Convex relaxation of a class of vertex penalizing functionals

K. Bredies   T. Pock   B. Wirth

A–8010 GRAZ, HEINRICHSTRASSE 36, AUSTRIA

Convex relaxation of a class of vertex penalizing functionals

Kristian Bredies · Thomas Pock · Benedikt Wirth

Abstract We investigate a class of variational problems that incorporate in some sense curvature information of the level lines. The functionals we consider incorporate metrics defined on the orientations of pairs of line segments that meet in the vertices of the level lines. We discuss two particular instances: One instance that minimizes the total number of vertices of the level lines and another instance that minimizes the total sum of the absolute exterior angles between the line segments. In case of smooth level lines, the latter corresponds to the total absolute curvature. We show that these problems can be solved approximately by means of a tractable convex relaxation in higher dimensions. In our numerical experiments we present preliminary results for image segmentation, image denoising and image inpainting.

Kristian Bredies
Institute of Mathematics and Scientific Computing
University of Graz
Heinrichstraße 36
8010 Graz, Austria
E-mail: kristian.bredies@uni-graz.at

Thomas Pock
Institute for Computer Graphics and Vision
Graz University of Technology
Inffeldgasse 16
8010 Graz, Austria
E-mail: pock@icg.tugraz.at

Benedikt Wirth
Courant Institute of Mathematical Sciences
New York University
251 Mercer Street
New York 10012, NY, USA
E-mail: benedikt.wirth@cims.nyu.edu

1 Introduction

For almost three decades, smoothness of first-order derivatives has been the dominating regularization framework to solve ill-posed low-level vision problems such as image denoising, image segmentation, inpainting [19,4,28,33]. It has first been observed in the seminal work of Mumford [27] that higher order features such as the curvature of an object boundary gives a much stronger prior for tasks such as boundary completion and object disocclusion [29]. Since then, there is an increased interest in functionals incorporating higher order information, see for example [8,5].

There is also a strong evidence that curvature plays a dominant role in the human visual system [24,12]. A common choice, which is known to coincide with the main properties of Gestalt principles is to minimize the elastica functional

$$\int_\gamma (\alpha + \beta \kappa^2)\, \mathrm{d}\gamma, \tag{1.1}$$

where $\alpha > 0$, $\beta > 0$ are weighting parameters, γ is a smooth curve and κ is its curvature. A related and well studied functional is the so-called Willmore energy [40]

$$\frac{1}{2} \int_\gamma \kappa^2\, \mathrm{d}\gamma, \tag{1.2}$$

where now γ is a smooth, closed surface, κ is its mean curvature and $\mathrm{d}\gamma$ is the induced surface measure. Gradient flows of this energy have been studied for geometric problems based on level-set formulations [14], convolution thresholding schemes [22] or more recently by a two-step time discretization scheme [18].

The concept of curvature regularity of object boundaries has been generalized to whole images in [26,2]. The main

idea is to impose the curvature regularity to each single level line of a gray value image which leads to a functional

$$\int_\Omega |\nabla u| \left(\alpha + \beta \left(\operatorname{div} \frac{\nabla u}{|\nabla u|} \right)^2 \right) \, dx \, , \qquad (1.3)$$

where Ω is the image domain and $u \in \mathscr{C}_c^2(\Omega, \mathbb{R})$ is the image function. In [26], the application to image inpainting problems is shown, where the model yields faithful reconstructions of missing image data. Remarkably, the proposed algorithm based on dynamic programming computes a globally optimal solution. For related work see also [3, 10, 38] where different algorithms are proposed to minimize the elastica functional. Also related, researchers study partial differential equations (PDEs) related to the gradient flow of (1.3). See for example [39, 11].

The reason that held researchers off from using curvature depending functionals for practical imaging problems is its strong non-convexity, which makes a global minimization a very hard task. It is only recently that researchers started to work on global minimization algorithms for general curvature depending energy functionals. In [34, 35], the authors proposed in a discrete setting an integer linear programming formulation of curvature minimizing energy functionals. The method basically works by discretizing all possible combinations of oriented boundary elements which results in a huge integer linear program (LP). The integer LP is then solved by standard LP relaxation which for many practical image segmentation and inpainting problems leads to near-optimal solutions. A related, but simpler approach has been presented in [16] by means of a Markov random field formulation incorporating higher-order cliques. Recently, a convex relaxation of the Menger curvature of a characteristic function has been proposed in [20], but a generalization to general images remains unclear. Image segmentations with curvature regularization can also be obtained for certain active contour-type models which are based on so-called ratio functionals [36]. The minimizer is obtained by a minimum ratio cycle algorithm applied to a graph that represents all possible discrete curves in the image. Although this approach can lead to globally optimal solutions, it is still a relaxation of the original problem in the sense that it does not explicitly exclude self-intersections. Furthermore, the model is restricted to binary segmentation problems and cannot incorporate region-based fidelity terms.

In a paper by Citti and Sarti [12] it has been shown that by lifting (1.3) to the so-called *roto-translation space*, (1.3) reduces to a minimal surface problem in higher dimensions. The idea is to consider the original problem in a higher dimensional space, where the additional dimension is given by a local orientation $\vartheta \in S^1$ tangential to the image gradient. The level lines of the function u lifted to this space are then tangent to the vector fields $(\vartheta, 0)$ and $(0, 1)$ and
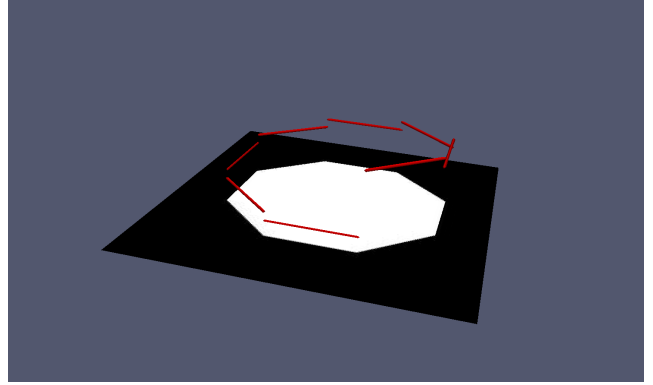

Fig. 1 Lifting of a binary image u of an octagon to the roto-translation space $\Omega \times S^1$. The red lines refer to the measure $|\nabla u|$ which is only supported at the edges of the octagon. According to their orientations, they appear as straight lines at different heights in the roto-translation space.

its directional derivatives are given by $X_1 = \vartheta_1 \partial_{x_1} + \vartheta_2 \partial_{x_2}$ and $X_2 = \partial_\vartheta$. The authors consider a generalized function $v : \Omega \times S^1 \to \mathbb{R}^+$ that can be identified with the measure ∇u in the lifted space. See Figure 1 for an example where a simple binary image is lifted to the roto-translation space. Surprisingly, it turns out that in this representation, the non-convex higher order two-dimensional problem (1.3) can be written in terms of a non-convex energy depending only on the first order derivatives X_1 and X_2. The authors showed applications to disocclusion problems using an iterative diffusion and concentration process by application of the so-called sub-Laplacian operator in the lifted space.

In this paper we make use of the idea of Citti and Sarti and show that it can be used to find a convex representation of a certain class of vertex penalizing (related to curvature minimizing) functionals. We show that the lifted representation v of the image can be related to the image u itself by means of linear constraints. This allows us to utilize the convex representation of the curvature in a convex regularization framework for general imaging problems. The paper is organized as follows: In Section 2 we give a precise mathematical definition of the functional lifting idea of Citti and Sarti using measures. In Section 3 we present the class of vertex penalizing functionals by means of a lower semi-continuous metric on the continuous label space of orientations of the level lines of the image. In Section 4 we describe a convex relaxation of the functionals, which makes the functional amenable as a regularizer for general imaging problems. In Section 5 we give a finite differences discretization of the energies and show how to efficiently minimize the resulting convex programs using a first-order primal-dual algorithm. In the last section we give a conclusion and discuss possible directions for future investigations.

2 Functional lifting

We propose, in the following, a representation of the gradient of a function of bounded variation which is lifted to a positive measure on the higher-dimensional space $\Omega \times S^1$ where $\Omega \subset \mathbb{R}^2$ is a bounded domain and S^1 denotes the unit circle in $\mathbb{R}^2$. But first, as we make frequent use of it, let us shortly recall the definition of the space $\mathrm{BV}(\Omega)$, for detail we refer to the literature, for example [1].

Definition 2.1. Let $\Omega \subset \mathbb{R}^2$ be a domain and $u \in L^1(\Omega)$. Then, the *total variation* of u is

$$\mathrm{TV}(u) = \sup \left\{ \int_\Omega u \operatorname{div} \varphi \, \mathrm{d}x \,\Big|\, \varphi \in \mathscr{C}_c^\infty(\Omega, \mathbb{R}^2), \|\varphi\|_\infty \le 1 \right\}.$$

The space of functions of *bounded variation* is the set

$$\mathrm{BV}(\Omega) = \{ u \in L^1(\Omega) \mid \mathrm{TV}(u) < \infty \}$$

endowed with the norm $\|u\|_{\mathrm{BV}} = \|u\|_1 + \mathrm{TV}(u)$.

Note that $\mathrm{BV}(\Omega)$ is a Banach space. For further properties, let us also mention some basic measure-theoretic notions.

Definition 2.2. Denote by $\mathscr{B}(\Omega)$ the Borel algebra generated by the open subsets of Ω. An $\mathbb{R}^d$-valued *finite Radon measure* is a countably additive, regular set function $\mu : \mathscr{B}(\Omega) \to \mathbb{R}^d$ with $\mu(\emptyset) = 0$.

A $\mathbb{R}$-valued finite Radon measure μ is *positive*, denoted $\mu \ge 0$, if $\mu(E) \ge 0$ for all $E \in \mathscr{B}(\Omega)$.

The *total variation measure* $|\mu|$ of a finite Radon measure is defined as

$$|\mu|(E) = \sup \left\{ \sum_{n=0}^\infty |\mu(E_n)| \,\Big|\, E_n \in \mathscr{B}(\Omega) \text{ pairwise disjoint}, \right.$$
$$\left. \Omega = \bigcup_{n=0}^\infty E_n \right\}.$$

The space of Radon measures is the set

$$\mathscr{M}(\Omega, \mathbb{R}^d) = \{ \mu : \mathscr{B}(\Omega) \to \mathbb{R}^d \mid \mu \text{ Radon measure} \}$$

endowed with the norm $\|\mu\|_{\mathscr{M}} = |\mu|(\Omega)$.

The total-variation measure of a $\mathbb{R}^d$-valued finite Radon measure is always a positive finite Radon measure. With the norm $\|\cdot\|_{\mathscr{M}}$, $\mathscr{M}(\Omega, \mathbb{R}^d)$ becomes a Banach space. It is well-known that $\mathscr{M}(\Omega, \mathbb{R}^d)$ can be identified with the dual space of $\mathscr{C}_0(\Omega, \mathbb{R}^d)$ by the integral

$$\varphi \mapsto \int_\Omega \varphi \, \mathrm{d}\mu.$$

In particular, for $u \in \mathrm{BV}(\Omega)$, the distributional derivative is a Radon measure, i.e., $\nabla u \in \mathscr{M}(\Omega, \mathbb{R}^2)$. It is absolutely continuous with respect to its total variation measure, hence, there is a *density* $\sigma \in L^1_{|\nabla u|}(\Omega, \mathbb{R}^2)$ such that $\nabla u = \sigma |\nabla u|$.

It can be shown that $\sigma(x) \in S^1$ $|\nabla u|$-almost everywhere. Therefore, the pair $(\sigma, |\nabla u|)$ is called the *polar decomposition* of ∇u.

Finally, let us define the operation which rotates a $\vartheta \in S^1$ counterclockwise by $\frac{\pi}{2}$:

$$\vartheta^\perp = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \vartheta.$$

This way, $-\sigma(x)^\perp$ points tangential to the level sets of u oriented in such a way that the function u is increasing on the "left-hand side".

With these prerequisites, the definition of the functional lifting of ∇u for $u \in \mathrm{BV}(\Omega)$ reads as follows.

Definition 2.3. Let $u \in \mathrm{BV}(\Omega)$, denote by $|\nabla u| \in \mathscr{M}(\Omega)$ the total variation measure of $\nabla u \in \mathscr{M}(\Omega, \mathbb{R}^2)$ and by $\sigma \in L^\infty_{|\nabla u|}(\Omega, \mathbb{R}^2)$ the density of ∇u with respect to $|\nabla u|$. We define the *functional lifting of* ∇u as the measure $\mu = \mu(\nabla u) \in \mathscr{M}(\Omega \times S^1)$ with

$$\int_{\Omega \times S^1} \varphi \, \mathrm{d}\mu = \int_\Omega \varphi\big(x, -\sigma(x)^\perp\big) \, \mathrm{d}|\nabla u|,$$

for each $\varphi \in \mathscr{C}_0(\Omega \times S^1)$.

One can easily see that $x \mapsto \big(x, -\sigma(x)^\perp\big)$ is measurable (with respect to $|\nabla u|$) between $\Omega \to \Omega \times S^1$. Thus, the functional μ indeed defines a measure in $\mathscr{M}(\Omega \times S^1)$ which can easily verified to be positive. Note again that the function φ is only integrated on the corresponding tangential direction $-\sigma^\perp$.

Also, observe that the measures $|\nabla u|$ and ∇u can be recovered from μ by

$$\int_\Omega \varphi \, \mathrm{d}|\nabla u| = \int_{\Omega \times S^1} \varphi(x) \, \mathrm{d}\mu(x, \vartheta)$$

for all $\varphi \in \mathscr{C}_0(\Omega)$, as well as, since $-\vartheta^{\perp\perp} = \vartheta$,

$$\int_\Omega \varphi \cdot \mathrm{d}\nabla u = \int_{\Omega \times S^1} \varphi(x) \cdot \vartheta^\perp \, \mathrm{d}\mu(x, \vartheta)$$

for all $\varphi \in \mathscr{C}_0(\Omega, \mathbb{R}^2)$. Here, $\vartheta \in S^1$ denotes the angular component in the space $\Omega \times S^1$, a notation we will frequently use throughout the paper. In the next section, Section 3, we aim at establishing functionals which, applied to μ, incorporate, up to a certain extend, curvature information from the level sets of u. For this purpose, we proceed as follows. First, for a given metric ρ on $S^1 \times S^1$ we introduce a "norm" on the space $\mathscr{M}(S^1)$. We will then use this norm to introduce a functional T_ρ on $\mathscr{M}(\Omega \times S^1)$ with the following property. For u being the characteristic function of a polygon P and μ being the functional lifting of ∇u, the penalty $T_\rho(\mu)$ corresponds to summing up, over all vertices, $\rho(\vartheta_1, \vartheta_2)$ where ϑ_1, ϑ_2 are the unit directions associated with the line segments meeting in a vertex.

In particular, this allows us to define a T_{ρ_0} such that

$$T_{\rho_0}(\mu) = \#\{\text{vertices of } P\}$$

and a T_{ρ_1} such that

$$T_{\rho_1}(\mu) = \sum_{x \in \text{vertices in } P} \gamma(x)$$

where $\gamma(x)$ is the unsigned external angle in x, i.e., the absolute value of the external angle in x. For the latter we will also see that for u and μ the characteristic function of a smooth set Ω' and the functional lifting of ∇u, respectively, it holds that

$$T_{\rho_1}(\mu) = \int_{\partial \Omega'} |\kappa| \, \mathrm{d}\mathscr{H}^1$$

where κ denotes the curvature of the boundary of Ω'. Afterwards, in Section 4, these notions will be used to define relaxed functionals acting on images rather than the functional lifting of the gradient. Such functionals turn out to be proper, convex and lower semi-continuous and it will be shown that they are also suitable for the solution of various imaging problems.

3 A vertex penalization functional

Let us mention once again that our motivation is to derive a suitable class of functionals which act, on the one hand, on the functional lifting of the gradient of characteristic functions and, on the other hand, are able to penalize the "vertices" of the boundary of a set. We will, for this purpose, discuss a suitable class of polygons. But tentatively, for motivation, suppose that $\Omega = B_1(0)$ is the unit disc and $\mu \in \mathscr{M}(\Omega \times S^1)$ represents the integration on two unit line segments meeting in 0. See Figure 2 for an example of a vertex which is formed by two line segments with orientations $\vartheta_1, \vartheta_2 \in S^1$. Then

$$\int_{\Omega \times S^1} \varphi \, \mathrm{d}\mu = \int_{-1}^0 \varphi(t\vartheta_1, \vartheta_1) \, \mathrm{d}t + \int_0^1 \varphi(t\vartheta_2, \vartheta_2) \, \mathrm{d}t$$

for each $\varphi \in \mathscr{C}_0(\Omega \times S^1)$. This corresponds to the functional lifting of the gradient of a polygon with a single vertex in 0. Now, consider the distributional directional derivative of μ with respect to $(-\vartheta, 0)$. In order to compute this, we have to test with $(x, \vartheta) \mapsto \nabla_x \psi(x, \vartheta) \cdot \vartheta$ for $\psi \in \mathscr{C}_c^\infty(\Omega \times S^1)$:

$$\int_{\Omega \times S^1} \nabla_x \psi(x, \vartheta) \cdot \vartheta \, \mathrm{d}\mu$$
$$= \int_{-1}^0 \frac{\partial}{\partial t} \psi(t\vartheta_1, \vartheta_1) \, \mathrm{d}t + \int_0^1 \frac{\partial}{\partial t} \psi(t\vartheta_2, \vartheta_2) \, \mathrm{d}t$$
$$= \psi(0, \vartheta_1) - \psi(0, \vartheta_2) = \langle \delta_{\vartheta_1} - \delta_{\vartheta_2}, \psi(0, \cdot) \rangle$$

where δ_ϑ denotes the delta distribution at ϑ. Hence, in order to penalize vertices of a polygon, we have to find suitable functionals acting on $\mathscr{M}(S^1)$. Ideally, we would like
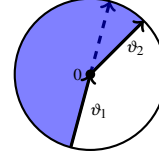

Fig. 2 Orientations of the line segments that form a vertex of a polygon (blue). According to Definition 2.3, the orientations $\vartheta_1, \vartheta_2 \in S^1$ are given such that the characteristic function of the polygon is increasing on the left-hand side of the line segments.

to prescribe the penalty for each difference of delta peaks, i.e., each pair of orientations. As it turns out in Subsection 3.1, this is indeed possible for a certain class of metrics on $S^1 \times S^1$. In some sense, the δ_ϑ correspond to the "corners" of the "unit simplex" in $\mathscr{M}(S^1)$ or, more precisely, are the extremal points of the set of probability measures in $\mathscr{M}(S^1)$. Interpreting each element in S^1 as a label, the problem is therefore to find a functional which measures the distance between these labels in a prescribed way. Therefore, we speak of S^1 as a continuous labeling space in analogy to the finite labeling spaces utilized, for instance, for multiclass labeling in image segmentation and stereo [7,31,25]. Having discussed this, we address, in Subsection 3.2, how to derive a vertex-penalizing functional to more general polygons and, eventually, to a functional which acts on the functional lifting of the gradient of characteristic functions.

3.1 The unit circle as a continuous labeling space

In the following, let ρ be given according to the following assumption.

Assumption 3.1. *Let $\rho : S^1 \times S^1 \to [0, \infty[$ such that*

1. *ρ defines a metric on S^1,*
2. *ρ is lower semi-continuous, i.e., $\rho = \sup_{i \in I} \rho_i$ where I is a non-empty index set and $\rho_i \in \mathscr{C}(S^1 \times S^1)$ is a continuous metric for each $i \in I$.*

As mentioned above, we would like to define a norm on the space $\mathscr{M}(S^1)$ for which the distance between two delta peaks matches the given metric ρ. For this purpose, we define the set

$$C_\rho = \{\varphi \in \mathscr{C}(S^1) \mid \varphi(\eta_1) - \varphi(\eta_2) \leq \rho(\eta_1, \eta_2) \quad \text{for all } (\eta_1, \eta_2) \in S^1 \times S^1\} \quad (3.1)$$

which is the predual unit ball of the weak* sequentially lower semi-continuous norm $\|\cdot\|_\rho$ on $\mathscr{M}(S^1)$, i.e., for $\mu \in \mathscr{M}(S^1)$,

$$\|\mu\|_\rho = \sup_{\varphi \in C_\rho} \langle \mu, \varphi \rangle. \quad (3.2)$$

Here, we say that a weak* sequentially lower semi-continuous norm is a non-negative, weak* sequentially

lower semi-continuous, positively homogeneous and positive definite functional which satisfies the triangle inequality but may also take the value ∞. The following proposition shows that it is at least justified to call $\|\cdot\|_\rho$ a weak* sequentially lower semi-continouos semi-norm.

Proposition 3.2. *The functional $\|\cdot\|_\rho : \mathscr{M}(S^1) \to [0,\infty]$ is sequentially weak* lower semi-continuous, positively homogeneous and satisfies the triangle inequality.*

Proof. By symmetry of ρ, we have that $\varphi \in C_\rho$ implies $-\varphi \in C_\rho$. Hence,

$$\|\mu\|_\rho = \sup_{\varphi \in C_\rho} |\langle \mu, \varphi \rangle| \geq 0,$$

so $\|\cdot\|_\rho$ is non-negative. The sequential weak* lower semi-continuity follows from the fact the functional is a pointwise supremum of sequentially weak* continuous functionals. For $\lambda \in \mathbb{R}$ we moreoever see that, utilizing $C_\rho = -C_\rho$,

$$\|\lambda\mu\|_\rho = \sup_{\varphi \in C_\rho} \langle |\lambda|\mu, \mathrm{sgn}(\lambda)\varphi \rangle$$
$$= |\lambda| \sup_{\varphi \in \mathrm{sgn}(\lambda)C_\rho} \langle \mu, \varphi \rangle = |\lambda| \|\mu\|_\rho$$

establishing the positive homogeneity. Finally, if $\mu_1, \mu_2 \in \mathscr{M}(S^1)$, then, for each $\varphi \in C_\rho$ we have

$$\langle \mu_1 + \mu_2, \varphi \rangle \leq \|\mu_1\|_\rho + \|\mu_2\|_\rho$$

which implies the triangle inequality by taking the supremum on the left-hand side. □

Remark 3.3. The functional $\|\mu\|_\rho$ is only finite if $\int_{S^1} 1 \, d\mu = 0$. Otherwise, since the constant functions are in C_ρ,

$$\sup_{\varphi \in C_\rho} \langle \mu, \varphi \rangle \geq \sup_{c \in \mathbb{R}} \int_{S^1} c \, d\mu = \infty.$$

Note that we did not yet show the positive definiteness as the proof is more involved. We first like to establish the identity

$$\|\delta_{\vartheta_1} - \delta_{\vartheta_2}\|_\rho = \rho(\vartheta_1, \vartheta_2)$$

for each $\vartheta_1, \vartheta_2 \in S^1$. This is first done for continuous metrics.

Proposition 3.4. *Let $\rho : S^1 \times S^1 \to [0,\infty[$ satisfy Assumption 3.1. Then, for each $\vartheta_1, \vartheta_2 \in S^1$ it holds that*

$$\sup_{\varphi \in C_\rho} \varphi(\vartheta_1) - \varphi(\vartheta_2) = \rho(\vartheta_1, \vartheta_2).$$

Proof. Note that according to the definition of C_ρ in (3.1), we have $\varphi(\vartheta_1) - \varphi(\vartheta_2) \leq \rho(\vartheta_1, \vartheta_2)$ for all $\varphi \in C_\rho$, so

$$\sup_{\varphi \in C_\rho} \varphi(\vartheta_1) - \varphi(\vartheta_2) \leq \rho(\vartheta_1, \vartheta_2).$$

For the converse inequality, first assume that $\rho : S^1 \times S^1 \to [0,\infty[$ is continuous. Then, $\varphi^* \in \mathscr{C}(S^1)$, given by $\varphi^*(\vartheta) = \rho(\vartheta, \vartheta_2)$ is contained in C_ρ: For $\eta_1, \eta_2 \in S^1$ it holds that

$$\varphi^*(\eta_1) - \varphi^*(\eta_2) = \rho(\eta_1, \vartheta_2) - \rho(\eta_2, \vartheta_2) \leq \rho(\eta_1, \eta_2)$$

since ρ satisfies the triangle inequality. Hence,

$$\sup_{\varphi \in C_\rho} \varphi(\vartheta_1) - \varphi(\vartheta_2) \geq \varphi^*(\vartheta_1) - \varphi^*(\vartheta_2) = \rho(\vartheta_1, \vartheta_2).$$

If $\rho = \sup_{i \in I} \rho_i$ for I being a non-empty index set and each ρ_i being continuous, it follows, as each $C_{\rho_i} \subset C_\rho$, that

$$\sup_{\varphi \in C_\rho} \varphi(\vartheta_1) - \varphi(\vartheta_2) \geq \sup_{i \in I} \sup_{\varphi \in C_{\rho_i}} \varphi(\vartheta_1) - \varphi(\vartheta_2)$$
$$\geq \sup_{i \in I} \rho_i(\vartheta_1, \vartheta_2) = \rho(\vartheta_1, \vartheta_2).$$

This shows the desired identity. □

As an immediate consequence, we get the positive definiteness of $\|\cdot\|_\rho$.

Proposition 3.5. *For ρ satisfying Assumption 3.1, the functional $\|\cdot\|_\rho$ is positive definite.*

Proof. Let $\mu \in \mathscr{M}(S^1)$ such that $\|\mu\|_\rho = 0$, i.e., $\langle \mu, \varphi \rangle \leq 0$ for all $\varphi \in C_\rho$. It is then clear that with

$$V = \bigcup_{\lambda \geq 0} \lambda C_\rho$$

it also follows that $\langle \mu, \varphi \rangle \leq 0$ for all $\varphi \in V$. We aim at proving that V is dense in $\mathscr{C}(S^1)$. For this purpose, the prerequisites of the Stone-Weierstraß theorem are verified [13, Theorem 17.1]. First, from Proposition 3.4, it follows that C_ρ is separating as for each $\vartheta_1, \vartheta_2 \in S^1$ with $\vartheta_1 \neq \vartheta_2$, there is an $i \in I$ and a function $\varphi \in C_\rho$ such that $\varphi(\vartheta_1) - \varphi(\vartheta_2) = \rho_i(\vartheta_1, \vartheta_2) > 0$. Moreover, C_ρ obviously contains the constant functions. Hence, V is separating and contains the constant functions.

Next, we like to verify that V is a subalgebra of $\mathscr{C}(S^1)$. For this purpose, let $\varphi \in V$ and $\alpha \in \mathbb{R}$. Then, $\varphi = \lambda \varphi_0$ for some $\lambda \geq 0$ and $\varphi_0 \in C_\rho$. Since $C_\rho = -C_\rho$, it follows that $\mathrm{sgn}(\alpha)\varphi_0 \in C_\rho$ and hence, $\alpha\varphi = |\alpha|\lambda \, \mathrm{sgn}(\alpha)\varphi_0 \in V$. For $\varphi, \psi \in V$ we can find $\lambda_1, \lambda_2 \geq 0$ and $\varphi_0, \psi_0 \in C_\rho$ such that $\varphi = \lambda_1 \varphi_0$ and $\psi = \lambda_2 \psi_0$. Therefore, for each $\vartheta_1, \vartheta_2 \in S^1$, it holds that

$$(\varphi + \psi)(\vartheta_1) - (\varphi + \psi)(\vartheta_2)$$
$$= \lambda_1 (\varphi_0(\vartheta_1) - \varphi_0(\vartheta_2)) + \lambda_2 (\psi_0(\vartheta_1) - \psi_0(\vartheta_2))$$
$$\leq (\lambda_1 + \lambda_2)\rho(\vartheta_1, \vartheta_2)$$

implying that $\lambda^{-1}(\varphi + \psi) \in C_\rho$ for any $\lambda > \lambda_1 + \lambda_2$ and, consequently, that $\varphi + \psi \in V$. Also,

$$(\varphi\psi)(\vartheta_1) - (\varphi\psi)(\vartheta_2) = \lambda_1\lambda_2\Big(\psi_0(\vartheta_1)\big(\varphi_0(\vartheta_1) - \varphi_0(\vartheta_2)\big)$$
$$+ \varphi_0(\vartheta_2)\big(\psi_0(\vartheta_1) - \psi_0(\vartheta_2)\big)\Big)$$
$$\leq \lambda_1\lambda_2\big(\|\varphi_0\|_\infty + \|\psi_0\|_\infty\big)\rho(\vartheta_1, \vartheta_2)$$

from which follows that $\lambda^{-1}\varphi\psi \in C_\rho$ for each $\lambda > \lambda_1\lambda_2\big(\|\varphi_0\|_\infty + \|\psi_0\|_\infty\big)$. Hence $\varphi\psi \in V$.

The Stone-Weierstraß theorem now gives the density of V in $\mathscr{C}(S^1)$. As $\varphi \mapsto \langle\mu, \varphi\rangle$ is a continuous functional, it follows that $\langle\mu, \varphi\rangle \leq 0$ for all $\varphi \in \mathscr{C}(S^1)$. This is only possible if $\mu = 0$. Hence, $\|\cdot\|_\rho$ is positive definite. $\qquad\square$

The results of this section can be summarized in the following theorem.

Theorem 3.6. *For ρ satisfying Assumption 3.1, $\|\cdot\|_\rho : \mathscr{M}(S^1) \to \mathbb{R} \cup \{\infty\}$ defines a functional with the following properties:*

1. *$\|\cdot\|_\rho$ is a weak* sequentially lower semi-continuous norm,*
2. *for each $\vartheta_1, \vartheta_2 \in S^1$, it holds that*

$$\|\delta_{\vartheta_1} - \delta_{\vartheta_2}\|_\rho = \rho(\vartheta_1, \vartheta_2).$$

Proof. The first item follows from the Propositions 3.2 and 3.5 while the second is just another way of writing Proposition 3.4. $\qquad\square$

Example 3.7. Let $\rho_0 : S^1 \times S^1 \to \{0, 1\}$ be the discrete metric, i.e. $\rho_0(\vartheta_1, \vartheta_2) = 0$ if $\vartheta_1 = \vartheta_2$ and 1 otherwise. This metric is admissible in the sense of Assumption 3.1 since

$$\rho_0(\vartheta_1, \vartheta_2) = \sup_{\lambda > 0} \min(1, \lambda^{-1}|\vartheta_1 - \vartheta_2|).$$

The set C_{ρ_0} then consists of all functions $\varphi \in \mathscr{C}(S^1)$ for which

$$\max_{\vartheta_1 \in S^1} \varphi(\vartheta_1) - \min_{\vartheta_2 \in S^1} \varphi(\vartheta_2) \leq 1.$$

This is exactly, up to a constant function, the ∞-ball of radius $\frac{1}{2}$, i.e.,

$$C_{\rho_0} = \mathbf{1}\mathbb{R} + \{\|\varphi\|_\infty \leq \tfrac{1}{2}\} = \mathbf{1}\mathbb{R} + \overline{\{\varphi \in \mathscr{C}^\infty(S^1), \|\varphi\|_\infty \leq \tfrac{1}{2}\}}$$

where $\mathbf{1}$ denotes the function constant 1 and the closure is being taken in $\mathscr{C}(S^1)$. Consequently, for $\mu \in \mathscr{M}(S^1)$ with $\int_{S^1} 1\,\mathrm{d}\mu = 0$ we have

$$\sup_{\varphi \in C_{\rho_0}} \langle\mu, \varphi\rangle = \sup_{\|\varphi\|_\infty \leq 1/2} \langle\mu, \varphi\rangle = \tfrac{1}{2}\|\mu\|_{\mathscr{M}}.$$

Hence, we can characterize $\|\cdot\|_{\rho_0}$ as follows (also see Remark 3.3):

$$\|\mu\|_{\rho_0} = \begin{cases} \frac{1}{2}\|\mu\|_{\mathscr{M}} & \text{if } \int_{S^1} 1\,\mathrm{d}\mu = 0 \\ \infty & \text{otherwise.} \end{cases}$$

Example 3.8. Choose $\rho_1 : S^1 \times S^1 \to [0, \pi]$ as the metric measuring geodesic distances in S^1. This can be expressed, for instance, as

$$\rho_1(\vartheta_1, \vartheta_2) = \min\big\{|t_1 - t_2| \,\big|\, \vartheta_i = \big(\cos(t_i), \sin(t_i)\big)$$
$$\text{for } i = 1, 2\big\}.$$

The set C_{ρ_1} then consists of all Lipschitz continuous functions in $\mathscr{C}(S^1)$ with Lipschitz constant not exceeding 1. This set can also be characterized by:

$$C_{\rho_1} = \{\varphi \in \mathscr{C}(S^1) \,\big|\, \varphi'(\vartheta) \text{ exists}$$
$$\text{for almost every } \vartheta \in S^1 \text{ and } \|\varphi'\|_\infty \leq 1\}$$
$$= \overline{\{\varphi \in \mathscr{C}^\infty(S^1) \,\big|\, \|\varphi'\|_\infty \leq 1\}}$$

where the closure is again taken in $\mathscr{C}(S^1)$. The functional $\|\cdot\|_{\rho_1}$ then corresponds to a dual Lipschitz norm.

3.2 Penalizing the vertices of a polygon

Let $\Omega \subset \mathbb{R}^2$ be a bounded Lipschitz domain. In the following, we consider a connected polygon according to the following definition.

Definition 3.9. We say that a relatively closed polygon $P \subset \Omega$ is an *admissible polygon* if

1. the line segments of P in Ω are (up to boundary points) given by

$$[x_i, y_i] = \{\lambda x_i + (1 - \lambda)y_i \mid \lambda \in [0, 1]\}$$

for $I \geq 3$ and $x_1, \ldots, x_I, y_1, \ldots, y_I \in \overline{\Omega}$,
2. the vertices are distinct, i.e., the collection x_i, $i = 1, \ldots, I$, as well as the collection y_i, $i = 1, \ldots, I$, are pairwise distinct,
3. the line segments are connected, i.e., for $i = 1, \ldots, I$ it holds that $x_{i+1} = y_i$ if $y_i \in \Omega$ and that y_i and x_{i+1} lie on the same connected component of $\partial\Omega$ if $y_i \in \partial\Omega$, where $x_{I+1} = x_1$,
4. the line segments $[x_i, y_i]$ are pairwise disjoint for $i = 1, \ldots, I$, i.e., the polygon does not intersect itself,
5. P lies on the left hand side with respect to the oriented segments $[x_i, y_i]$, i.e., for each $x = \lambda x_i + (1 - \lambda)y_i$ with $1 \leq i \leq I$ and $\lambda \in {]0, 1[}$ there exists a neighborhood of U of x such that

$$P \cap U = \{z \in U \mid (z - x) \cdot (y_i - x_i)^\perp \geq 0\}.$$

This notion is illustrated in Figure 3. Note that the characteristic function $u = \chi_P$ of each admissible polygon P
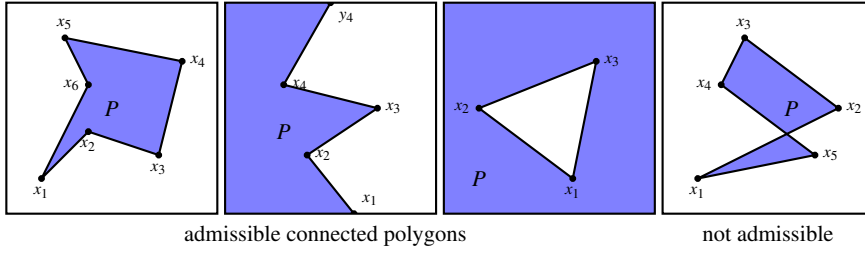
Fig. 3 Examples for polygons satisfying and contradicting Definition 3.9. The endpoints y_i are not marked if they coincide with a x_{i+1}.

admissible connected polygons   not admissible

is contained in $\mathrm{BV}(\Omega)$ with the derivative acting on a test function $\varphi \in \mathscr{C}_0(\Omega, \mathbb{R}^2)$ according to

$$\int_\Omega \varphi \cdot \mathrm{d}\nabla u = \sum_{i=1}^I \int_0^1 |y_i - x_i| \varphi\big(x_i(t)\big) \cdot \vartheta_i^\perp \,\mathrm{d}t,$$

$$x_i(t) = (1-t)x_i + ty_i, \qquad \vartheta_i = \frac{y_i - x_i}{|y_i - x_i|}, \qquad i = 1, \ldots, I. \tag{3.3}$$

Consequently, the functional lifting of ∇u according to Definition 2.3 can be expressed as follows: For $\varphi \in \mathscr{C}_0(\Omega \times S^1)$ we have

$$\int_{\Omega \times S^1} \varphi \,\mathrm{d}\mu = \sum_{i=1}^I \int_0^1 |y_i - x_i| \varphi\big(x_i(t), \vartheta_i\big) \,\mathrm{d}t.$$

Now, as we already did in the beginning of the section, take a function $\psi \in \mathscr{C}_0(\Omega \times S^1)$ for which $\nabla_x \psi \in \mathscr{C}_0(\Omega \times S^1, \mathbb{R}^2)$ and test μ against the function $\varphi \in \mathscr{C}_0(\Omega \times S^1)$ given by

$$\varphi(x, \vartheta) = \nabla_x \psi(x, \vartheta) \cdot \vartheta.$$

Then, with $\vartheta_{I+1} = \vartheta_1$,

$$\begin{aligned}
\int_{\Omega \times S^1} \varphi \,\mathrm{d}\mu &= \sum_{i=1}^I \int_0^1 |y_i - x_i| \nabla_x \psi\big(x_i(t), \vartheta_i\big) \cdot \vartheta_i \,\mathrm{d}t \\
&= \sum_{i=1}^I \int_0^1 \frac{\partial}{\partial t} \psi\big(x_i(t), \vartheta_i\big) \,\mathrm{d}t \\
&= \sum_{i=1}^I \psi(y_i, \vartheta_i) - \psi(x_i, \vartheta_i) \\
&= \sum_{\substack{1 \le i \le I, \\ y_i \in \Omega}} \psi(y_i, \vartheta_i) - \psi(y_i, \vartheta_{i+1}).
\end{aligned} \tag{3.4}$$

Choosing a metric according to Assumption 3.1 and requiring additionally that $\psi(x, \cdot) \in C_\rho$ for each $x \in \Omega$ where C_ρ is chosen according to (3.1), we can deduce

$$\int_{\Omega \times S^1} \varphi \,\mathrm{d}\mu \le \sum_{\substack{1 \le i \le I, \\ y_i \in \Omega}} \rho(\vartheta_i, \vartheta_{i+1}).$$

Taking the supremum over these functions one would expect, in view of Theorem 3.6, that the right-hand side will be attained. This motivates the following definition.

Definition 3.10. For ρ according to Assumption 3.1 and $\mu \in \mathscr{M}(\Omega \times S^1)$, let

$$T_\rho(\mu) = \sup_{\psi \in M_\rho(\Omega)} \int_{\Omega \times S^1} \nabla_x \psi(x, \vartheta) \cdot \vartheta \,\mathrm{d}\mu(x, \vartheta)$$

where

$$\begin{aligned}
M_\rho(\Omega) = \big\{ \psi \in \mathscr{C}_0(\Omega \times S^1) \,\big|\, &\nabla_x \psi \in \mathscr{C}_0(\Omega \times S^1, \mathbb{R}^2), \\
&\psi(x, \cdot) \in C_\rho \text{ for all } x \in \Omega \big\}.
\end{aligned}$$

Remark 3.11. The functional T_ρ can be interpreted as follows. Assume that μ is the product of the Lebesgue measure with a $\mathscr{C}^\infty(\overline{\Omega} \times S^1)$ density μ_0. Then, the integral in Definition 3.10 becomes, with ∇_ϑ denoting the directional derivative with respect to $(\vartheta, 0)$,

$$\begin{aligned}
\int_{\Omega \times S^1} &\nabla_x \psi(x, \vartheta) \cdot \vartheta \mu_0(x, \vartheta) \,\mathrm{d}(x, \vartheta) \\
&= -\int_{\Omega \times S^1} \nabla_\vartheta \mu_0(x, \vartheta) \psi(x, \vartheta) \,\mathrm{d}(x, \vartheta),
\end{aligned}$$

hence taking the supremum over all $\psi \in M_\rho(\Omega)$ yields, in view of (3.2),

$$T_\rho(\mu) = \int_\Omega \|\nabla_\vartheta \mu_0(x, \cdot)\|_\rho \,\mathrm{d}x.$$

Generally, T_ρ can thus be seen as a differentiation of μ, in the distributional sense, with respect to $-\nabla_\vartheta$ and the measurement of the derivative in the $\|\cdot\|_\rho$-sense in S^1 as well as in the Radon-norm sense in Ω.

Let us also note that T_ρ defines a weak* lower semicontinuous semi-norm on $\mathscr{M}(\Omega \times S^1)$.

Proposition 3.12. *The functional T_ρ is sequentially weak* lower semi-continuous, positively homogeneous and satisfies the triangle inequality.*

Proof. The statements can be deduced in analogy to Proposition 3.2 by replacing C_ρ with

$$\{(x, \vartheta) \mapsto \nabla_x \psi(x, \vartheta) \cdot \vartheta \mid \psi \in M_\rho(\Omega)\}. \qquad \square$$

The crucial observation now is that indeed, for admissible polygons, the functional T_ρ behaves like expected.

Proposition 3.13. *Let ρ be a metric satisfying Assumption 3.1. For P being an admissible polygon, $u = \chi_P$ and μ the lifting of ∇u it holds that*

$$T_\rho(\mu) = \sum_{\substack{1 \leq i \leq I, \\ y_i \in \Omega}} \rho\left(\frac{y_i - x_i}{|y_i - x_i|}, \frac{y_{i+1} - x_{i+1}}{|y_{i+1} - x_{i+1}|}\right).$$

Proof. As the vertices of P, i.e., $y_1, \ldots, y_I$ are pairwise distinct, one can find an $\varepsilon > 0$ such that $|y_i - y_j| > \varepsilon$ for each $1 \leq i, j \leq I$ and $|y_i - x| > \varepsilon$ for $y_i \in \Omega$ and $x \in \partial\Omega$. Choose a function $\zeta \in \mathscr{C}_c^\infty(B_\varepsilon(0))$ such that $0 \leq \zeta \leq 1$ as well as $\zeta(0) = 1$. Denote by $L = \{y_i \mid y_i \in \Omega, i = 1, \ldots, I\}$ the set of interior vertices and introduce $\vartheta^1, \vartheta^2 : L \to S^1$ such that, with $\vartheta_1, \ldots, \vartheta_{I+1}$ according to (3.3), $\vartheta^1(y) = \vartheta_i$ if $y = y_i$ as well as $\vartheta^2(y) = \vartheta_{i+1}$ if $y = y_i$. Then, for each $\bar{\psi} \in \mathscr{C}(L \times S^1)$ one can construct $\psi : \Omega \times S^1 \to \mathbb{R}$ as follows:

$$\psi(x, \vartheta) = \sum_{y \in L} \zeta(x - y)\bar{\psi}(y, \vartheta).$$

One can easily see that $\psi \in \mathscr{C}_0(\Omega \times S^1)$ with $\psi|_{L \times S^1} = \bar{\psi}$ and $\nabla_x \psi \in \mathscr{C}_0(\Omega \times S^1)$. Now, suppose that $\bar{\psi}(y, \cdot) \in C_\rho$ for each $y \in L$. Then, for each $x \in \Omega$ there is at most one y such that $\zeta(x - y) > 0$. In case no such y exists, then $\psi(x, \cdot) = 0 \in C_\rho$. Otherwise, for each $\eta_1, \eta_2 \in S^1$ we have, as $0 \leq \zeta \leq 1$,

$$\psi(x, \eta_1) - \psi(x, \eta_2) = \zeta(x - y)\big(\bar{\psi}(y, \eta_1) - \bar{\psi}(y, \eta_2)\big)$$
$$\leq \rho(\eta_1, \eta_2)$$

implying that $\psi(x, \cdot) \in C_\rho$. We already saw in (3.4) that

$$\int_{\Omega \times S^1} \nabla_x \psi(x, \vartheta) \cdot \vartheta \, d\mu = \sum_{y \in L} \psi(y, \vartheta^1(y)) - \psi(y, \vartheta^2(y)).$$

From Proposition 3.4 it follows that we can choose for each $N \geq 1$ a $\bar{\psi}^N \in \mathscr{C}(L \times S^1)$ such that $\bar{\psi}^N(y, \cdot) \in C_\rho$ and $\bar{\psi}^N(y, \vartheta^1(y)) - \bar{\psi}^N(y, \vartheta^2(y)) \geq \rho(\vartheta^1(y), \vartheta^2(y)) - 1/N$ for each $y \in L$. The corresponding $\psi^N : \Omega \times S^1 \to \mathbb{R}$ according to the above construction are in $M_\rho(\Omega)$ according to Definition 3.10, hence

$$T_\rho(\mu) \geq \sup_{N \geq 1} \sum_{y \in L} \rho(\vartheta^1(y), \vartheta^2(y)) - \frac{1}{N}$$
$$= \sum_{y \in L} \rho(\vartheta^1(y), \vartheta^2(y)).$$

This gives the result as the converse inequality is immediate by (3.4). $\qquad\square$

Example 3.14. For ρ_0 the discrete metric, see Example 3.7, it is easy to deduce that $T_{\rho_0}(\mu)$ corresponds to

$$T_{\rho_0}(\mu) = \#\left\{y_i \in \Omega \,\middle|\, \frac{y_i - x_i}{|y_i - x_i|} \neq \frac{y_{i+1} - x_{i+1}}{|y_{i+1} - x_{i+1}|}\right\}$$

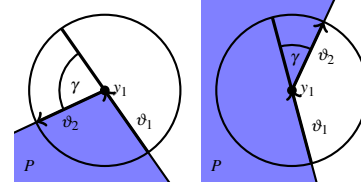which is the number of "genuine" vertices of P.

Fig. 4 Illustration of the undirected angle $\gamma(y_i) = \rho_1(\vartheta_1, \vartheta_2)$ with line segments meeting, in the end-point y_1, at an angle less and greater than π, respectively.

Example 3.15. For ρ_1 the geodesic metric on S^1, see Example 3.8, we have for $y_i \in \Omega$,

$$\rho_1\left(\frac{y_i - x_i}{|y_i - x_i|}, \frac{y_{i+1} - x_{i+1}}{|y_{i+1} - x_{i+1}|}\right) = \gamma(y_i)$$

where γ is the *unsigned external angle*, i.e., the absolute value of the external angle, between the line segments of P meeting in y_i, see Figure 4 for an illustration. Consequently,

$$T_{\rho_1}(\mu) = \sum_{\substack{1 \leq i \leq I \\ y_i \in \Omega}} \gamma(y_i)$$

and in particular, for a convex polygon P whose vertices are entirely contained in Ω it follows that $T_{\rho_1}(\mu) = 2\pi$ since the external angle is non-negative in each vertex.

The last example gives the intuition that $T_{\rho_1}(\mu)$ measures the total curvature. Indeed, if we replace each polygon corner by smoothly connecting the line segments meeting in y_i via a circular arc A_i corresponding to a radius $r > 0$, then the length of that arc would be $\gamma(y_i)r$ and the absolute value of the curvature $|\kappa|$ of A_i would be $\frac{1}{r}$, hence

$$\int_{A_i} |\kappa| \, d\mathscr{H}^1 = \gamma(y_i).$$

In this sense, $T_{\rho_1}(\mu)$ measures the total curvature of the polygon P. If P is not a polygon, but a set whose boundary can be parametrized by a smooth curve, then this can also be made precise. Likewise the fact that T_{ρ_0} counts the vertices of a polygon can be generalized in the way that if P has a smooth curved boundary (which might be interpreted as the limit case of a polygon with infinitely many vertices) then $T_{\rho_0}(\mu) = \infty$

Proposition 3.16. *Let ρ_0 be the discrete and ρ_1 be the geodesic metric on S^1. Let $P \subset \Omega$ with piecewise $\mathscr{C}^2$ boundary $\partial P \subset \Omega$, i.e. ∂P contains a set $x_1, \ldots, x_I$ of vertices which are connected by $\mathscr{C}^2$ arcs. For simplicity, assume ∂P to be homeomorphic to S^1. Then, for μ being the lifted gradient of the characteristic function χ_P, it holds that*

$$T_{\rho_0}(\mu) = \begin{cases} \#\{x_i \,|\, \partial P \text{ is not } \mathscr{C}^1 \text{ at } x_i\} & \text{if } \kappa = 0 \text{ on} \\ & \partial P \setminus \{x_1, \ldots, x_I\}, \\ \infty & \text{else,} \end{cases}$$

$$T_{\rho_1}(\mu) = \int_{\partial P \setminus \{x_1, \ldots, x_I\}} |\kappa| \, d\mathscr{H}^1 + \sum_{1 \leq i \leq I} \gamma(x_i),$$

where κ is the curvature of the curve ∂P and $\gamma(x_i)$ its unsigned external angle at x_i.

Proof. We can parameterize ∂P by arclength via a continuous function $x : [0,L[\to \Omega$ which can be periodically extended to $\mathbb{R}$ and which is twice differentiable except at points $s_1, \ldots, s_I \in [0,L[$ with $s_1 < \ldots < s_I$ and $x(s_i) = x_i$. Without loss of generality we may assume $s_1 = 0$, and we introduce $s_{I+1} = L$ and $S = \{s_1, \ldots, s_I\}$. Also, we assume x to be oriented such that P locally lies left of it. The parametrization x satisfies $|x'(s)| = 1$ for all $s \in [0,L[\setminus S$, and the curvature can be expressed as $\kappa(x(s)) = x'(s)^\perp \cdot x''(s)$ for all $s \in [0,L[\setminus S$.

Note that $|\nabla \chi_P| = \mathscr{H}^1 \llcorner \partial P$ and that the density σ of $\nabla \chi_P$ with respect to $|\nabla \chi_P|$ satisfies $\sigma(x(s)) = x'(s)^\perp$ for almost every $s \in [0,L[\setminus S$. Therefore, the lifting μ corresponds to

$$\int_{\Omega \times S^1} \varphi \, \mathrm{d}\mu = \sum_{i=1}^{I} \int_{s_i}^{s_{i+1}} \varphi(x(s), x'(s)) \, \mathrm{d}s.$$

By the fundamental theorem of calculus we have

$$\psi(x(s_{i+1}), x'(s_{i+1})^-) - \psi(x(s_i), x'(s_i)^+)$$
$$= \int_{s_i}^{s_{i+1}} \frac{\partial}{\partial s} \psi(x(s), x'(s)) \, \mathrm{d}s$$
$$= \int_{s_i}^{s_{i+1}} \nabla_x \psi(x(s), x'(s)) \cdot x'(s) \, \mathrm{d}s$$
$$\quad + \int_{s_i}^{s_{i+1}} \frac{\partial}{\partial \vartheta} \psi(x(s), x'(s)) \cdot x''(s) \, \mathrm{d}s,$$

where $x'(s_i)^\pm$ denotes the derivative from the right and from the left, respectively.

Example 3.14 shows $T_{\rho_0}(\mu) = \#\{x_i \,|\, \partial P \text{ is not } \mathscr{C}^1 \text{ at } x_i\}$ if ∂P is polygonal, i. e. it has only straight line segments with $\kappa = 0$. Else there is some $s \in [0,L[\setminus S$ with $\kappa(x(s)) \neq 0$ ($\kappa(x(s)) > 0$ without loss of generality), and by continuity there is a non-singleton interval $[s_l, s_u] \subset [0,L[\setminus S$ on which $\kappa(x(s)) > \hat{\kappa}$ for some $\hat{\kappa} > 0$. Denote by $\hat{\rho} : S^1 \times S^1 \to]-\pi, \pi]$ the signed angle function, i.e.,

$$\hat{\rho}(\vartheta_1, \vartheta_2) = \begin{cases} \rho_1(\vartheta_1, \vartheta_2) & \text{if the geodesic from } \vartheta_1 \\ & \text{to } \vartheta_2 \text{ is counterclockwise,} \\ -\rho_1(\vartheta_1, \vartheta_2) & \text{else,} \end{cases}$$

where we agree always to choose the counterclockwise geodesic in the ambiguous case. It is easy to see that $\tilde{\psi}(x(s), \vartheta) = \hat{\rho}(x'(s), \vartheta)$ can be extended to a $\mathscr{C}^1$-function in a neighborhood U of $\{(x(s), x'(s)) \,|\, s \in [s_l, s_u]\}$. Choosing a suitable smooth cutoff function $\eta : \Omega \times S^1 \to [0,1]$ such that $\mathrm{supp}\,\eta \subset U$, $\eta(x(s), x'(s)) = 1$ and $\frac{\partial \eta}{\partial \vartheta}(x(s), x'(s)) = 0$ for $s \in [s_l, s_u]$, we can achieve that for each $\lambda \in \mathbb{R}$,

$$\psi_\lambda : \Omega \times S^1 \to \mathbb{R}, \qquad \psi_\lambda(x, \vartheta) = \tfrac{1}{2}\eta(x, \vartheta)\sin(\lambda \tilde{\psi}(x, \vartheta))$$

is in $\mathscr{C}_0^1(\Omega \times S^1)$. Hence, $\psi_\lambda \in M_{\rho_0}$ for each $\lambda \in \mathbb{R}$, and consequently,

$$T_{\rho_0}(\mu) \geq \sup_{\lambda \in \mathbb{R}} \sum_{i=1}^{I} \int_{s_i}^{s_{i+1}} \nabla_x \psi_\lambda(x(s), x'(s)) \cdot x'(s) \, \mathrm{d}s$$
$$= \sup_{\lambda \in \mathbb{R}} -\sum_{i=1}^{I} \int_{s_i}^{s_{i+1}} \frac{\partial}{\partial \vartheta} \psi_\lambda(x(s), x'(s)) \cdot x''(s) \, \mathrm{d}s$$
$$\geq \sup_{\lambda \in \mathbb{R}} \frac{\lambda}{2} \int_{s_l}^{s_u} x'(s)^\perp \cdot x''(s) \, \mathrm{d}s$$
$$\geq \sup_{\lambda \in \mathbb{R}} \frac{\lambda}{2}(s_u - s_l)\hat{\kappa} = \infty.$$

As for T_{ρ_1}, suppose that $\psi \in \mathscr{C}_0^1(\Omega \times S^1)$ with $\|\frac{\partial \psi}{\partial \vartheta}\|_\infty \leq 1$. For each $(x, \vartheta) \in \Omega \times S^1$, $\frac{\partial \psi}{\partial \vartheta}(x, \vartheta) = c(x, \vartheta)\vartheta^\perp$ for some $c \in \mathscr{C}_0(\Omega \times S^1)$ with $\|c\|_\infty \leq 1$, hence

$$\sum_{i=1}^{I} \int_{s_i}^{s_{i+1}} \nabla_x \psi(x(s), x'(s)) \cdot x'(s) \, \mathrm{d}s$$
$$= \sum_{i=1}^{I} \psi(x(s_{i+1}), x'(s_{i+1})^-) - \psi(x(s_i), x'(s_i)^+)$$
$$\quad - \int_{s_i}^{s_{i+1}} \frac{\partial}{\partial \vartheta} \psi(x(s), x'(s)) \cdot x''(s) \, \mathrm{d}s$$
$$= \sum_{i=1}^{I} \psi(x(s_i), x'(s_i)^-) - \psi(x(s_i), x'(s_i)^+)$$
$$\quad - \int_{s_i}^{s_{i+1}} c(x(s), x'(s))\kappa(x(s)) \, \mathrm{d}s,$$

where $x'(s_1)^-$ and $x'(s_I)^+$ have to be interpreted as $x'(s_I)^-$ and $x'(s_1)^+$, respectively. By Example 3.8, to obtain $T_{\rho_1}(\mu)$ it suffices to test with each of the above ψ so that

$$T_{\rho_1}(\mu) \leq \sum_{i=1}^{I} \rho_1(x'(s_i)^-, x'(s_i)^+) + \int_{s_i}^{s_{i+1}} |\kappa(x(s))| \, \mathrm{d}s$$
$$= \sum_{i=1}^{I} \gamma(x_i) + \int_{\partial P \setminus x(S)} |\kappa| \, \mathrm{d}\mathscr{H}^1.$$

To prove that the value on the right-hand side is attained, consider the function $\tilde{\psi} : \partial P \times S^1 \to \mathbb{R}$ defined by

$$\tilde{\psi}(x(s), \vartheta) = \begin{cases} -\operatorname{sgn}(\kappa(x(s)))\hat{\rho}(x'(s), \vartheta) \\ \quad \text{if } s \in [0,L[\setminus S \text{ and } \rho_1(x'(s), \vartheta) < \frac{\pi}{2}, \\ -\operatorname{sgn}(\kappa(x(s)))\hat{\rho}(\vartheta, -x'(s)) \\ \quad \text{if } s \in [0,L[\setminus S \text{ and } \rho_1(x'(s), \vartheta) \geq \frac{\pi}{2}, \\ -\rho_1(x'(s)^-, \vartheta) \quad \text{if } s \in S. \end{cases}$$

Obviously, $\tilde{\psi}(x(s), \cdot) \in C_{\rho_1}$ with

$$\frac{\partial \tilde{\psi}}{\partial \vartheta}(x(s), x'(s)) = -\operatorname{sgn}(\kappa(x(s)))x'(s)^\perp$$

and

$$\tilde{\psi}(x(s_i), x'(s_i)^-) - \tilde{\psi}(x(s_i), x'(s_i)^+) = \rho_1(x'(s_i)^-, x'(s_i)^+).$$

However, as $\text{sgn}(\kappa)$ is not necessarily continuous, $\tilde{\psi}$ is in general not smooth, so that it cannot be extended to a function in M_{ρ_1}. Nevertheless, $\tilde{\psi}$ can readily be approximated by functions $\psi^N \in M_{\rho_1}$ such that

$$\lim_{N\to\infty} \frac{\partial \psi^N}{\partial \vartheta}\big(x(s),x'(s)\big) = -\text{sgn}(\kappa(x(s)))x'(s)^\perp$$

for a.e. $s \in [0,L[\,\backslash\, S$,

$$\lim_{N\to\infty} \psi^N\big(x(s_i),x'(s_i)^-\big) - \psi^N\big(x(s_i),x'(s_i)^+\big)$$
$$= \rho_1\big(x'(s_i)^-,x'(s_i)^+\big), \qquad i = 1,\ldots,I.$$

Consequently,

$$T_{\rho_1}(\mu) \geq \lim_{N\to\infty} \sum_{i=1}^{I} \psi^N\big(x(s_i),x'(s_i)^-\big) - \psi^N\big(x(s_i),x'(s_i)^+\big)$$
$$+ \int_{s_i}^{s_{i+1}} \frac{\partial \psi^N}{\partial \vartheta}\big(x(s),x'(s)\big)\, \mathrm{d}s = \sum_{i=1}^{I} \gamma(x_i) + \int_{\partial P\backslash x(S)} |\kappa|\, \mathrm{d}\mathscr{H}^1$$

by uniform boundedness and pointwise a.e. convergence. $\qquad\square$

Remark 3.17. Let us finally mention some connection of T_{ρ_1} to the total cyclic variation from [37]. Identifying $S^1 \sim [0,2\pi[$, it can be defined for functions $v : \Omega \to S^1$ of special bounded variation on the domain $\Omega \subset \mathbb{R}^d$ by setting

$$\text{TV}_{S^1}(v) = \int_\Omega |\nabla v|\, \mathrm{d}x + \int_{S_v} \rho_1(v^+,v^-)\, \mathrm{d}\mathscr{H}^{d-1}$$

where S_v denotes the jump set of v and v^- and v^+ are its values on each side of the jump set. In the setting of Proposition 3.16, i.e., $P \subset \Omega$ with piecewise smooth boundary, μ the lifted gradient of χ_P and $x : [0,L[\to \Omega$ as in the proof, we see that for $v = x'$ we have $\text{TV}_{S^1}(v) = T_{\rho_1}(\mu)$. Note that TV_{S^1} has to be interpreted in dimension one in order to make sense. As the authors show in [37], in terms of the sublevel set relaxation $\chi_{\{\vartheta < v(s)\}}$ on $[0,L[\times [0,2\pi[$, it can also be written in a dual formulation:

$$\text{TV}_{S^1}(v) = \sup_{\substack{\varphi \in \mathscr{C}^\infty_\#([0,L[\times[0,2\pi[) \\ \|\varphi\|_\infty \leq 1,\ \int_0^{2\pi}\varphi\,\mathrm{d}\vartheta = 0}} \int_0^L \int_0^{v(s)} \frac{\partial \varphi}{\partial s}(s,\vartheta)\, \mathrm{d}\vartheta\, \mathrm{d}s$$

where $\mathscr{C}^\infty_\#([0,L[\times [0,2\pi[)$ denotes the set of arbitrarily differentiable periodic functions. Employing integration by parts with respect to ϑ and substituting $\psi(s,\vartheta) = \int_0^\vartheta \varphi(s,t)\, \mathrm{d}t$ as well as $\tilde{\mu} = -\partial_\vartheta \chi_{\{\vartheta < v(s)\}}$ the supremum can be rewritten to

$$\text{TV}_{S^1}(v) = \sup_{\substack{\psi \in \mathscr{C}^\infty_\#([0,L[\times[0,2\pi[) \\ \|\partial_\vartheta \psi\|_\infty \leq 1,\ \psi(\cdot,0)=0}} \int_0^L \int_0^{2\pi} \frac{\partial \psi}{\partial s}\, \mathrm{d}\tilde{\mu}(s,\vartheta).$$

Up to periodic boundary conditions and density, the set over which the supremum is taken coincides with $M_{\rho_1}(]0,L[)$

from Definition 3.10 (also see Example 3.8). In order to obtain the functional in [37] for general $\Omega \subset \mathbb{R}^d$, one has to choose the dual formulation

$$\text{TV}_{S^1}(v) = \sup_{\substack{\psi \in \mathscr{C}^\infty_c(\Omega \times [0,2\pi[,\mathbb{R}^d) \\ \|\partial_\vartheta \psi\|_\infty \leq 1}} \int_\Omega \int_0^{2\pi} \text{div}_x\, \psi\, \mathrm{d}\partial_\vartheta(\chi_{\{\vartheta < v(x)\}})(x,\vartheta).$$

If $d = 2$, one could be tempted with plugging in a v such that $-\partial_\vartheta(\chi_{\{\vartheta < v(x)\}}) = \mu$, the gradient lifting of χ_P. In this case, however, μ is too singular, leading to an infinite supremum. In fact, to make such a supremum coincide with T_{ρ_1} one has to take scalar test functions ψ and test against $\nabla_x \psi \cdot \vartheta$ as it is done in Definition 3.10.

4 Generalization and convex relaxation

So far we have considered the vertex penalization functional only for characteristic functions of polygons or sets with piecewise smooth boundary. We would now like to generalize it to a preferably large class of functions. Throughout the section, we assume that ρ is an admissible metric according to Assumption 3.1 and that Ω is a bounded Lipschitz domain.

The first observation is obvious: For binary $u \in \text{BV}(\Omega)$, the functional lifting μ according to Definition 2.3 is in $\mathscr{M}(\Omega \times S^1)$, therefore, $T_\rho(\mu)$ still makes sense. As we interpret T_ρ as a functional which penalizes object boundaries, it would be meaningful to have a functional which acts on the sublevel sets of a function $u \in \text{BV}(\Omega)$ analogous to, for example, (1.3). Denoting by μ_t the functional lifting of $\nabla(\chi_{\{u<t\}})$, this corresponds to, letting $\alpha,\beta > 0$,

$$R_\rho^{\alpha,\beta}(u) = \int_{\mathbb{R}} \alpha\|\mu_t\|_{\mathscr{M}} + \beta T_\rho(\mu_t)\, \mathrm{d}t \tag{4.1}$$

where $\|\mu_t\|_{\mathscr{M}}$ corresponds to the perimeter of the sublevel set $\{u < t\}$. Indeed, this makes sense for images of bounded variation, since almost every $\nabla(\chi_{\{u<t\}}) \in \mathscr{M}(\Omega,\mathbb{R}^2)$. However, similarly to the well-known curvature-dependent functionals like, for instance, Euler's elastica, this functional is hard to tract algorithmically, in particular, in terms of global minimization. This is due to non-convexity which is caused by two non-linear operations since $T_\rho(\mu)$ is convex in μ. These can be identified to be the operation which extracts the sublevel sets, i.e., $(u,t) \mapsto \chi_{\{u<t\}}$ as well as the functional lifting operation $u \mapsto \mu(\nabla u)$. Our goal is therefore to relax the functional R according to (4.1) such that it becomes convex, lower semi-continuous and admits a structure that is suitable for global optimization in terms of computational tractability.

4.1 Relaxing the sublevel set formulation

The first step of relaxation is to get rid of the sublevel set operation. For this purpose, observe that the coarea formula gives, for $u \in \mathrm{BV}(\Omega)$ and $\varphi \in \mathscr{C}_0(\Omega, \mathbb{R}^2)$,

$$\int_\Omega \varphi \cdot \sigma \, \mathrm{d}|\nabla u| = \int_\mathbb{R} \int_{\partial^*\{u<t\}} \varphi(x) \cdot v_t(x) \, \mathrm{d}\mathscr{H}^1(x) \, \mathrm{d}t$$

where, σ is again the density of ∇u with respect to $|\nabla u|$, ∂^* denotes the essential boundary and v_t is the generalized outer normal to $\{u<t\}$, see for instance [1]. This implies that $v : (x,t) \mapsto v_t(x)$ extended by 0 outside of $\partial^*\{u<t\}$ can be regarded as a $\mathscr{H}^1 \otimes \mathscr{L}^1$ measurable function on $\Omega \times \mathbb{R}$. Disintegration then gives, for $|\nabla u|$-almost every $x \in \Omega$, a probability measure $\mu_x \in \mathscr{M}(\mathbb{R})$ such that

$$\int_\mathbb{R} v(x,t) \, \mathrm{d}\mu_x(t) = \sigma(x).$$

Now, $|\sigma(x)| = 1$ also holds $|\nabla u|$-almost everywhere, and since $|v(x,t)|$ is either 1 or 0 it follows that

$$1 = |\sigma(x)| \le \int_\mathbb{R} |v(x,t)| \, \mathrm{d}\mu_x(t) = \int_\mathbb{R} |v(x,t)|^2 \, \mathrm{d}\mu_x(t) \le 1$$

which implies with Jensen's inequality and the fact that the squared Euclidean norm $|\cdot|^2$ is strictly convex that $t \mapsto v(x,t)$ has to be constant μ_x-almost everywhere. Consequently, $v(x,t) = \chi_{\mathrm{supp}\,\mu_x}(t)\sigma(x)$ holds $\mathscr{H}^1 \otimes \mathscr{L}^1$-almost everywhere on $\Omega \times \mathbb{R}$.

This implies for the functional lifting μ of ∇u by virtue of the coarea formula, the definitions of μ_t, i.e., the functional liftings of $\nabla(\chi_{\{u<t\}}) = v_t \mathscr{H}^1 \llcorner \partial^*\{u<t\}$:

$$\begin{aligned}
\int_{\Omega \times S^1} \varphi \, \mathrm{d}\mu &= \int_\Omega \varphi\big(x, -\sigma(x)^\perp\big) \, \mathrm{d}|\nabla u| \\
&= \int_\mathbb{R} \int_{\partial^*\{u<t\}} \varphi\big(x, -\sigma(x)^\perp\big) \, \mathrm{d}\mathscr{H}^1 \, \mathrm{d}t \\
&= \int_\mathbb{R} \int_{\partial^*\{u<t\}} \varphi\big(x, -v(x,t)^\perp\big) \, \mathrm{d}\mathscr{H}^1 \, \mathrm{d}t \\
&= \int_\mathbb{R} \int_{\Omega \times S^1} \varphi \, \mathrm{d}\mu_t \, \mathrm{d}t
\end{aligned}$$

for each $\varphi \in \mathscr{C}_0(\Omega \times S^1)$. Hence, the functional lifting operation respects, in a certain sense, the level sets of u. Consequently, in view of Definition 3.10 if we test μ with a $\psi \in M_\rho(\Omega)$, then

$$\begin{aligned}
\int_{\Omega \times S^1} &\nabla_x \psi(x,\vartheta) \cdot \vartheta \, \mathrm{d}\mu(x,\vartheta) \\
&= \int_\mathbb{R} \int_{\Omega \times S^1} \nabla_x \psi(x,\vartheta) \cdot \vartheta \, \mathrm{d}\mu_t(x,\vartheta) \, \mathrm{d}t \\
&\le \int_\mathbb{R} T_\rho(\mu_t) \, \mathrm{d}t.
\end{aligned}$$

Likewise,

$$\alpha\|\mu\|_\mathscr{M} = \int_\Omega \alpha \, \mathrm{d}\mu = \int_\mathbb{R} \int_{\Omega \times S^1} \alpha \, \mathrm{d}\mu_t \, \mathrm{d}t = \int_\mathbb{R} \alpha\|\mu_t\|_\mathscr{M} \, \mathrm{d}t.$$

Therefore, we have, for each $u \in \mathrm{BV}(\Omega)$,

$$\alpha\|\mu\|_\mathscr{M} + \beta T_\rho(\mu) \le R_\rho^{\alpha,\beta}(u). \tag{4.2}$$

In other words: Plugging in the functional lifting of ∇u for general $u \in \mathrm{BV}(\Omega)$ instead of the sublevel sets gives a relaxation of $R_\rho^{\alpha,\beta}$.

4.2 Relaxing the functional lifting

Next, we derive a suitable convex relaxation for the functional lifting operation. For this purpose, consider the graph of $u \mapsto \mu$, i.e.,

$$G_\nabla = \big\{(u,\mu) \in \mathrm{BV}(\Omega) \times \mathscr{M}(\Omega \times S^1) \, \big| \, \mu \text{ is the functional}$$
$$\text{lifting of } \nabla u\big\}.$$

The convex relaxation of this graph would be the closed convex hull of G_∇ in an appropriate topology. Clearly, this set exists in an abstract sense but is, however, not computationally accessible. We therefore seek a suitable convex set which contains G_∇. For this purpose, observe that if $u \in \mathrm{BV}(\Omega)$, then μ is always a positive measure, i.e., $\mu \ge 0$. Additionally, for each $\varphi \in \mathscr{C}_c^\infty(\Omega, \mathbb{R}^2)$ it holds, by definition of the gradient lifting μ as well as the distributional derivative,

$$\begin{aligned}
\int_\Omega u \operatorname{div} \varphi \, \mathrm{d}x &= -\int_\Omega \varphi \cdot \sigma \, \mathrm{d}|\nabla u| \\
&= -\int_{\Omega \times S^1} \varphi(x) \cdot \vartheta^\perp \, \mathrm{d}\mu(x,\vartheta)
\end{aligned}$$

and, consequently,

$$\int_\Omega u \operatorname{div} \varphi \, \mathrm{d}x + \int_{\Omega \times S^1} \varphi(x) \cdot \vartheta^\perp \, \mathrm{d}\mu(x,\vartheta) = 0.$$

These two properties are necessary for $(u,\mu) \in G_\nabla$, therefore the set

$$\begin{aligned}
M_\nabla = \Big\{ &(u,\mu) \in L^1(\Omega) \times \mathscr{M}(\Omega \times S^1) \, \Big| \, \mu \ge 0, \\
&\int_\Omega u \operatorname{div} \varphi \, \mathrm{d}x + \int_{\Omega \times S^1} \varphi(x) \cdot \vartheta^\perp \, \mathrm{d}\mu(x,\vartheta) = 0 \\
&\qquad \text{for all } \varphi \in \mathscr{C}_c^\infty(\Omega, \mathbb{R}^2)\Big\} \quad (4.3)
\end{aligned}$$

is a superset of G_∇. We take this as a relaxation although, of course, there may exist other choices which provide a tighter relaxation. Let us note some basic properties of M_∇.

Proposition 4.1. *The set M_∇ according to (4.3) is nonempty, convex and sequentially closed with respect to weak convergence in $L^1(\Omega)$ and weak* convergence in $\mathscr{M}(\Omega \times S^1)$. Morevoer, for each $(u,\mu) \in M_\nabla$ it follows that $u \in \mathrm{BV}(\Omega)$.*

Proof. Obviously, M_∇ is non-empty and convex. For a sequence $\{(u^n, \mu^n)\}$ in M_∇ for which $u^n \rightharpoonup u$ in $L^1(\Omega)$ and $\mu^n \overset{*}{\rightharpoonup} \mu$ in $\mathscr{M}(\Omega \times S^1)$ for $(u, \mu) \in L^1(\Omega) \times \mathscr{M}(\Omega \times S^1)$ there holds for each $\psi \in \mathscr{C}_0(\Omega \times S^1)$ with $\psi \geq 0$ that

$$\int_{\Omega \times S^1} \psi \, \mathrm{d}\mu = \lim_{n \to \infty} \int_{\Omega \times S^1} \psi \, \mathrm{d}\mu^n \geq 0$$

since each μ^n is positive. This in fact characterizes positive measures, hence $\mu \geq 0$. Analogously, as the chosen convergence implies convergence of the integrals, for each $\varphi \in \mathscr{C}_c^\infty(\Omega, \mathbb{R}^2)$

$$\int_\Omega u \operatorname{div} \varphi \, \mathrm{d}x + \int_{\Omega \times S^1} \varphi(x) \cdot \vartheta^\perp \, \mathrm{d}\mu(x, \vartheta)$$
$$= \lim_{n \to \infty} \int_\Omega u^n \operatorname{div} \varphi \, \mathrm{d}x + \int_{\Omega \times S^1} \varphi(x) \cdot \vartheta^\perp \, \mathrm{d}\mu^n(x, \vartheta) = 0.$$

Thus, M_∇ is closed in the claimed sense. Finally, note that for each $(u, \mu) \in M_\nabla$ and $\varphi \in \mathscr{C}_c^\infty(\Omega, \mathbb{R}^2)$ with $\|\varphi\|_\infty \leq 1$ it holds that

$$\int_\Omega u \operatorname{div} \varphi \, \mathrm{d}x \leq \int_{\Omega \times S^1} |\varphi(x) \cdot \vartheta^\perp| \, \mathrm{d}\mu(x, \vartheta) \leq \|\mu\|_\mathscr{M}$$

meaning that u is of bounded variation, i.e., $u \in \mathrm{BV}(\Omega)$. □

In view of Subsection 4.1, the idea for the relaxation of R is now to plug into T_ρ, for a $u \in L^1(\Omega)$, all corresponding μ such that $(u, \mu) \in M_\nabla$ and take the infimum. This gives:

Definition 4.2. Let $\alpha > 0$ and $\beta > 0$. For $u \in L^1(\Omega)$ define

$$\overline{R}_\rho^{\alpha, \beta}(u) = \inf_{\substack{\mu \in \mathscr{M}(\Omega \times S^1) \\ (u, \mu) \in M_\nabla}} \alpha \|\mu\|_\mathscr{M} + \beta T_\rho(\mu)$$

where we set the infimum of the empty set to ∞.

Remark 4.3. In [34, 35], Schoenemann et al. proposed relaxations of curvature-penalizing functionals on graphs which represent discretized images. They introduce a variable y^{l_1, l_2} for each oriented pair (l_1, l_2) of adjacent graph edges and express the functional in terms of these variables. Although this approach is inherently discrete, some connections to our continuous model may be established. On the one hand, a possible counterpart of $(y^{l_1, l_2})_{l_1, l_2}$ in our model would be an extension of the measure μ which also incorporates curvature information in addition to position and orientation of the image edges. The consistency between measure μ and image u is expressed as the constraint $(u, \mu) \in M_\nabla$ which parallels the so-called surface continuation constraint in [35]. Reducing the consistency to one single linear constraint represents one principal source of convex relaxation in both models, the other being the joint treatment of all sublevel sets as discussed above.

Definition 4.2 yields a functional with convenient properties.

Proposition 4.4. *The functional $\overline{R}_\rho^{\alpha, \beta} : L^1(\Omega) \to [0, \infty]$ is lower semi-continuous, convex and positively one-homogeneous, i.e., $\overline{R}_\rho^{\alpha, \beta}(\lambda u) = \lambda \overline{R}_\rho^{\alpha, \beta}(u)$ for $\lambda \geq 0$ and $u \in L^1(\Omega)$. It moreover obeys the estimate*

$$\alpha \mathrm{TV} \leq \overline{R}_\rho^{\alpha, \beta} \leq R_\rho^{\alpha, \beta}.$$

Proof. First note that $\alpha \|\mu\|_\mathscr{M} + \beta T_\rho(\mu) \geq 0$ for each $\mu \in \mathscr{M}(\Omega \times S^1)$ by Definition 3.10 and consequently, $\overline{R}_\rho^{\alpha, \beta}(u) \in [0, \infty]$ for each $u \in L^1(\Omega)$. The lower semi-continuity can be deduced as follows. For $\{u^n\}$ a converging sequence in $L^1(\Omega)$ with limit $u \in L^1(\Omega)$, choose, for each n, corresponding minimizing sequences $\{(\mu^n)^m\}$ in $\mathscr{M}(\Omega \times S^1)$, i.e., $(u^n, (\mu^n)^m) \in M_\nabla$ and

$$\lim_{m \to \infty} \alpha \|(\mu^n)^m\|_\mathscr{M} + \beta T_\rho((\mu^n)^m) = \overline{R}_\rho^{\alpha, \beta}(u^n).$$

We can assume without loss of generality that $\{\overline{R}_\rho^{\alpha, \beta}(u^n)\}$ is a sequence in $[0, \infty[$ which converges to a finite value. Consider a diagonal sequence of $\{(\mu^n)^m\}$, denoted by $\{\tilde{\mu}^n\}$, such that

$$\alpha \|\tilde{\mu}^n\|_\mathscr{M} + \beta T_\rho(\tilde{\mu}^n) \leq \overline{R}_\rho^{\alpha, \beta}(u^n) + \tfrac{1}{n} \qquad \text{for all} \quad n \geq 1.$$

This sequence is bounded as, for $n \geq 1$,

$$\|\tilde{\mu}^n\|_\mathscr{M} \leq \alpha^{-1} \left(\alpha \|\tilde{\mu}^n\|_\mathscr{M} + \beta T_\rho(\tilde{\mu}^n) \right)$$
$$\leq \alpha^{-1} \left(\sup_{n \geq 1} \overline{R}_\rho^{\alpha, \beta}(u^n) + 1 \right) < \infty.$$

Hence, there exists a subsequence, also denoted by $\{\tilde{\mu}^n\}$, and a $\mu \in \mathscr{M}(\Omega \times S^1)$ with $\tilde{\mu}^n \overset{*}{\rightharpoonup} \mu$. From the closedness property stated in Proposition 4.1 it follows that $(u, \mu) \in M_\nabla$. Consequently, as $\|\cdot\|_\mathscr{M}$ as well as T_ρ are sequentially weak* lower semi-continuous, see Proposition 3.12,

$$\alpha \|\mu\|_\mathscr{M} + \beta T_\rho(\mu) \leq \liminf_{n \to \infty} \alpha \|\tilde{\mu}^n\|_\mathscr{M} + \beta T_\rho(\tilde{\mu}^n)$$
$$= \lim_{n \to \infty} \overline{R}_\rho^{\alpha, \beta}(u^n)$$

which implies the desired lower semi-continuity.

For the convexity, assume that $u^1, u^2 \in L^1(\Omega)$ such that $\overline{R}_\rho^{\alpha, \beta}(u^1) < \infty$ as well as $\overline{R}_\rho^{\alpha, \beta}(u^2) < \infty$. Denote by $\{(\mu^1)^n\}$ and $\{(\mu^2)^n\}$ the corresponding minimizing sequences, i.e., $(u^i, (\mu^i)^n) \in M_\nabla$ and $\alpha \|(\mu^i)^n\|_\mathscr{M} + \beta T_\rho((\mu^i)^n) \to \overline{R}_\rho^{\alpha, \beta}(u^i)$ as $n \to \infty$ for $i = 1, 2$. Then, for an arbitrary $\lambda \in [0, 1]$ we have, by convexity of M_∇, that for each n

$$\left(\lambda u^1 + (1 - \lambda) u^2, \lambda (\mu^1)^n + (1 - \lambda)(\mu^2)^n \right) \in M_\nabla$$

and, consequently, with the convexity of T_ρ (which follows from the positive homogeneity and triangle inequality, see

Proposition 3.12),

$$
\begin{aligned}
\overline{R}_\rho^{\alpha,\beta}&\left(\lambda u^1+(1-\lambda)u^2\right)\\
&\leq \liminf_{n\to\infty}\alpha\left\|\lambda(\mu^1)^n+(1-\lambda)(\mu^2)^n\right\|_{\mathcal{M}}\\
&\quad+\beta T_\rho\left(\lambda(\mu^1)^n+(1-\lambda)(\mu^2)^n\right)\\
&\leq \lambda\left(\lim_{n\to\infty}\alpha\|(\mu^1)^n\|_\alpha+\beta T_\rho\left((\mu^1)^n\right)\right)\\
&\quad+(1-\lambda)\left(\lim_{n\to\infty}\alpha\|(\mu^2)^n\|_{\mathcal{M}}+\beta T_\rho\left((\mu^2)^n\right)\right)\\
&=\lambda\overline{R}_\rho^{\alpha,\beta}(u^1)+(1-\lambda)\overline{R}_\rho^{\alpha,\beta}(u^2).
\end{aligned}
$$

The positive one-homogeneity can be seen as follows. First observe that $\overline{R}_\rho^{\alpha,\beta}(0)=0$ since $(0,0)\in M_\nabla$ and $T_\rho(0)=0$. This shows in particular that $\overline{R}_\rho^{\alpha,\beta}$ is proper. For $\lambda>0$, $u\in L^1(\Omega)$ we see that $(\lambda u,\lambda\mu)\in M_\nabla$ if and only if $(u,\mu)\in M_\nabla$ and by positive homogeneity of $\|\cdot\|_{\mathcal{M}}$ as well as T_ρ (see Proposition 3.12) it follows

$$
\begin{aligned}
\overline{R}_\rho^{\alpha,\beta}(\lambda u)&=\inf_{\substack{\lambda\mu\in\mathcal{M}(\Omega\times S^1)\\(\lambda u,\lambda\mu)\in M_\nabla}}\alpha\|\lambda\mu\|_{\mathcal{M}}+\beta T_\rho(\lambda\mu)\\
&=\lambda\inf_{\substack{\mu\in\mathcal{M}(\Omega\times S^1)\\(u,\mu)\in M_\nabla}}\alpha\|\mu\|_{\mathcal{M}}+\beta T_\rho(\mu)=\lambda\overline{R}_\rho^{\alpha,\beta}(u).
\end{aligned}
$$

Finally, note that for each $u\in\mathrm{BV}(\Omega)$ we have, by definition of M_∇ and as $T_\rho\geq 0$, for $\mu\in\mathcal{M}(\Omega\times S^1)$ with $(u,\mu)\in M_\nabla$ and $\varphi\in\mathscr{C}_c^\infty(\Omega,\mathbb{R}^2)$ with $\|\varphi\|_\infty\leq 1$ that

$$
\begin{aligned}
\alpha\int_\Omega u\,\mathrm{div}\,\varphi\,\mathrm{d}x&\leq\alpha\int_{\Omega\times S^1}|\varphi(x)\cdot\vartheta^\perp|\,\mathrm{d}\mu\\
&\leq\alpha\|\mu\|_{\mathcal{M}}+\beta T_\rho(\mu).
\end{aligned}
$$

Taking the supremum over all the above φ and the infimum over all the above μ then yields $\alpha\,\mathrm{TV}(u)\leq\overline{R}_\rho^{\alpha,\beta}(u)$. If $u\notin\mathrm{BV}(\Omega)$, then there exists no μ such that $(u,\mu)\in M_\nabla$: Otherwise, it would follow, by Proposition 4.1, that $u\in\mathrm{BV}(\Omega)$, a contradiction. Therefore, $\alpha\,\mathrm{TV}(u)=\infty=\overline{R}_\rho^{\alpha,\beta}(u)$ as we agreed that the infimum over the empty set is infinity. At last, if $u\in\mathrm{BV}(\Omega)$, then the gradient lifting μ according to Definition 2.3 exists and $(u,\mu)\in M_\nabla$, hence, according to (4.2),

$$
\overline{R}_\rho^{\alpha,\beta}(u)\leq\alpha\|\mu\|_{\mathcal{M}}+\beta T_\rho(\mu)\leq R_\rho^{\alpha,\beta}(u).
$$

In case $u\notin\mathrm{BV}(\Omega)$, the coarea formula gives, for μ_t being the functional lifting of $\nabla(\chi_{\{u<t\}})$,

$$
\infty=\alpha\,\mathrm{TV}(u)=\int_\mathbb{R}\alpha\|\mu_t\|_{\mathcal{M}}\,\mathrm{d}t\leq R_\rho^{\alpha,\beta}(u)
$$

which concludes the proof. $\square$

Of course, the metrics ρ_0 and ρ_1 introduced in Examples 3.7 and 3.8, respectively, can be used for $\overline{R}_\rho^{\alpha,\beta}$. Depending on the choice, we call the resulting functionals $\mathrm{TVX}_0^{\alpha,\beta}$ and $\mathrm{TVX}_1^{\alpha,\beta}$, respectively. Let us take a closer look at those functionals in the following examples.

Example 4.5. First, consider $\overline{R}_{\rho_0}^{\alpha,\beta}$, the relaxed functional associated with the discrete metric ρ_0. Remembering the characterization of C_{ρ_0} from Example 3.7 and using mollification as well as smooth cut-off techniques, we see that

$$
\begin{aligned}
\tilde{M}_{\rho_0}(\Omega)=\{(x,\vartheta)\mapsto\psi(x,\vartheta)+\varphi(x)\mid\psi\in\mathscr{C}_c^\infty(\Omega\times S^1),\\
\|\psi\|_\infty\leq\tfrac{1}{2},\ \varphi\in\mathscr{C}_c^\infty(\Omega)\}
\end{aligned}
$$

is densely contained in M_{ρ_0} according to Definition 3.10. Introduce $\nabla_\vartheta\mu$ as the distributional directional derivative of μ with respect to $(\vartheta,0)$ (see also Remark 3.11) which is uniquely determined by

$$
\langle\nabla_\vartheta\mu,\,\psi\rangle=-\int_{\Omega\times S^1}\nabla_x\psi(x,\vartheta)\cdot\vartheta\,\mathrm{d}\mu
$$

for all $\psi\in\mathscr{C}_c^\infty(\Omega\times S^1)$. With this notion and using $\tilde{M}_{\rho_0}(\Omega)$ instead of $M_{\rho_0}(\Omega)$, we can express T_{ρ_0} as

$$
\begin{aligned}
T_{\rho_0}(\mu)=&\sup_{\substack{\psi\in\mathscr{C}_c^\infty(\Omega\times S^1)\\\|\psi\|_\infty\leq\frac{1}{2}}}\int_{\Omega\times S^1}\nabla_x\psi(x,\vartheta)\cdot\vartheta\,\mathrm{d}\mu(x,\vartheta)\\
&+\sup_{\varphi\in\mathscr{C}_c^\infty(\Omega)}\int_{\Omega\times S^1}\nabla_x\varphi(x)\cdot\vartheta\,\mathrm{d}\mu(x,\vartheta).
\end{aligned}
$$

The first supremum amounts to $\frac{1}{2}\|\nabla_\vartheta\mu\|_{\mathcal{M}}$ with infinity attained if $\nabla_\vartheta\mu$ is not in $\mathcal{M}(\Omega\times S^1)$. Regarding the second supremum, assume that $(u,\mu)\in M_\nabla$, so for each $\varphi\in\mathscr{C}_c^\infty(\Omega)$ we have

$$
\begin{aligned}
\int_{\Omega\times S^1}\nabla\varphi(x)\cdot\vartheta\,\mathrm{d}\mu(x,\vartheta)&=\int_{\Omega\times S^1}\nabla\varphi(x)^\perp\cdot\vartheta^\perp\,\mathrm{d}\mu(x,\vartheta)\\
&=-\int_\Omega u\,\mathrm{div}\left((\nabla\varphi)^\perp\right)\mathrm{d}x=0
\end{aligned}
$$

since $\mathrm{div}\left((\nabla\varphi)^\perp\right)=\mathrm{curl}\,\nabla\varphi=0$. Hence, the supremum is actually 0 and the relaxed functional $\mathrm{TVX}_0^{\alpha,\beta}=\overline{R}_{\rho_0}^{\alpha,\beta}$ may be described as follows:

$$
\mathrm{TVX}_0^{\alpha,\beta}(u)=\inf_{\substack{\mu\in\mathcal{M}(\Omega\times S^1)\\(u,\mu)\in M_\nabla}}\alpha\|\mu\|_{\mathcal{M}}+\tfrac{\beta}{2}\|\nabla_\vartheta\mu\|_{\mathcal{M}}. \qquad (4.4)
$$

Example 4.6. Let us now turn to $\overline{R}_{\rho_1}^{\alpha,\beta}$ with the geodesic metric ρ_1 according to Example 3.8. First, introduce for each $\psi\in\mathscr{C}^\infty(\Omega\times S^1)$ the projection

$$
(\pi_x\psi)(x)=\int_{S^1}\psi(x,\vartheta)\,\mathrm{d}\vartheta.
$$

In the case of ρ_1, the set

$$\tilde{M}_{\rho_1}(\Omega) = \{(x,\vartheta) \mapsto \psi(x,\vartheta) + \varphi(x) \mid \psi \in \mathscr{C}_c^\infty(\Omega \times S^1),$$
$$\pi_x \psi = 0, \ \|\partial_\vartheta \psi\|_\infty \leq 1, \ \varphi \in \mathscr{C}_c^\infty(\Omega)\}$$

can be identified to be sufficient to test with in Definition 3.10 in order to obtain T_{ρ_1}. Consequently, $\text{TVX}_1^{\alpha,\beta} = \overline{R}_{\rho_1}^{\alpha,\beta}$ may be written as

$$\text{TVX}_1^{\alpha,\beta}(u) = \inf_{\substack{\mu \in \mathscr{M}(\Omega \times S^1) \\ (u,\mu) \in M_\nabla}} \sup_{\substack{\psi \in \mathscr{C}_c^\infty(\Omega \times S^1) \\ \pi_x \psi = 0 \\ \|\partial_\vartheta \psi\|_\infty \leq 1}} \alpha\|\mu\|_{\mathscr{M}} + \beta\langle \nabla_\vartheta \mu, \psi\rangle$$

(4.5)

again since $\sup_{\varphi \in \mathscr{C}_c^\infty(\Omega)} \int_{\Omega \times S^1} \nabla_x \varphi(x) \cdot \vartheta \ \mathrm{d}\mu(x,\vartheta) = 0$ for $(u,\mu) \in M_\nabla$ (see Example 4.5). Strictly speaking, the constraint $\pi_x \psi = 0$ is unnecessary ($\tilde{M}_{\rho_1}(\Omega)$ does not change if this constraint is neglected); it only serves to remove the ambiguity associated with adding an offset to $\psi(x,\cdot)$ and thereby reduces the space of allowed test functions which might lead to better numerical stability. Informally, one can also use Fenchel-Rockafellar duality to turn the supremum into an infimum. Denoting by $\partial_\vartheta \nu$ the distributional partial derivative of a $\nu \in \mathscr{M}(\Omega \times S^1)$ with respect to the S^1-direction, this results in

$$\text{TVX}_1^{\alpha,\beta}(u) = \inf_{\substack{\mu \in \mathscr{M}(\Omega \times S^1) \\ (u,\mu) \in M_\nabla}} \inf_{\substack{\nu \in \mathscr{M}(\Omega \times S^1) \\ \partial_\vartheta \nu + \nabla_\vartheta \mu = 0}} \alpha\|\mu\|_{\mathscr{M}} + \beta\|\nu\|_{\mathscr{M}}.$$

However, this interpretation is currently just an informal observation.

4.3 Application to imaging problems

We conclude the section by showing basic existence and uniqueness results for imaging problems regularized with $\overline{R}_\rho^{\alpha,\beta}$ followed by examples from imaging.

Theorem 4.7. *Let $G : L^2(\Omega) \to \ ]-\infty, \infty]$ be bounded from below, convex, lower semi-continuous and such that*

$$G(u^n) \to \infty \quad \text{whenever} \quad \begin{cases} |\int_\Omega u^n \ \mathrm{d}x| \to \infty \ \text{and} \\ \{\|u^n - |\Omega|^{-1} \int_\Omega u^n \ \mathrm{d}x\|_2\} \\ \text{is bounded.} \end{cases}$$

Then, for each $\alpha > 0$ and $\beta > 0$ there exists a solution u^* of the variational problem

$$\min_{u \in L^2(\Omega)} G(u) + \overline{R}_\rho^{\alpha,\beta}(u). \tag{4.6}$$

In case that G is strictly convex, the solution is unique if the minimum is finite.

Proof. In order to prove the result, we use the direct method. If the objective functional $F = G + \overline{R}_\rho^{\alpha,\beta}$ is constant ∞, the statement is trivial. Hence, we assume in the following that the objective functional F is proper. From the assumptions on G as well as Proposition 4.1 it is immediate that it is also bounded from below, hence, there exists a minimizing sequence $\{u^n\}$ whose functional values $\{F(u^n)\}$ converge to the infimum of F which is finite. Consequently, as

$$\alpha \text{TV} \leq \overline{R}_\rho^{\alpha,\beta} \leq F - \inf_{u \in L^2(\Omega)} G(u)$$

according to Proposition 4.1, $\{\text{TV}(u^n)\}$ is bounded. Now, with the Poincaré-Friedrichs inequality for TV (see, for instance, [17]), it holds that, for a suitable $C > 0$,

$$\left\|u^n - \frac{1}{|\Omega|}\int_\Omega u^n \ \mathrm{d}x\right\|_2 \leq C \sup_{n \in \mathbb{N}} \text{TV}(u^n) < \infty$$

The assumption on G now tells us that, $\{\int_\Omega u^n \ \mathrm{d}x\}$ is bounded since otherwise, for a subsequence, $G(u^n) \to \infty$ which implies $F(u^n) \to \infty$, a contradiction. Consequently, we have

$$\|u^n\|_2 \leq \sup_{n \in \mathbb{N}} |\Omega|^{-1/2}\left|\int_\Omega u^n \ \mathrm{d}x\right| + \left\|u^n - \frac{1}{|\Omega|}\int_\Omega u^n \ \mathrm{d}x\right\|_2$$
$$< \infty$$

which implies that there exists a weakly convergent subsequence, still denoted by $\{u^n\}$ with limit $u^* \in L^2(\Omega)$. Since G and $\overline{R}_\rho^{\alpha,\beta}$ are convex and lower semi-continuous (by assumption and Proposition 4.1, respectively), we can conclude, as usual:

$$F(u^*) \leq \liminf_{n \to \infty} F(u^n) = \inf_{u \in L^2(\Omega)} F(u)$$

which implies that u^* is a minimizer. In case that G is strictly convex and that the minimum is finite, the uniqueness of the minimizer follows by virtue of the standard contradiction argument. $\qquad\square$

Example 4.8. The problem of binary image segmentation, i.e., partitioning the domain Ω in foreground and background, aims at minimizing, in the convex relaxed form, the functional

$$G(u) = \int_\Omega fu \ \mathrm{d}x + \iota_C(u),$$
$$C = \{u \in L^2(\Omega) \mid 0 \leq u(x) \leq 1 \text{ a.e. in } \Omega\}.$$

Here, ι_C denotes the indicator functional with respect to C, i.e., $\iota(u) = 0$ if $u \in C$ and $\iota(u) = \infty$ otherwise. Furthermore, $f \in L^1(\Omega)$ is an external segmentation field that is negative in points which are more likely to be background and positive in points which are more likely to be foreground. We would like to regularize this problem with $\overline{R}_\rho^{\alpha,\beta}$, i.e., solve

$$\min_{\substack{u \in L^2(\Omega) \\ 0 \leq u \leq 1 \text{ a.e.}}} \int_\Omega fu \ \mathrm{d}x + \overline{R}_\rho^{\alpha,\beta}(u). \tag{4.7}$$

In order to apply Theorem 4.7, verify that $G(u) \geq -\|f\|_1$ for all u such that $u(x) \in [0,1]$ a.e. and that G is convex. Moreover, for a sequence $\{u^n\}$ in $L^2(\Omega)$ with $|\int_\Omega u^n \, dx| \to \infty$ it holds that $\iota_C(u^n) = \infty$ for all but finitely many n since otherwise, $|\int_\Omega u^n \, dx| \leq |\Omega|$ for infinitely many n, a contradiction. Consequently, $G(u^n) \to \infty$ as $n \to \infty$. Finally, G is lower semi-continuous: For each converging sequence $\{u^n\}$ in $L^2(\Omega)$ with $u^n(x) \in [0,1]$ a.e. for all n and limit $u \in L^2(\Omega)$ we can find a pointwise a.e. converging subsequence (not relabeled). Thus the limit u has to satisfy $u(x) \in [0,1]$ a.e. and as $|fu^n| \leq |f|$ a.e., Lebesgue's dominated convergence theorem yields

$$\lim_{n\to\infty} \int_\Omega fu^n \, dx = \int_\Omega fu \, dx.$$

The subsequence was arbitrary, hence $G(u) \leq \liminf_{n\to\infty} G(u^n)$. This shows the applicability of Theorem 4.7, consequently, (4.7) has a solution.

Example 4.9. For the problem of denoising an image, we consider data fidelity functionals G which correspond to the usual L^p discrepancy: For $1 \leq p \leq 2$ let

$$G(u) = \frac{1}{p} \int_\Omega |u - f|^p \, dx.$$

Here, $f \in L^p(\Omega)$ denotes the noisy image. Obviously, G is non-negative, convex, continuous and satisfies $G(u^n) \to \infty$ if $|\int_\Omega u^n \, dx| \to \infty$ since

$$\|u^n - f\|_p^p \geq 2^{1-p} \|u^n\|_p^p - \|f\|_p^p$$
$$\geq (2|\Omega|)^{1-p} \left| \int_\Omega u^n \, dx \right|^p - \|f\|_P^p.$$

Hence, Theorem 4.7 is applicable meaning that the denoising problem

$$\min_{u \in L^2(\Omega)} \frac{1}{p} \int_\Omega |u - f|^p \, dx + \overline{R}_\rho^{\alpha,\beta}(u) \tag{4.8}$$

has a solution. In case $p > 1$, the solution is unique.

Example 4.10. Let us consider the inpainting problem for an incomplete image $f \in L^2(\Omega')$ given only on a non-null subset Ω' of Ω. Using the regularizer $\overline{R}_\rho^{\alpha,\beta}$, the goal is to minimize the functional

$$\min_{\substack{u \in L^2(\Omega) \\ u|_{\Omega'} = f}} \overline{R}_\rho^{\alpha,\beta}(u). \tag{4.9}$$

The corresponding data term $G : L^2(\Omega) \to]-\infty, \infty]$ reads as

$$G(u) = \begin{cases} 0 & \text{if } u|_{\Omega'} = f, \\ \infty & \text{else.} \end{cases}$$

It is easy to see that G is bounded from below and convex. The lower semi-continuity follows from the fact that the pointwise a.e. constraints $u(x) = f(x)$ on Ω' form

a closed subset of $L^2(\Omega)$. Now let $\{u^n\}$ be given such that $|\int_\Omega u^n \, dx| \to \infty$ and that $\{\|u^n - |\Omega|^{-1} \int_\Omega u^n \, dx\|_2\}$ is bounded. Suppose that $u^n|_{\Omega'} = f$ for infinitely many n, without loss of generality, we may assume that this is the case for the whole sequence $\{u^n\}$. Denote by $v^n = u^n - |\Omega|^{-1} \int_\Omega u^n \, dx$ and observe that, as $u^n|_{\Omega'} = f$,

$$\sup_{n\in\mathbb{N}} \left| \int_\Omega u^n \, dx \right| = \sup_{n\in\mathbb{N}} \left| \frac{|\Omega|}{|\Omega'|} \int_{\Omega'} (f - v^n) \, dx \right|$$
$$\leq C \sup_{n\in\mathbb{N}} \left(\|f\|_2 + \|v^n\|_2 \right) < \infty$$

which is a contradiction. Hence, $u^n|_{\Omega'} = f$ for only finitely many n and, consequently, $G(u^n) \to \infty$. By Theorem 4.7, we have existence of a minimizer. However, the minimum will only be finite if the data f is regular enough, i.e., if there is a $u \in L^2(\Omega)$ such that $u|_{\Omega'} = f$ and $\overline{R}_\rho^{\alpha,\beta}(u) < \infty$.

5 Numerical Results

In this section we give a finite differences discretization of the imaging problems (4.6) and we show how to efficiently minimize the resulting saddle-point problems. In particular, we will present applications to the imaging problems introduced in Section 4: image segmentation, image restoration and inpainting.

5.1 Discrete Setting

An image u will be discretized using a two-dimensional regular Cartesian grid of size $M \times N$:

$$\{(ih, jh) \mid 1 \leq i \leq M, 1 \leq j \leq N\},$$

where M and N denote the height and width of the image, h denotes the size of the spacing and (i, j) denote the indices of the discrete locations (ih, jh) in the image domain. Likewise, the discretized version of the lifted quantity μ will be discretized on a three-dimensional Cartesian grid of size $M \times N \times K$:

$$\{(ih, jh, kg) \mid 1 \leq i \leq M, 1 \leq j \leq N, 1 \leq k \leq K\},$$

where K denotes the number of discrete points on the unit circle S^1 and g denotes the spacing.

Next we describe the finite differences approximations of the different linear operators we will need to define the discrete versions of the proposed energy functionals. The linear operator $A \in \mathbb{R}^{2MN \times MN}$ is a discretized version of the derivative operator. We use standard forward differences with Neumann boundary conditions

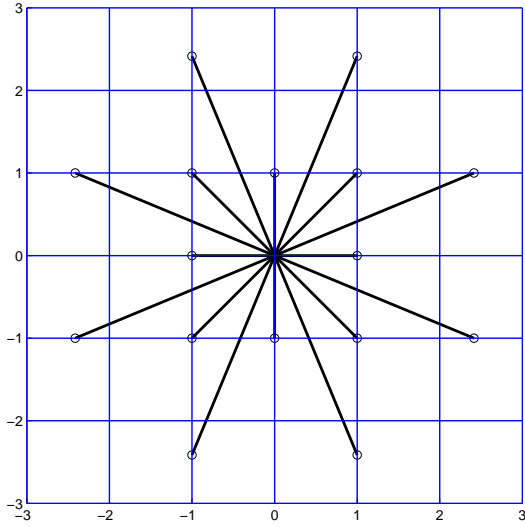$$(Au)_{i,j} = \begin{pmatrix} (Au)_{i,j}^1 \\ (Au)_{i,j}^2 \end{pmatrix},$$

Fig. 5 Finite differences scheme to discretize the directional derivatives. We use a non-local finite differences scheme to better capture boundary directions that cannot be well represented by a regular grid.

where

$$(Au)^1_{i,j} = \begin{cases} \dfrac{u_{i+1,j} - u_{i,j}}{h} & \text{if } i < M \\ 0 & \text{if } i = M \end{cases},$$

$$(Au)^2_{i,j} = \begin{cases} \dfrac{u_{i,j+1} - u_{i,j}}{h} & \text{if } j < N \\ 0 & \text{if } j = N \end{cases}.$$

The discretization of the directional derivative operators $\nabla_{\vartheta^\perp}$ is critical since a standard finite differences scheme for non grid-aligned directions might not exactly hit the grid points. We therefore use a non-local finite differences scheme in combination with a linear interpolation to eliminate interfering dissipation effects as much as possible. Let $\vartheta_k = (\vartheta^1_k, \vartheta^2_k) \in S^1$, $k = 1, \dots, K$ be uniformly distributed points on the unit circle. We compute the displacement vector for the finite differences scheme of the directional derivatives from ϑ_k by rescaling it such that its smaller component is equal to one. The respective rescaling constants are computed via

$$\delta_k = \begin{cases} \min\{|\vartheta^1_k|, |\vartheta^2_k|\} & \text{if } \vartheta_k \bmod \pi/2 \neq 0 \\ 1 & \text{else} \end{cases}.$$

The final displacement vectors $t_k = (t^1_k, t^2_k)$ are computed as $t_k = \vartheta_k / \delta_k$. See Figure 5 for an example of the displacement vectors t_k for $K = 16$ directions. The finite differences approximation of the directional derivative operator is a linear operator $B \in \mathbb{R}^{MNK \times MNK}$ which is defined as

$$(B\mu)_{i,j,k} = \frac{\mu_{i+t^1_k, j+t^2_k, k} - \mu_{i,j,k}}{|t_k|},$$

where $\mu_{i+t^1_k, j+t^2_k}$ refers to the result of a linear interpolation with zero boundary conditions.

We will also need to discretize the derivative operator in label direction. It is defined as the linear operator $C \in \mathbb{R}^{MNK \times MNK}$. We again use simple forward differences with periodic boundary conditions

$$(C\mu)_{i,j,k} = \begin{cases} \dfrac{\mu_{i,j,k+1} - \mu_{i,j,k}}{g} & \text{if } k < K \\ \dfrac{\mu_{i,j,1} - \mu_{i,j,k}}{g} & \text{if } k = K \end{cases}.$$

Finally we will need to discretize the linear constraint in the set M_∇ (see (4.3)), that relates the discrete image gradient Au to the lifted quantity μ. The issue here is that the image gradient is computed at half grid points while μ is defined on grid points. We found it best to relate the image gradient with μ linearly interpolated at the corresponding half points. We therefore define a linear operator $D \in \mathbb{R}^{2MN \times MNK}$ by

$$(D\mu)_{i,j} = \begin{pmatrix} (D\mu)^1_{i,j} \\ (D\mu)^2_{i,j} \end{pmatrix},$$

where

$$(D\mu)^1_{i,j,k} = \begin{cases} \dfrac{g}{2} \sum_{k=1}^{K} (-\vartheta^\perp_k)^1 (\mu_{i+1,j,k} + \mu_{i,j,k}) & \text{if } i < M \\ \dfrac{g}{2} \sum_{k=1}^{K} (-\vartheta^\perp_k)^1 \mu_{i,j,k} & \text{if } i = M \end{cases},$$

and

$$(D\mu)^2_{i,j,k} = \begin{cases} \dfrac{g}{2} \sum_{k=1}^{K} (-\vartheta^\perp_k)^2 (\mu_{i,j+1,k} + \mu_{i,j,k}) & \text{if } j < N \\ \dfrac{g}{2} \sum_{k=1}^{K} (-\vartheta^\perp_k)^2 \mu_{i,j,k} & \text{if } j = N \end{cases}.$$

5.2 Discrete energies

Having defined the discrete operators we are now ready to define the discrete versions of (4.6). For simplicity we will assume $h = g = 1$ for the rest of this section. We denote by $u \in \mathbb{R}^{MN}$ the discrete image, by $\mu \in \mathbb{R}^{MNK}$ the lifted gradient and by $\psi \in \mathbb{R}^{MNK}$ the dual variable. Now, combining the general model (4.6) with the functionals of Definition 3.10 and Definition 4.2 we arrive at the convex-concave saddle point problem

$$\min_{(u,\mu) \in M_\nabla} \max_{\psi \in M_\rho} G(u) + \alpha \sum_{i,j,k} \mu_{i,j,k} + \beta \langle B\mu, \psi \rangle \qquad (5.1)$$

where according to (4.3) the discrete variant of the convex set M_∇ is defined as

$$M_\nabla = \left\{ (\mu, u) \,\middle|\, \mu_{i,j,k} \geq 0, \quad (Au)_{i,j} = (D\mu)_{i,j} \right\}.$$

According to Example 4.5 and Example 4.6 and the discrete variants of the convex sets M_ρ are defined as

$$\text{TVX}_0^{\alpha,\beta} : M_{\rho_0} = \left\{ \psi \,\middle|\, |\psi_{i,j,k}| \leq \frac{1}{2}, \, \forall i,j,k \right\} \tag{5.2}$$

$$\text{TVX}_1^{\alpha,\beta} : M_{\rho_1} = \left\{ \psi \,\middle|\, |(C\psi)_{i,j,k}| \leq 1, \, \forall i,j,k, \right.$$
$$\left. \sum_k \psi_{i,j,k} = 0, \, \forall i,j \right\}. \tag{5.3}$$

5.3 Numerical optimization

We use the first-order primal-dual algorithm proposed in [9] which can be used to minimize convex problems with known saddle-point structure of the the general form

$$\min_{x \in X} \max_{y \in Y} \langle \mathscr{K}x, y \rangle + \mathscr{G}(x) - \mathscr{F}^*(y), \tag{5.4}$$

where X and Y are finite-dimensional vector spaces $\mathscr{K}$ is a linear operator and $\mathscr{G}$ and $\mathscr{F}^*$ are proper, convex and lower semi-continuous functions. See [9] for more information. The most important restriction of the algorithm is that $\mathscr{G}$ and $\mathscr{F}^*$ need to be of simple structure, in the sense that their proximity operators have to be easy to compute. We also make use of the diagonal preconditioning technique [30] that has been shown to perform better on problems with complicated $\mathscr{K}$ and avoids computing $\|\mathscr{K}\|$. Following [30], we assume that $\mathscr{K} \in \mathbb{R}^{m \times n}$ and compute the diagonal preconditioning matrices $\mathscr{T} \in \mathbb{R}^{n \times n}$ and $\mathscr{S} \in \mathbb{R}^{m \times m}$ according to the rules

$$\mathscr{T}_{b,b} = \frac{1}{\sum_{a=1}^m |\mathscr{K}_{a,b}|}, \quad \mathscr{S}_{a,a} = \frac{1}{\sum_{b=1}^n |\mathscr{K}_{a,b}|}. \tag{5.5}$$

In recent work [23], it has been shown that the primal-dual algorithm [9] can be written in form of a proximal point algorithm [32], which can be generalized by means of an additional overrelaxation step [21,15], which is controlled by an overrelaxation parameter γ. Very recently it has been shown in [6] that indeed, the overrelaxation step speeds up the convergence of the algorithm.

The basic iterations of the algorithm are as follows: Set $x = 0$, $y = 0$, choose $\mathscr{T}, \mathscr{S}$ according to (5.5) and $\gamma \in [0, 1[$. For $l \geq 0$ let

$$\begin{cases} x^{l+\frac{1}{2}} = \text{prox}_{\mathscr{T},\mathscr{G}}\left(x^l - \mathscr{T}\mathscr{K}^T y^l\right) \\ y^{l+\frac{1}{2}} = \text{prox}_{\mathscr{S},\mathscr{F}^*}\left(y^l + \mathscr{S}\mathscr{K}(2x^{l+\frac{1}{2}} - x^l)\right) \\ (x^{l+1}, y^{l+1}) = (x^{l+\frac{1}{2}}, y^{l+\frac{1}{2}}) + \gamma(x^{l+\frac{1}{2}} - x^l, y^{l+\frac{1}{2}} - y^l). \end{cases} \tag{5.6}$$

According to [30], the proximal mappings including the preconditioning are defined as

$$\text{prox}_{\mathscr{T},\mathscr{G}}(\hat{x}) = \arg\min_x \frac{1}{2}\left\langle \mathscr{T}^{-1}(x-\hat{x}), x-\hat{x} \right\rangle + \mathscr{G}(x),$$

$$\text{prox}_{\mathscr{S},\mathscr{F}^*}(\hat{y}) = \arg\min_y \frac{1}{2}\left\langle \mathscr{S}^{-1}(y-\hat{y}), y-\hat{y} \right\rangle + \mathscr{F}^*(y).$$

For more details see [30]. As shown in [9], the algorithm converges with rate $\mathscr{O}(1/l)$ for the averages $\bar{x} = (\sum_l x^l)/l$ and $\bar{y} = (\sum_l y^l)/l$ but we found it more efficient to use the final iterates x^l and y^l instead. The overrelaxation parameter is set to $\gamma = 0.9$ in all experiments. We stop the iteration as soon as the quantities $\|x^l - x^{l-1}\|$ and $\|y^l - y^{l-1}\|$ are below a certain threshold.

In order to make the algorithm (5.6) applicable to the constrained problem (5.1) we have to transform it into the generic form (5.4). This is shown in the next sections.

5.3.1 The functional $\text{TVX}_0^{\alpha,\beta}$

In this section we specialize to the functional that utilizes the discrete metric ρ_0. Observe that the set M_{ρ_0} is of simple structure, meaning that we can efficiently project on this set. However, the set M_∇ is more complicated, so that we have to perform an operator splitting. We introduce Lagrange multipliers $\phi \in \mathbb{R}^{MN}$ to account for the linear constraints $Au = D\mu$, which gives the saddle-point problem

$$\min_{u,\mu} \max_{\psi,\phi} \alpha \sum_{i,j,k} \mu_{i,j,k} + \beta \langle B\mu, \psi \rangle + \langle Au - D\mu, \phi \rangle + G(u)$$

$$\text{s.t.} \quad \mu_{i,j,k} \geq 0, \, |\psi_{i,j,k}| \leq \frac{1}{2} \tag{5.7}$$

Now, we set $x = (\mu, u)^T$ and $y = (\psi, \phi)$ and define the linear operator $\mathscr{K} \in \mathbb{R}^{m \times n}$, where $m = MNK + 2MN$ and $n = MNK + MN$ as

$$\mathscr{K} = \begin{pmatrix} \beta B & 0 \\ -D & A \end{pmatrix}.$$

Furthermore we define the functions $\mathscr{G}$ and $\mathscr{F}^*$ on the extended vectors x and y, respectively, as

$$\mathscr{G}(x) = \alpha \sum_{i,j,k} \mu_{i,j,k} + \iota_{[0,\infty)^{MNK}}(\mu) + G(u), \tag{5.8}$$

$$\mathscr{F}^*(y) = \iota_{[-\frac{1}{2},\frac{1}{2}]^{MNK}}(\psi), \tag{5.9}$$

where ι_S is again the indicator functional associated with the set S. The proximity operators with respect to $\mathscr{G}$ and $\mathscr{F}^*$ are identified as

$$x = \text{prox}_{\mathscr{T},\mathscr{G}}(\hat{x}) \tag{5.10}$$
$$\Longleftrightarrow$$
$$\mu_{i,j,k} = \max(0, \hat{\mu}_{i,j,k} - \alpha \mathscr{T}_{i,j,k}^\mu)$$
$$u = \text{prox}_{\mathscr{T}^u, G},$$

and

$$y = \text{prox}_{\mathscr{S},\mathscr{F}^*}(\hat{y}) \tag{5.11}$$
$$\Longleftrightarrow$$
$$\psi_{i,j,k} = \max(-\frac{1}{2}, \min(\frac{1}{2}, \hat{\psi}_{i,j,k}))$$
$$\phi_{i,j,k} = \hat{\phi}_{i,j,k},$$

where $\mathscr{T}$ and $\mathscr{S}$ are the preconditioning matrices and by a slight abuse of notation we denote for example by $\mathscr{T}^\mu$

the submatrix of $\mathscr{T}$ that corresponds to the vector μ and by $\mathscr{T}^\mu_{i,j,k}$ the entry of $\mathscr{T}^\mu$ that corresponds to the element $\mu_{i,j,k}$. With the above transformations we can now cast (5.7) into the generic form (5.4).

5.3.2 The functional $\mathrm{TVX}_1^{\alpha,\beta}$

Let us now turn to the geodesic metric ρ_1. The main difference to the previous model is that the convex set M_{ρ_1} in (5.3) includes an additional linear operator and hence the projection onto this set is not simple. Besides the Lagrange multipliers ϕ that account for the linear constraint $Au = D\mu$, we additionally introduce an auxiliary vector $\zeta \in \mathbb{R}^{MNK}$ and attach another set of Lagrange multipliers $v \in \mathbb{R}^{MNK}$ to account for the linear constraint $C\psi = \zeta$. Note that the constraint $\sum_k \psi_{i,j,k} = 0$ in M_{ρ_1} does not require any additional treatment since a projection onto this set can be performed efficiently by subtracting its average. We therefore arrive at the following saddle-point problem

$$\min_{u,\mu,v} \max_{\psi,\phi,\zeta} \alpha \sum_{i,j,k} \mu_{i,j,k} + \beta \langle B\mu, \psi \rangle + \langle Au - D\mu, \phi \rangle$$
$$+ \langle C\psi - \zeta, v \rangle + G(u)$$
$$\text{s.t.} \quad \mu_{i,j,k} \geq 0, \ |\zeta_{i,j,k}| \leq 1, \ \sum_k \psi_{i,j,k} = 0. \qquad (5.12)$$

Similarly to above we set $x = (\mu, v, u)^T$ and $y = (\psi, \zeta, \phi)$ and define the matrix $\mathscr{K} \in \mathbb{R}^{m \times n}$, where $m = 2MNK + 2MN$ and $n = 2MNK + MN$ as

$$\mathscr{K} = \begin{pmatrix} \beta B & C^T & 0 \\ 0 & -I & 0 \\ -D & 0 & A \end{pmatrix},$$

where $I \in \mathbb{R}^{MNK \times MNK}$ denotes the diagonal matrix with all diagonal entries equal to 1.

Likewise to the previous model we can easily identify the functions $\mathscr{G}$ and F^* in the generic form (5.4) as

$$\mathscr{G}(x) = \alpha \sum_{i,j,k} \mu_{i,j,k} + \iota_{[0,\infty)^{MNK}}(\mu) + G(u), \qquad (5.13)$$

$$\mathscr{F}^*(y) = \iota_Z(\psi) + \iota_{[-1,1]^{MNK}}(\zeta), \qquad (5.14)$$

where $Z = \{\psi : \sum_k \psi_{i,j,k} = 0, \ \forall i,j\}$ and the corresponding proximity operators are given by

$$x = \mathrm{prox}_{\mathscr{T},\mathscr{G}}(\hat{x}) \qquad (5.15)$$
$$\Longleftrightarrow$$
$$\mu_{i,j,k} = \max(0, \hat{\mu}_{i,j,k} - \alpha \mathscr{T}^\mu_{i,j,k})$$
$$v_{i,j,k} = \hat{v}_{i,j,k}$$
$$\hat{u} = \mathrm{prox}_{\mathscr{T}^u,G}(\hat{u}),$$

and

$$y = \mathrm{prox}_{\mathscr{S},\mathscr{F}^*}(\hat{y}) \qquad (5.16)$$
$$\Longleftrightarrow$$
$$\psi_{i,j,k} = \hat{\psi}_{i,j,k} - \left(\sum_k \hat{\psi}_{i,j,k}\right)/K$$
$$\zeta_{i,j,k} = \max(-1, \min(1, \hat{\zeta}_{i,j,k}))$$
$$\phi_{i,j,k} = \hat{\phi}_{i,j,k},$$

such that (5.12) can be cast into the generic form (5.4).

5.4 Application to imaging problems

In this section we show preliminary results of the finite differences implementation of the general imaging model (4.6). For simplicity, we will set $\alpha = 0.1$ and $\beta = 1$ for the regularizer $\mathrm{TVX}_{0,1}^{\alpha,\beta}$ in all experiments and hence we will skip the superscripts α and β for notational convenience. A study of the effects of varying the weights α and β is left for future work. For comparison we will also implement pure total variation regularization (TV), which can be easily realized by setting $\alpha = 1$ and $\beta = 0$. Finally, unless mentioned differently, we will set the discretization accuracy of the unit circle to $K = 32$ equally distributed points.

5.4.1 Binary image segmentation

In our first application we apply the proposed functional as a regularizer for binary image segmentation problems. As shown in Example 4.8, we use a simple linear data term of the form

$$G(u) = \lambda \sum_{i,j} f_{i,j} u_{i,j} + \iota_{[0,1]^{MN}}(u),$$

where $f \in \mathbb{R}^{MN}$ is an external segmentation field that is negative if a pixel is more likely to be background and positive if a pixel is more likely to be foreground. The parameter $\lambda > 0$ is used to control the weight of the data term with respect to the regularization term. The proximity operator with respect to G is given by

$$u = \mathrm{prox}_{\mathscr{T}^u,G}(\hat{u}) \Longleftrightarrow u_{i,j} = \max(0, \min(1, \hat{u}_{i,j} - \lambda \mathscr{T}^u_{i,j} f_{i,j})).$$

In the first segmentation experiment we study the effect of varying the number of discrete orientations K. Figure 6 shows the result of applying the proposed regularizers to an image segmentation problem. The segmentation field is computed as $f_{i,j} = (I_{i,j} - \mu_f)^2 - (I_{i,j} - \mu_b)^2$, where $I \in \mathbb{R}^{MN}$ is the input image and $\mu_{f,b}$ are the mean values of the fore- and background regions. We set $\mu_f = 0.0$, $\mu_b = 0.5$ and $\lambda = 1$. As expected, for a low number of orientations e.g. $K = 4$ and 8, the segmentation results favor certain directions. For a larger number of orientations e.g.
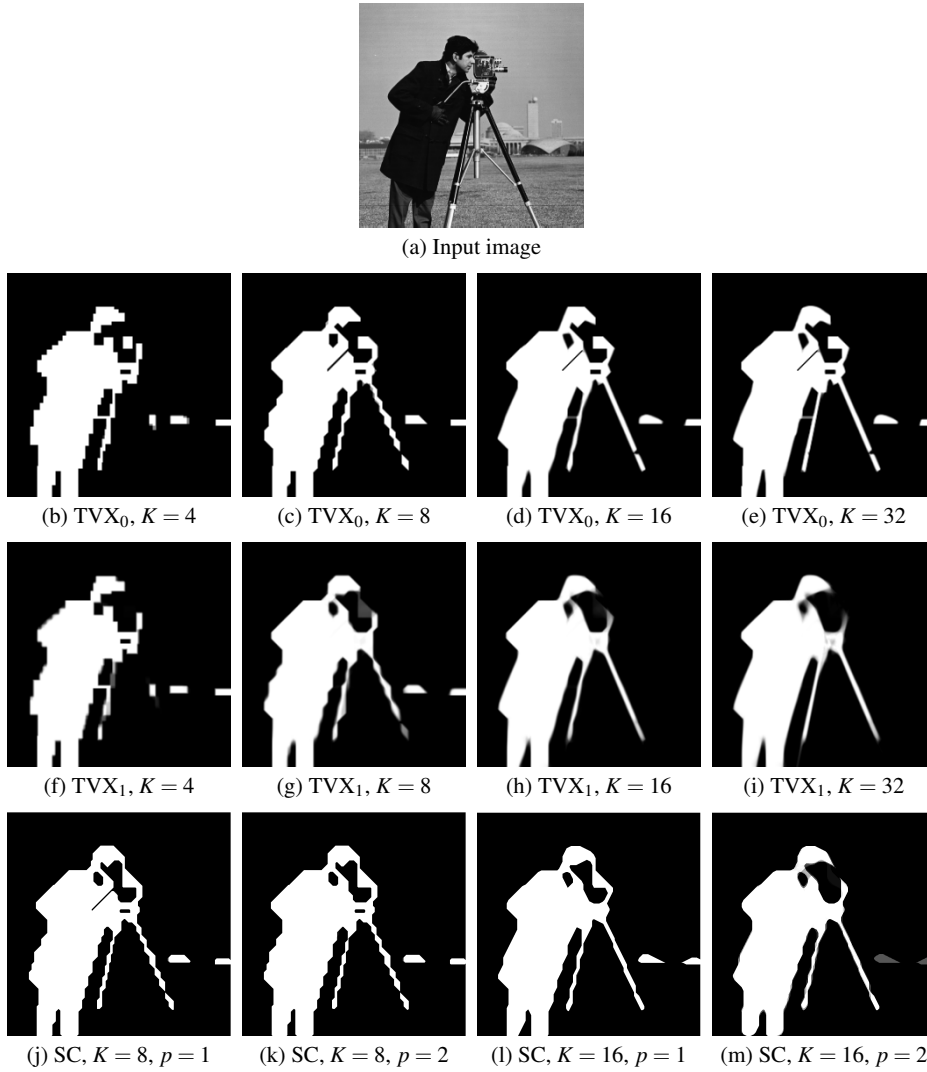
(a) Input image

(b) TVX$_0$, $K = 4$ | (c) TVX$_0$, $K = 8$ | (d) TVX$_0$, $K = 16$ | (e) TVX$_0$, $K = 32$

(f) TVX$_1$, $K = 4$ | (g) TVX$_1$, $K = 8$ | (h) TVX$_1$, $K = 16$ | (i) TVX$_1$, $K = 32$

(j) SC, $K = 8$, $p = 1$ | (k) SC, $K = 8$, $p = 2$ | (l) SC, $K = 16$, $p = 1$ | (m) SC, $K = 16$, $p = 2$

Fig. 6 Effect of varying the number of discrete orientations K. (a) shows the input image and (b)-(i) show the segmented images. (j)-(m) show results of the discrete approach [34] (SC) for different numbers of discrete orientations and different exponents p in the elastica functional.

$K = 16$ and $K = 32$, the model behaves almost isotropically. Note that for $K = 4$ the behavior of TVX$_0$ and TVX$_1$ is almost identical. One can also observe that the vertex counting functional leads to a strong preference of a polygonal shape, while the total curvature functional leads to smoother shapes. For comparison, we provide in the last row of Figure 6 some results of the discrete approach of Schoenemann and Cremers [34]. We used the primal dual algorithm [30] as a solver and implemented the algorithmic scheme on a graphics processing unit (GPU). We again used $\alpha = 0.1$ for the weight of the length term and $\beta = 1$ for the weight of the curvature term. In contrast to our TVX functionals, the approach of [34] can handle more general curvature depending functionals, and hence we ran the algorithm for different exponents p in the elastica functional. Running times for $K = 16$ discrete orientations where around 40 minutes on a Nvidia GTX 480 GPU. For comparison, a Matlab implementation of our approach takes around 5 minutes for problems with the same number of discrete orientations. Vi-

sually, the results of both approaches are comparable, although our approach leads to less strong orientation artifacts (see, for example, the results for $K = 16$). Furthermore we observed that Schoenemann's and Cremers's approach leads to strong edge cancellation artifacts for $K = 16$ orientations. This can, for instance, be seen from the spiky corners at the head of the cameraman which are, in fact, connected by overlapping edges. Due to memory restrictions, we were not able to perform experiments for $K = 32$ discrete orientations.

In the second experiment, we investigate the difference between the total vertex regularization models TVX$_{0,1}$ and standard total variation based regularization TV. Figure 7 show the differences between the different model by varying the parameter λ. The segmentation force was computed as $f_{i,j} = 0.5 - I_{i,j}$. Pure total variation regularization (TV) leads to a strong shrinkage of the objects since it penalizes the total length of the object boundary. TVX$_0$-based regularization penalizes the number of vertices of the segmenta-
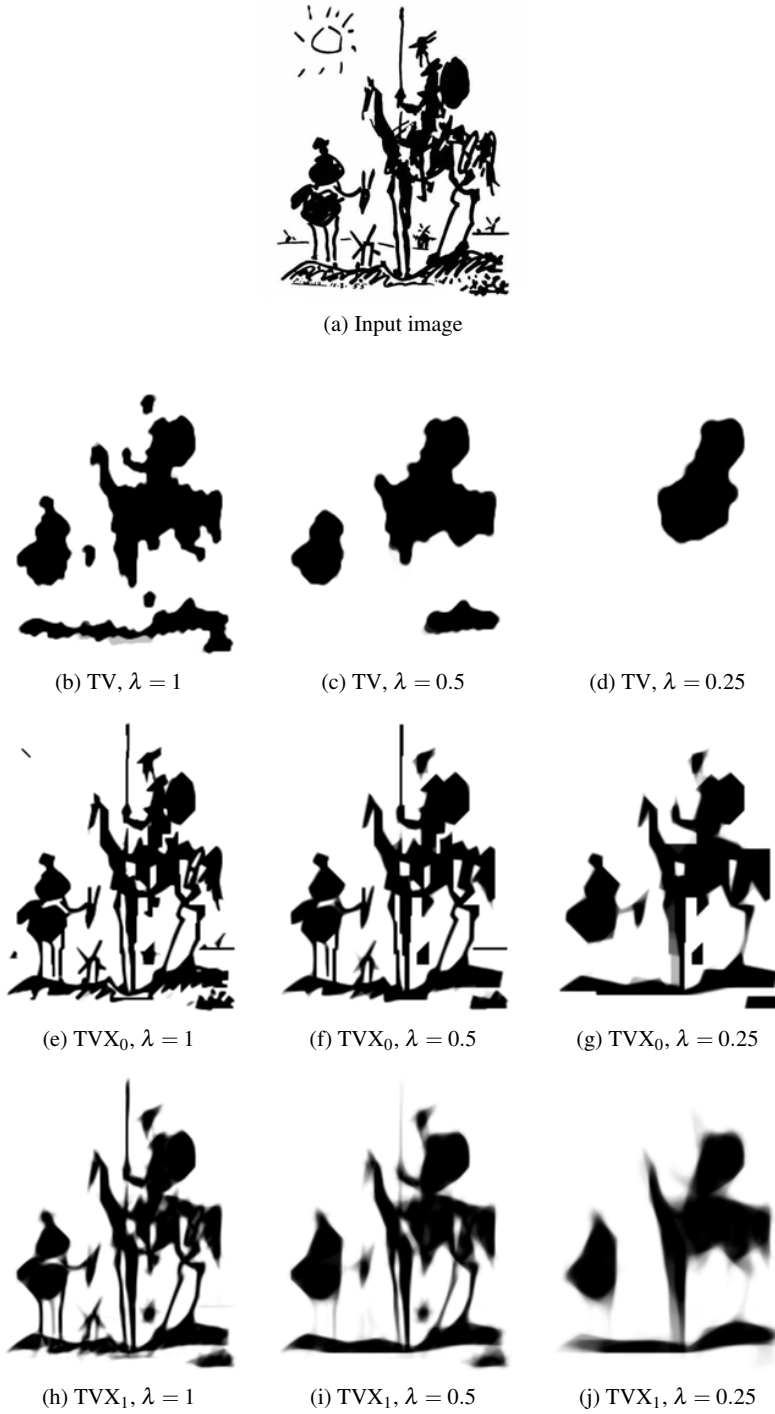
(a) Input image

(b) TV, $\lambda = 1$ (c) TV, $\lambda = 0.5$ (d) TV, $\lambda = 0.25$

Fig. 7 Comparison between TV, TVX$_0$ and TVX$_1$ regularization for different parameters λ. (a) shows the input image and (b)-(j) show the segmented images. While TV based regularization leads to a strong shrinkage of the objects, TVX$_0$ and TVX$_1$ based regularization much better preserve elongated structures. Note that TVX$_0$ leads to the development of simple polygonal structures.

(e) TVX$_0$, $\lambda = 1$ (f) TVX$_0$, $\lambda = 0.5$ (g) TVX$_0$, $\lambda = 0.25$

(h) TVX$_1$, $\lambda = 1$ (i) TVX$_1$, $\lambda = 0.5$ (j) TVX$_1$, $\lambda = 0.25$

tion boundary and hence leads to a simplification in terms of polygonal structures with a minimum number of vertices. TVX$_1$-based regularization yields to smooth object boundaries while preserving to some extent long elongates structures. For a stronger regularization, one can observe less binary results which is explained by the non-uniqueness of the minimizer and the less tightness of the relaxation for weaker data terms.

5.4.2 Image denoising

Next, we consider the task of image denoising (see also Example 4.9 for more information). In the first denoising example we assume the image to be corrupted by Gaussian noise and hence a squared ℓ_2 norm gives a suitable data model:

$$G(u) = \frac{\lambda}{2}\|u - f\|_2^2,$$

where $f \in \mathbb{R}^{MN}$ is the noisy image and $\lambda > 0$ is a regularization parameter. The proximity operator with respect to G is given by

$$u = \text{prox}_{\mathscr{T}^u, G}(\hat{u}) \iff u_{i,j} = \frac{\hat{u}_{i,j} + \lambda \, \mathscr{T}_{i,j}^u f_{i,j}}{1 + \lambda \, \mathscr{T}_{i,j}^u}.$$

Figure 8 shows the results of denoising an image that has been degraded by Gaussian noise of standard deviation $\sigma = 0.1$. For TV regularization we set $\lambda = 15$ and for TVX_1 regularization where we set $\lambda = 4$. We observe that the TVX_1 model leads to a better continuation of line-like structures and hence to an improved denoising quality.

In the second denoising example we consider the case of impulse noise such as salt and pepper noise. In this case, a ℓ_1 data term is known to perform better. It is defined as

$$G(u) = \lambda \|u - f\|_1,$$

and the corresponding proximity operator is given by the well-known shrinkage formula

$$u = \text{prox}_{\mathscr{T}^u, G}(\hat{u}) \iff$$
$$u_{i,j} = \begin{cases} \hat{u}_{i,j} - \lambda \, \mathscr{T}_{i,j}^u & \text{if } \hat{u}_{i,j} - f_{i,j} > \lambda \, \mathscr{T}_{i,j}^u \\ \hat{u}_{i,j} + \lambda \, \mathscr{T}_{i,j}^u & \text{if } \hat{u}_{i,j} - f_{i,j} < -\lambda \, \mathscr{T}_{i,j}^u \\ f_{i,j} & \text{if } |\hat{u}_{i,j} - f_{i,j}| \leq \lambda \, \mathscr{T}_{i,j}^u. \end{cases}$$

Figure 9 shows the results of denoising a natural image that has been corrupted by 25% salt and pepper noise. For the TV model we set $\lambda = 2.5$ and for the TVX_1 we set $\lambda = 1.5$. The results show that in this example, the TVX_1 regularizer performs significantly better than the standard TV regularizer. The reason for the improvement is twofold. Firstly, the noisy pixels strongly contribute to the total number of vertices of the image and hence can be effectively removed by the regularizer. Secondly, the image contains a lot of line-like structures which are nicely preserved by the TVX_1 regularizer.

5.4.3 Image inpainting

In the last application we investigate the proposed regularizers for image inpainting. Similar to Example 4.10 let us denote by $\Omega = \{(i,j) \,|\, 1 \leq i \leq M, \, 1 \leq j \leq N\}$ the index set of the discrete image $f \in \mathbb{R}^{MN}$ and by $\Omega' \subset \Omega$ the index set of given image pixels. We consider the following data model:

$$G(u) = \sum_{(i,j) \in \Omega'} \iota_{\{f_{i,j}\}}(u_{i,j}).$$

The corresponding proximity operator is given by

$$u = \text{prox}_{\mathscr{T}^u, G}(\hat{u}) \iff u_{i,j} = \begin{cases} \hat{u}_{i,j} & \text{if } (i,j) \in \Omega \setminus \Omega' \\ f_{i,j} & \text{if } (i,j) \in \Omega'. \end{cases}$$

Figure 10 shows the result of an inpainting problem with 95% lost pixels. TVX_0 and TVX_1 perform significantly better than standard TV inpainting. The improved performance of the vertex penalizing functionals is mainly contributed to their ability to complete object boundaries. We can also see that while TVX_0 prefers polygonal shapes, TVX_1 leads to smoother and hence more natural boundaries.

6 Conclusions

In this paper, we proposed and studied functionals that penalizes a metric defined on the vertices of the level lines of a two dimensional image.

The quintessential starting point of this work was formed by a generic class of functionals on polygons which penalize the polygon vertices. Each functional depends on a metric on S^1 and is defined as the sum over all polygon vertices x_i of the metric distance between the orientations of the polygon segments meeting in x_i. These functionals can be reformulated as a sum of certain line integrals along the polygon segments, an observation which leads to an equivalent expression using so-called functional lifting of the gradient of the polygon's characteristic function. This new formulation lends itself as a generalization of the functional to general gray scale images for which the functional is related to penalizing the curvature of the image level lines.

The discrete and the geodesic metric have exemplarily been considered as two special cases of metrics on S^1. The corresponding functionals on polygons respectively count the number of vertices or sum up the absolute values of all external vertex angles. While the first functional yields infinity for non-polygonal shapes, the second also has a meaning for shapes with curved boundaries in which case it computes the total absolute curvature of the boundary. Consequently, in applications the functionals can be observed to prefer polygonal and convex shapes, respectively.

To enable global minimization and hence the applicability to general imaging problems, we proposed a convex relaxation that separately treats two nonlinear operations that are involved in the functional: The extraction of all image level lines and the functional lifting of the image gradient. It is still an open problem whether and how a tighter convex relaxation can be found. For practical implementation of the functionals, we proposed a finite differences approximation of the energies and minimized it using a preconditioned first-order primal dual algorithm. We showed promising results for binary image segmentation, image denoising and the recovery of lost image information.

A natural next step for future investigation would be to consider intermediate metrics such as for instance the pointwise minimum of the discrete and the geodesic metric, which for a given shape would lead to an approximation
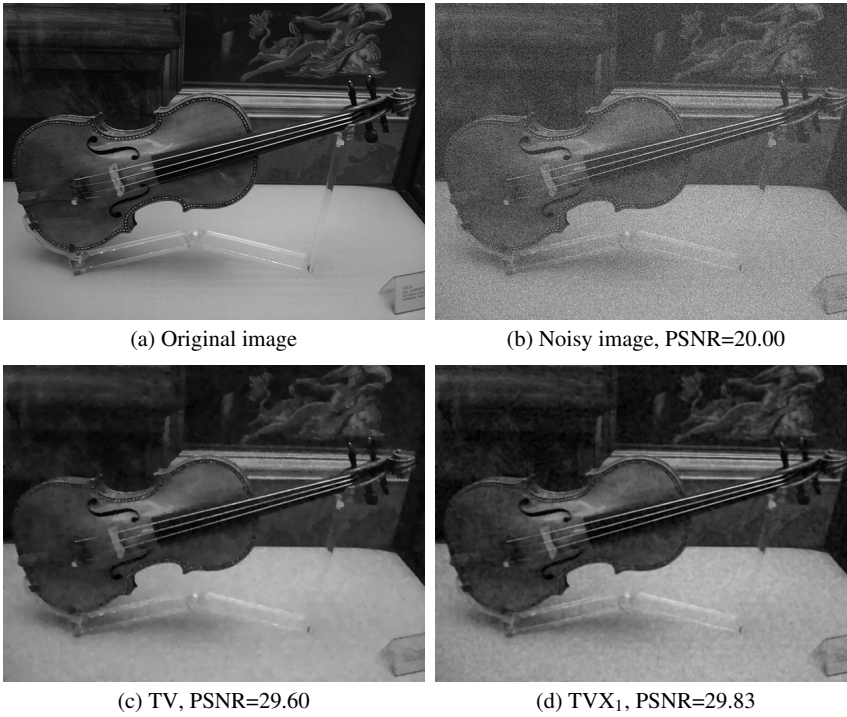
(a) Original image   (b) Noisy image, PSNR=20.00

(c) TV, PSNR=29.60   (d) TVX$_1$, PSNR=29.83

Fig. 8 Denoising of an image degraded by Gaussian noise. (a) and (b) show the original and the noisy image that has been degraded by adding Gaussian noise of standard deviation $\sigma = 0.1$. (c) and (d) show the results of TV and TVX$_1$ denoising. Observe that the TVX$_1$ regularizer leads to a better continuation of line-like structures which improves the denoising quality.
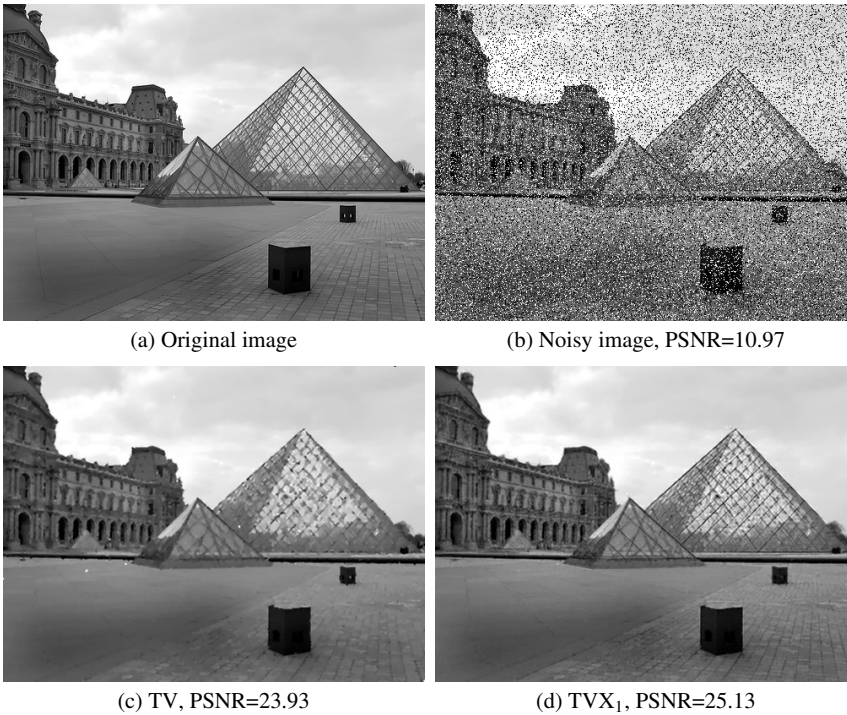

(a) Original image   (b) Noisy image, PSNR=10.97

(c) TV, PSNR=23.93   (d) TVX$_1$, PSNR=25.13

Fig. 9 Denoising of an image degraded by salt and pepper noise. (a) and (b) show the original image and the noisy image that has been degraded by 25% salt and pepper noise. (c) and (d) show the results of TV and TVX$_1$ denoising. Note that the TVX$_1$ model effectively removes the noise while nicely preserving line-like structures.

of the total absolute curvature of all smooth arcs plus the number of vertices. Another possibility would consist in the use of inhomogeneous metrics (for example by dilating distances only in certain regions of S^1). Such metrics would make directional changes of level lines more difficult in certain directions, which might be of interest for example if an image exhibits many lines along certain predominant orientations. In principle one could even relax the restriction

of using metrics on S^1 which might enable to give preference to certain vertex angles or certain combinations of line orientations. However, the structure of the functionals only allows to penalize curvature of image level lines in a positively one-homogeneous manner. Therefore, the approach cannot be employed to approximate for instance the Willmore functional or similar functionals, which is a topic of further investigation.
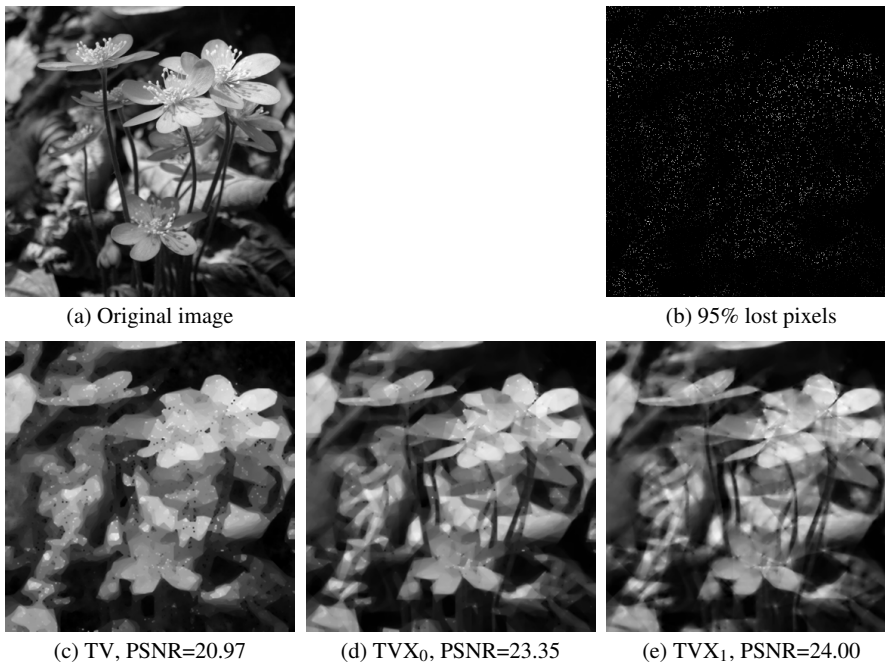
(a) Original image

(b) 95% lost pixels

Fig. 10 Recovery of lost information in an image. (a) and (b) show the original image and the image where 95% of the pixels are lost. (c)-(e) show the results of TV, TVX_0 and TVX_1 inpainting. One can observe that the proposed regularizers are much more efficient in recovering lost information in natural images.

(c) TV, PSNR=20.97

(d) TVX_0, PSNR=23.35

(e) TVX_1, PSNR=24.00

An extension of the proposed approach to three dimensions is not straightforward (one might for example think of penalizing the edge lengths or vertex numbers of polyhedra). The functional lifting of image gradients can be performed (note that $\sigma^\perp$ would have to be replaced by σ in Definition 2.3), but it is not obvious how to obtain information about image level set edges via testing the lifted image gradients with smooth functions. One might for example think of employing Stokes' theorem and replacing the directional derivative in Definition 3.10 by $\mathrm{curl}\,\psi(x,\vartheta)\cdot\vartheta$ for certain smooth test functions $\psi:\Omega\times S^2\to\mathbb{R}^3$.

Finally, future work will concentrate on the study of the proposed functionals in the context of more complex imaging problems such as multi-label image segmentation [7], disocclusion models [29] and general inverse problems.

References

1. Ambrosio, L., Fusco, N., Pallara, D.: Functions of bounded variation and free discontinuity problems. The Clarendon Press Oxford University Press, New York (2000)
2. Ambrosio, L., Masnou, S.: A direct variational approach to a problem arising in image reconstruction. Interfaces and Free Boundaries **5**, 63–81 (2003)
3. Ballester, C., Bertalmio, M., Caselles, V., Sapiro, G., Verdera, J.: Filling-in by joint interpolation of vector fields and gray levels. IEEE Trans. Image Processing **10**, 1200–1211 (2001)
4. Blake, A., Zisserman, A.: Visual Reconstruction. MIT Press (1987)
5. Bredies, K., Kunisch, K., Pock, T.: Total generalized variation. SIAM Journal on Imaging Sciences **3**(3), 492–526 (2010)
6. Cai, X., Gu, G., He, B., Yuan, X.: A relaxed customized proximal point algorithm for separable convex programming. Tech. rep., Department of Mathematics, Hong Kong Baptist University, China (2011)
7. Chambolle, A., Cremers, D., Pock, T.: A convex approach for computing minimal partitions. Tech. Rep. 649, CMAP, Ecole Polytechnique, France (2008)
8. Chambolle, A., Lions, P.L.: Image recovery via total variation minimization and related problems. Numer. Math. **76**(2), 167–188 (1997)
9. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. Journal of Mathematical Imaging and Vision **40**(1), 120–145 (2010)
10. Chan, T.F., Kang, S.H., Shen, J.: Euler's elastica and curvature based inpaintings. SIAM J. Appl. Math. **63**, 564–594 (2002)
11. Chan, T.F., Shen, J.: Nontexture inpainting by curvature driven diffusion (cdd). J. Visul Comm. Image Rep. **12**, 436–449 (2001)
12. Citti, G., Sarti, A.: A cortical based model of perceptual completion in the roto-translation space. Journal of Mathematical Imaging and Vision **24**(3), 307–326 (2006)
13. DiBenedetto, E.: Real analysis. Birkhäuser Advanced Texts: Basler Lehrbücher. [Birkhäuser Advanced Texts: Basel Textbooks]. Birkhäuser Boston Inc., Boston, MA (2002)
14. Droske, M., Rumpf, M.: A level set formulation for Willmore flow. Interfaces and Free Boundaries **6**(3), 361–378 (2004)
15. Eckstein, J., Bertsekas, D.P.: On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators. Math. Programming **55**(3, Ser. A), 293–318 (1992). DOI 10.1007/BF01581204
16. El-Zehiry, N., Grady, L.: Fast global optimization of curvature. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR2010), pp. 3257–3264 (2010)
17. Evans, L.C., Gariepy, R.F.: Measure theory and fine properties of functions. CRC Press, Boca Raton, FL (1992)
18. Franken, M., Rumpf, M., Wirth, B.: A phase field based PDE constraint optimization approach to time discrete Willmore flow. International Journal of Numerical Analysis and Modeling (2011)

19. Geman, S., Geman, D.: Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. IEEE Trans. PAMI **6**(6), 721–741 (1984)

20. Goldluecke, B., Cremers, D.: Introducing total curvature for image processing. In: IEEE International Conference on Computer Vision (ICCV) (2011)

21. Gol'shtein, E., Tret'yakov, N.: Modified Lagrangians in convex programming and their generalizations. Mathematical Programming Study **10**, 86–97 (1979)

22. Grzhibovskis, R., Heintz, A.: A convolution-thresholding approximation of generalized curvature flows. SIAM J. Numer. Anal. **42**, 2652–2670 (2004). DOI 10.1137/S0036142903431316

23. He, B., Yuan, X.: Convergence analysis of primal-dual algorithms for total variation image restoration. Tech. rep., Nanjing University, China (2010)

24. Kanizsa, G.: Organization in Vision. Praeger, New York (1979)

25. Lellmann, J., Schnörr, C.: Continuous multiclass labeling approaches and algorithms. SIAM Journal on Imaging Sciences **4**(4), 1049–1096 (2011). DOI 10.1137/100805844

26. Masnou, S., Morel, J.M.: Level-lines based disocclusion. In: Proceedings of 5th IEEE International Conference on Image Processing (ICIP), pp. 259–263 (1998)

27. Mumford, D.: Elastica and computer vision. Algebraic Geometry and Its Applications pp. 491–506 (1994)

28. Mumford, D., Shah, J.: Optimal approximation by piecewise smooth functions and associated variational problems. Comm. Pure Appl. Math. **42**, 577–685 (1989)

29. Nitzberg, M., Mumford, D., Shiota, T.: Filtering, segmentation, and depth,. In: Lecture Notes in Comp. Sci. 662 (1993)

30. Pock, T., Chambolle, A.: Diagonal preconditioning for first order primal-dual algorithms. In: International Conference of Computer Vision (ICCV 2011), pp. 1762–1769 (2011)

31. Pock, T., Cremers, D., Bischof, H., Chambolle, A.: Global solutions of variational models with convex regularization. SIAM Journal on Imaging Sciences **3**(4), 1122–1145 (2010). DOI 10.1137/090757617

32. Rockafellar, R.T.: Monotone operators and the proximal point algorithm. SIAM Journal on Control and Optimization **14**(5), 877–898 (1976)

33. Rudin, L., Osher, S.J., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60**, 259–268 (1992). [also in *Experimental Mathematics: Computational Issues in Nonlinear Science* (Proc. Los Alamos Conf. 1991)]

34. Schoenemann, T., Cremers, D.: Introducing curvature into globally optimal image segmentation: Minimum ratio cycles on product graphs. In: International Conference on Computer Vision (ICCV2007). Rio de Janeiro, Brazil (2007)

35. Schoenemann, T., Kahl, F., Masnou, S., Cremers, D.: A linear framework for region-based image segmentation and inpainting involving curvature penalization. International Journal of Computer Vision (2012). To appear

36. Schoenemann, T., Masnou, S., Cremers, D.: The elastic ratio: Introducing curvature into ratio-based image segmentation. IEEE Transactions on Image Processing **20**(9), 2565–2581 (2011)

37. Strekalovskiy, E., Cremers, D.: Total variation for cyclic structures: Convex relaxation and efficient minimization. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR2011), pp. 1905–1911 (2011)

38. Tai, X.C., Hahn, J., Chung, G.J.: A fast algorithm for Euler's elastica model using augmented Lagrangian method. SIAM J. Imaging Sciences **4**(1), 313–344 (2011)

39. Weickert, J.: Anisotropic diffusion in image processing. Kaiserslautern (1996)

40. Willmore, T.: Riemannian Geometry. Clarendon Press (1993)