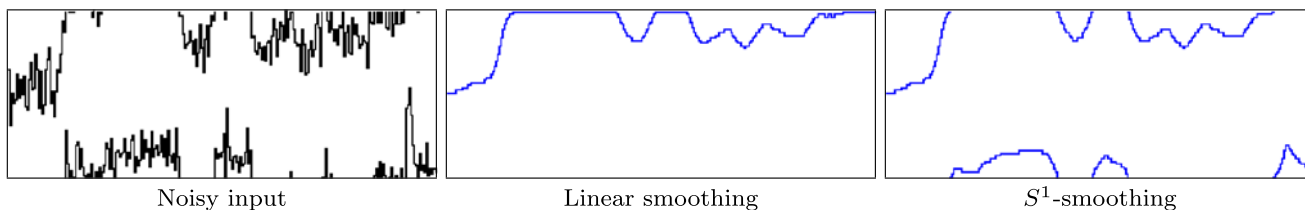
Fig. 7 Smoothing of angle values. The proposed quadratic penalizer (*right*) for S^1 structures in Sect. 6.2 correctly recognizes wrap-arounds of cyclic values and produces a smoothed signal which naturally resembles the original noisy one (*left*). In contrast, the usual quadratic penalizer (*center*) for functions with linear range does not allow any value jumps at all. The resulting smoothed signal (with the cyclic data term as in (72)) tends to stick either at value 0 or 1 in the presence of wrap-arounds, and forfeits essential signal details

Input TV-inpainting TV_{S^1}-inpainting

Fig. 8 Inpainting of periodic hue values. While total variation regularization (*center*) does not handle wrap-arounds correctly (shrinking the interface with the highest value jump, here from *magenta* to *red*) the proposed TV_{S^1} formulation (*right*) is designed to provide an optimal solution for cyclic structures at no additional cost (Color figure online)

value 0 and ending with the top right magenta region with value 8/9. We use model (72) with $\lambda = 1$ and the weighted variant of the data term $\varrho(x, t) = g(x)d_{S^1}(t, f(x))^2$, setting $g(x) = 0$ for points x within the inpainting circle area and $g(x) = \infty$ otherwise.

As Fig. 8 shows, TV_{S^1} produces the expected symmetric result, as opposed to TV. Thus, the image prior for cyclic values in regions with no available data is more natural with TV_{S^1} than with TV.

7.4 Phase Denoising

Synthetic aperture radars (SAR) [8] capture terrain elevations by means of interferometry. A satellite sends out radio waves of certain wavelength and records the echoed waves going back from earth surface, producing a complex valued image. Two such images taken from different view points or at different times are combined to a *phase differences* image of the respective complex numbers at each point. The phase differences $\Delta\phi \in [0, 1)$ are essentially proportional to the *wrapped* unknown ground height H,

$$\Delta\phi = c(H \bmod 1) \tag{73}$$

with some factor $c > 0$. Phase unwrapping techniques aim to reconstruct H from $\Delta\phi$. As the phase images $\Delta\phi$ are noisy, a denoising preprocessing step is typically required. Because the phases are cyclic, our framework of S^1 regularizers applies naturally here.

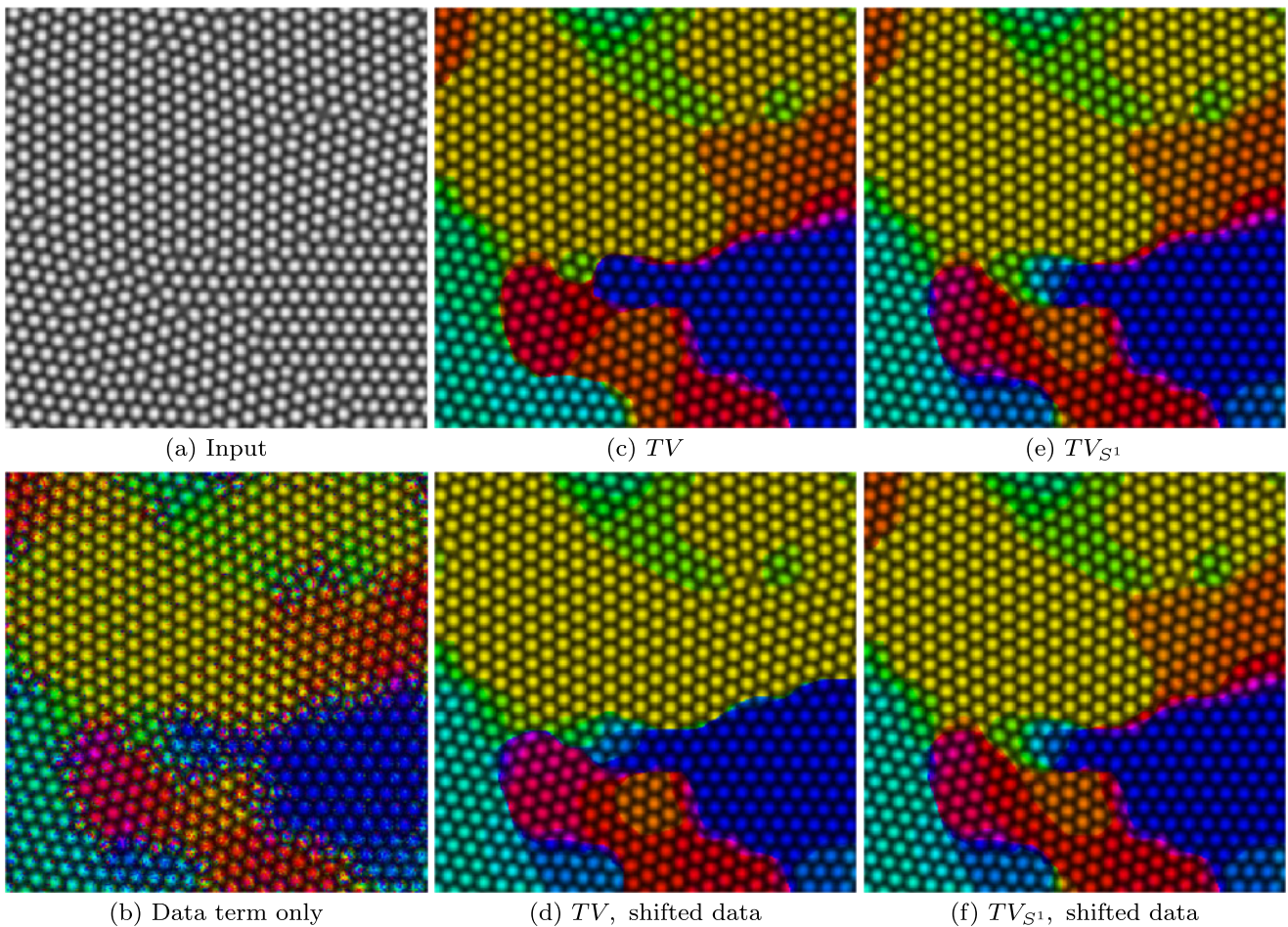Noisy phase TV_{S^1}-denoising
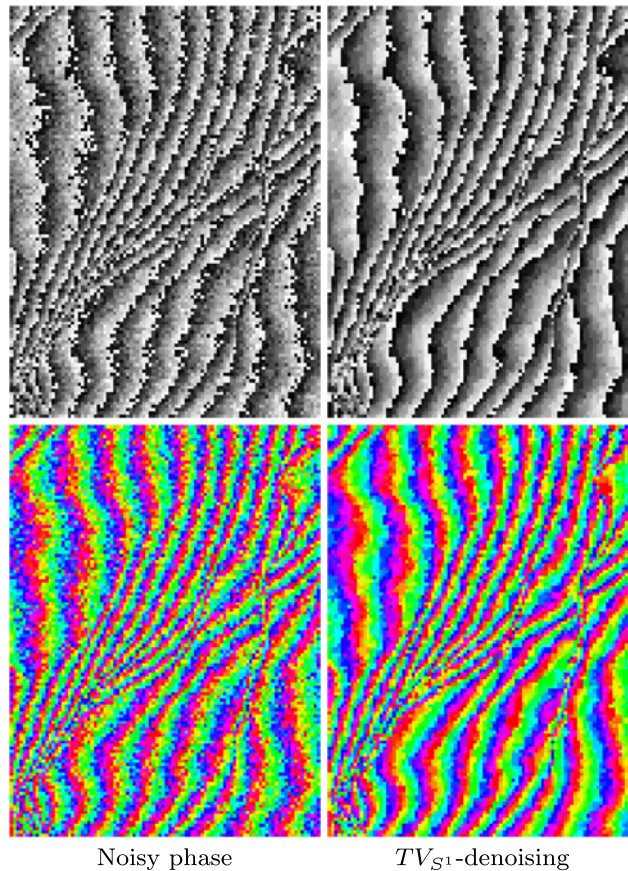
Fig. 9 Denoising of cyclic phase images from synthetic aperture radar (SAR). The phase measurements objects of interest (*left*), e.g. terrain elevation, are typically rather noisy. Proposed regularizers such as TV_{S^1} provide a natural method to obtain a denoised result (*right*). *Top row* shows the representatives of the phases from $\Gamma = [0, 1)$ and the *bottom row* the same phases using continuous hue coloring

Figure 9 shows phase denoising using $TV_{S^1}^g$ on the image "longs" from the publicly available data from [8].[1] We added 10 % Gaussian noise to this simulated clean phase image, wrapping the noisy values outside of [0, 1) back to this interval. We use the model (72) with $\lambda = 0.1$ and apply

---

[1] ftp://ftp.wiley.com/public/sci_tech_med/phase_unwrapping/data.zip.

(a) Input   (c) TV (e) TV_{S^1}

(b) Data term only   (d) TV, shifted data   (f) TV_{S^1}, shifted data

Fig. 10 Segmentation of regions with homogeneous atom lattice *orientation*, coloring different orientations by different hues. This example demonstrates the cyclic shift invariance of TV_{S^1} in Proposition 5. The results in (**e**) for the original data term $\varrho(x, t)$ and in (**f**) for the cyclically shifted one, $\varrho(x, t - 0.5 \bmod 1)$, are nearly indistinguishable using TV_{S^1} (coloring in (**f**) is shifted accordingly for comparison). In contrast, using TV the solution may be different for different zero orientations (**c**–**d**) (Color figure online)

the weighted variant $TV_{S^1}^g$ in (5). The weight g is set to the "correlation" which is available as part of the data. This is an image with $(0, 1]$ values which are inversely proportional to the local standard deviation of $\Delta\phi$ in a small window.

7.5 Atom Lattice Orientations

Many material properties such as deformation behavior can be deduced knowing the structures at atomic scale. In particular one is interested in the segmentation of grains, i.e. regions with homogeneous *orientation* of the atom lattice. To this end, when smooth transitions between neighboring regions are allowed Boerdgen et al. [4] used TV regularization with a nonconvex data term.

Since orientations are cyclic, the total *cyclic* variation TV_{S^1} is a more natural regularizer for this problem. Figure 10 shows a comparison of segmentations using TV_{S^1} and TV, applied to an image obtained by the phase field crystal simulation model [27]. This experiment shows the

solution behavior if one uses different representations of the orientation space S^1 as $[0, 1)$. The two solutions with TV_{S^1} are essentially identical due to the cyclic shift invariance in Proposition 5. On the contrary, with TV they may differ and important structures may not be recognized in the segmentation. For example, in Fig. 10 the upper right red area of (c) is missing in (d).

8 Conclusion

We have introduced a novel kind of total variation, TV_{S^1}, for cyclic structures as well as cyclic versions of other more general regularizers such as Huber-TV and Mumford-Shah. The regularizer TV_{S^1} penalizes the value jumps with the S^1 distance instead of linearly. A convex formulation is obtained through a recent theory for general functionals using functional lifting. We show that TV_{S^1} has a number of useful mathematical properties such as invariance to cyclic shifts

and lower-semicontinuity. The framework allows to couple the TV_{S^1} regularizer with arbitrary data terms, subject only to natural conditions. We show existence of minimizers and provide an equivalent formulation which allows to solve the optimization problem as efficiently as the usual total variation. Experiments on a variety of imaging problems show the clear advantage of cyclic regularizers in the correct handling of value wrap-arounds as opposed to noncyclic ones such as TV.

References

1. Alberti, G., Bouchitté, G., Maso, G.D.: The calibration method for the Mumford-Shah functional and free-discontinuity problems. Calc. Var. Partial Differ. Equ. **3**(16), 299–333 (2003)
2. Ambrosio, L., Fusco, N., Pallara, D.: Functions of Bounded Variation and Free Discontinuity Problems. Oxford University Press, Oxford (2000)
3. Bertalmio, M., Vese, L., Sapiro, G., Osher, S.J.: Simultaneous structure and texture image inpainting. IEEE Trans. Image Process. **12**(8), 882–889 (2003)
4. Boerdgen, M., Berkels, B., Rumpf, M., Cremers, D.: Convex relaxation for grain segmentation at atomic scale. In: VMV, pp. 179–186 (2010)
5. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Trans. Pattern Anal. Mach. Intell. **23**(11), 1222–1239 (2001)
6. Brox, T., Bruhn, A., Papenberg, N., Weickert, J.: High accuracy optical flow estimation based on a theory for warping. In: ECCV, pp. 25–36 (2004)
7. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. J. Math. Imaging Vis. **40**, 120–145 (2011)
8. Ghiglia, D.C., Pritt, M.D.: Two-Dimensional Phase Unwrapping: Theory, Algorithms and Software. Wiley, New York (1998)
9. Giaquinta, M., Modica, G., Souček, J.: Variational problems for maps of bounded variation with values in s^1. Calc. Var. Partial Differ. Equ. **1**(1), 87–121 (1993)
10. Giaquinta, M., Modica, G., Souček, J.: Cartesian Currents in the Calculus of Variations, I: Cartesian Currents. Springer, Berlin (1998)
11. Giaquinta, M., Mucci, D.: The BV-energy of maps into a manifold: relaxation and density results. Ann. Sc. Norm. Super. Pisa, Cl. Sci. **5**(4), 483–548 (2006)
12. Giaquinta, M., Mucci, D.: Maps of bounded variation with values into a manifold: total variation and relaxed energy. Pure Appl. Math. Q. **3**(2), 513–538 (2007)
13. Kleinberg, J., Tardos, E.: Approximation algorithms for classification problems with pairwise relationships: metric labeling and Markov random fields. In: FOCS, pp. 14–23 (1999)
14. Kolev, K., Klodt, M., Brox, T., Cremers, D.: Continuous global optimization in multiview 3d reconstruction. Int. J. Comput. Vis. **84**(1), 80–96 (2009)
15. Lellmann, J., Kappes, J.H., Yuan, J., Becker, F., Schnörr, C.: Convex multi-class image labeling by simplex-constrained total variation. Tech. rep, IWR, Univ. Heidelberg (2008)
16. Marquina, A., Osher, S.J.: Image super-resolution by TV regularization and Bregman iteration. J. Sci. Comput. **37**(3), 367–382 (2008)
17. Nikolova, M., Esedoglu, S., Chan, T.F.: Algorithms for finding global minimizers of image segmentation and denoising models. SIAM J. Appl. Math. **66**(5), 1632–1648 (2006)
18. Ostresh, L.M.: On the convergence of a class of iterative methods for solving the Weber location problem. Oper. Res. **26**(4) (1978)
19. Pock, T., Chambolle, A.: Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In: ICCV, pp. 1762–1769 (2011)
20. Pock, T., Chambolle, A., Bischof, H., Cremers, D.: A convex relaxation approach for computing minimal partitions. In: CVPR, pp. 810–817 (2009)
21. Pock, T., Cremers, D., Bischof, H., Chambolle, A.: An algorithm for minimizing the piecewise smooth Mumford-Shah functional. In: ICCV (2009)
22. Pock, T., Cremers, D., Bischof, H., Chambolle, A.: Global solutions of variational models with convex regularization. SIAM J. Imaging Sci. **3**(4), 1122–1145 (2010)
23. Pock, T., Schoenemann, T., Graber, G., Bischof, H., Cremers, D.: A convex formulation of continuous multi-label problems. In: ECCV, pp. 792–805 (2008)
24. Rockafellar, R.T.: Convex Analysis. Princeton University Press, Princeton (1996)
25. Rudin, L.I., Osher, S.J., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60**(1–4), 259–268 (1992)
26. Shulman, D., Herve, J.: Regularization of discontinuous flow fields. In: Visual Motion, pp. 81–86 (1989)
27. Stefanovic, P., Haataja, M., Provatas, N.: Phase field crystal study of deformation and plasticity in nanocrystalline materials. Phys. Rev. E **80**(4), 046,107 (2009)
28. Strekalovskiy, E., Cremers, D.: Total variation for cyclic structures: convex relaxation and efficient minimization. In: CVPR, pp. 1905–1911 (2011)

Daniel Cremers received a MS degree (Diplom) in Theoretical Physics (1997) from the University of Heidelberg and a PhD degree in Computer Science (2002) from the University of Mannheim, Germany. Subsequently he spent two years as a postdoctoral researcher at the UCLA and one year as a permanent researcher at Siemens Corporate Research in Princeton. From 2005 until 2009 Prof. Cremers headed the Computer Vision Group at the University of Bonn, Germany. Since 2009 he is full professor for computer science and mathematics at the TU München. He received several awards, including the *Best Paper of the Year 2003* by the Pattern Recognition Society, the *Olympus Award 2004*, and the *2005 UCLA Chancellor's Award*. In 2009 he was awarded an ERC Starting Grant. In December 2010 he was listed among "Germany's top 40 researchers below 40" (Capital).

Evgeny Strekalovskiy received a BS degree (Vordiplom) in Mathematics from the University of Jena in 2006 and a BS degree (Vordiplom) in Computer Science from the University of Bonn in 2007. He holds two MS degrees (Diplom), one in Mathematics received 2009, and one in Computer Science (with Honors) received 2010 from the University of Bonn, Germany. He is currently working towards his PhD degree in the Computer Vision Group at the Technical University of Munich, Germany. He was awarded the Intel Doctoral Student Honor Programe fellowship 2012. His current research focuses on convex optimization for imaging problems. His research interests include continuous variational methods, convex relaxation techniques and multilabel problems.