Second Order Differences of Cyclic Data and Applications in Variational Denoising

Ronny Bergmann[*], Friederike Laus[*], Gabriele Steidl[*], Andreas Weinmann[†]

October 12, 2014

Abstract

In many image and signal processing applications, as interferometric synthetic aperture radar (SAR), electroencephalogram (EEG) data analysis, ground based astronomy, and color image restoration in HSV or LCh spaces the data has its range on the one-dimensional sphere $\mathbb{S}^1$. Although the minimization of total variation (TV) regularized functionals is among the most popular methods for edge-preserving image restoration such methods were only very recently applied to cyclic structures. However, as for Euclidean data, TV regularized variational methods suffer from the so called staircasing effect. This effect can be avoided by involving higher order derivatives into the functional.

This is the first paper which uses higher order differences of cyclic data in regularization terms of energy functionals for image restoration. We introduce absolute higher order differences for $\mathbb{S}^1$-valued data in a sound way which is independent of the chosen representation system on the circle. Our absolute cyclic first order difference is just the geodesic distance between points. Similar to the geodesic distances the absolute cyclic second order differences have only values in $[0, \pi]$. We update the cyclic variational TV approach by our new cyclic second order differences. To minimize the corresponding functional we apply a cyclic proximal point method which was recently successfully proposed for Hadamard manifolds. Choosing appropriate cycles this algorithm can be implemented in an efficient way. The main steps require the evaluation of proximal mappings of our cyclic differences for which we provide analytical expressions. Under certain conditions we prove the convergence of our algorithm. Various numerical examples with artificial as well as real-world data demonstrate the advantageous performance of our algorithm.

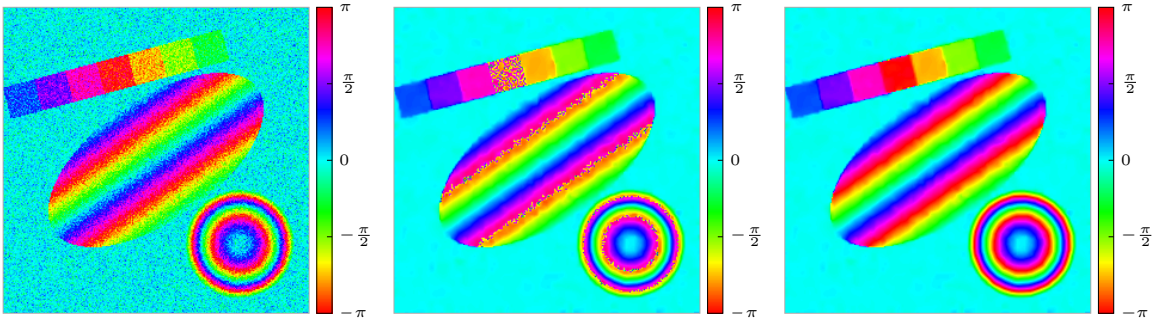
1 Introduction

A frequently used method for edge-preserving image denoising is the variational approach which minimizes the Rudin-Osher-Fatemi (ROF) functional [50]. In a discrete (penalized) form the ROF functional can be written as

$$\sum_{i,j}(f_{i,j} - x_{i,j})^2 + \lambda \sum_{i,j} |\nabla x_{i,j}|, \quad \lambda > 0,$$

[*]Fachbereich für Mathematik, Technische Universität Kaiserslautern, Paul-Ehrlich-Str. 31, 67663 Kaiserslautern, Germany, {bergmann, steidl, friederike.laus}@mathematik.uni-kl.de.

[†]Department of Mathematics, Technische Universität München and Fast Algorithms for Biomedical Imaging Group, Helmholtz-Zentrum München, Ingolstädter Landstr. 1, 85764 Neuherberg, Germany, andreas.weinmann@helmholtz-muenchen.de.



(a) Noisy phase valued data. (b) Restoration by a real-valued model. (c) Restoration by our cyclic model.

Figure 1. (a) Phase-valued noisy image. (b) Restored image using a real-valued model with first and second order differences. Since the cyclic nature of the data is ignored, e.g., the identical spherical data $-\pi$ and π are considered to have the Euclidean distance of 2π, there appear hard artifacts. (c) Restored image using our model (21), see Figure 6 (f). For further explanations of the experiment we refer to Subsection 5.2.

where $f \in \mathbb{R}^{N,M}$ is the given corrupted image and ∇ denotes the discrete gradient operator which contains usually first order forward differences in vertical and horizontal directions. The regularizing term $\sum_{i,j}|\nabla x_{i,j}|$ can be considered as discrete version of the total variation (TV) functional. Since the gradient does not penalize constant areas the minimizer of the ROF functional tends to have such regions, an effect known as staircasing. An approach to avoid this effect consists in the employment of higher order differences/derivatives. Since the pioneering work [12] which couples the TV term with higher order terms by infimal convolution various techniques with higher order differences/derivatives were proposed in the literature, among them [9, 14, 15, 19, 20, 32, 34, 37, 38, 44, 51, 52, 53].

In various applications in image processing and computer vision the functions of interest take values on the circle $\mathbb{S}^1$ or another manifold. Handling phase-valued data with techniques for real-valued data introduces artifacts since the cyclic nature of the data is neglected. Figure 1 demonstrates the denoising of phase-valued data by applying a model with real-valued first and second order differences versus the model proposed in this paper which takes the cyclic structure of the data into account.

Therefore, processing manifold-valued data has gained a lot of interest in recent years. Examples are wavelet-type multiscale transforms for manifold data [30, 47, 62], manifold-valued partial differential equations [16, 29], and ground-based astronomy [3, 13, 17, 28]. Finally we like to mention statistical issues on Riemannian manifolds [23, 24, 46], and in particular the statistics of circular data [22, 33]. The TV notation for functions with values on a manifold has been studied in [26, 27] using the theory of Cartesian currents. These papers were an extension of the previous work [25] were the authors focus on $\mathbb{S}^1$-valued functions and show in particular the existence of minimizers of certain energies in the space of functions with bounded total cyclic variation. The first work which applies a cyclic TV approach among other models for imaging tasks was recently published by Cremers and Strekalovskiy in [56, 57]. The authors unwrapped the function values to the real axis and proposed an algorithmic solution to ac-

count for the periodicity. An algorithm which solves TV regularized minimization problems on Riemannian manifolds was proposed by Lellmann et al. in [35]. They reformulate the problem as a multilabel optimization problem with an infinite number of labels and approximate the resulting hard optimization problem using convex relaxation techniques. The algorithm was applied for chromaticity-brightness denoising, denoising of rotation data and processing of normal fields for visualization. Another approach to TV minimization for manifold-valued data via cyclic and parallel proximal point algorithms was proposed by one of the authors and his colleagues in [63]. It does not require any labeling or relaxation techniques. The authors apply their algorithm in particular for diffusion tensor imaging and interferometric SAR imaging. For Cartan-Hadamard manifolds convergence of the algorithm was shown based on a recent result of Bačák [2]. Unfortunately, one of the simplest manifolds that is not of Cartan-Hadamard type is the circle $\mathbb{S}^1$.

In this paper we deal with the incorporation of higher order differences into the energy functionals to improve denoising results for $\mathbb{S}^1$-valued data. Note that the (second-order) total generalized variation was generalized for tensor fields in [58]. However, to the best of our knowledge this is the first paper which defines second order differences of cyclic data and uses them in regularization terms of energy functionals for image restoration. We focus on a discrete setting. First we provide a meaningful definition of higher order differences for cyclic data which we call *absolute cyclic differences*. In particular our absolute cyclic first order differences resemble the geodesic distance (arc length distance) on the circle. As the geodesics the absolute cyclic second order differences take only values in $[0, \pi]$. This is not necessary the case for differences of order larger than two. Following the idea in [63] we suggest a cyclic proximal point algorithm to minimize the resulting functionals. This algorithm requires the evaluation of certain proximal mappings. We provide analytical expression for these mappings. Further, we suggest an appropriate choice of the cycles such that the whole algorithm becomes very efficient. We apply our algorithm to artificial data as well as to real-world interferometric SAR data. Moreover we provide an example from electroencephalography, namely on the habituation in single–trial sequences of auditory event–related responses which were considered for example in [43, 55].

The paper is organized as follows: in Section 2 we propose a definition of differences on $\mathbb{S}^1$. Then, in Section 3, we provide analytical expressions for the proximal mappings required in our cyclic proximal point algorithm. The approach is based on unwrapping the circle to $\mathbb{R}$ and considering the corresponding proximal mappings on the Euclidean space. The cyclic proximal point algorithm is presented in Section 4 together with a convergence proof under certain assumptions. Further we describe a vectorization strategy which makes the Matlab implementation efficient and provides parallel computing. Section 5 demonstrates the advantageous performance of our algorithm by numerical examples. Finally, conclusions and directions of future work are given in Section 6.

2 Differences of $\mathbb{S}^1$–valued data

In this section we propose a definition of higher order differences for cyclic data using corresponding points on the tangential line where we have to take the ambiguity of the exponential map into account. It turns out (Proposition 2.5) that absolute second order cyclic differences can be just obtained by considering real valued second order differences modulo 2π which is no longer true for our higher order differences (Example 2.4).

Let $\mathbb{S}^1$ be the unit circle in the plane

$$\mathbb{S}^1 := \{p_1^2 + p_2^2 = 1 : p = (p_1, p_2)^{\mathrm{T}} \in \mathbb{R}^2\}$$

endowed with the *geodesic distance* (arc length distance)

$$d_{\mathbb{S}^1}(p, q) = \arccos(\langle p, q\rangle).$$

Given a base point $q \in \mathbb{S}^1$, the *exponential map* $\exp_q : \mathbb{R} \to \mathbb{S}^1$ from the tangent space $T_q\mathbb{S}^1 \simeq \mathbb{R}$ of $\mathbb{S}^1$ at q onto $\mathbb{S}^1$ is defined by

$$\exp_q(x) = R_x q, \qquad R_x := \begin{pmatrix} \cos x & -\sin x \\ \sin x & \cos x \end{pmatrix}.$$

This map is 2π-periodic, i.e., $\exp_q(x) = \exp_q((x)_{2\pi})$ for any $x \in \mathbb{R}$, where $(x)_{2\pi}$ denotes the unique point in $[-\pi, \pi)$ such that $x = 2\pi k + (x)_{2\pi}$, $k \in \mathbb{Z}$. Some useful properties of the mapping $(\cdot)_{2\pi} : \mathbb{R} \to [-\pi, \pi)$ (which can also be considered as mapping from $\mathbb{R}$ onto $\mathbb{R}/2\pi\mathbb{Z}$) are collected in the following remark.

Remark 2.1. *The following relations hold true:*

i) $\big((x)_{2\pi} \pm (y)_{2\pi}\big)_{2\pi} = (x \pm y)_{2\pi}$ *for all* $x, y \in \mathbb{R}$.

ii) If $z = (x - y)_{2\pi}$ *then* $x = (z + y)_{2\pi}$ *for all* $x \in [-\pi, \pi)$, $y \in \mathbb{R}$.

To guarantee the injectivity of the exponential map, we restrict its domain of definition from $\mathbb{R}$ to $[-\pi, \pi)$. Thus, for $p, q \in \mathbb{S}^1$, there is now a unique $x \in [-\pi, \pi)$ satisfying $\exp_q(x) = p$. In particular we have $\exp_q(0) = q$. Given such representation system $x_j \in [-\pi, \pi)$ of $p_j \in \mathbb{S}^1$, $j = 1, 2$ centered at an arbitrary point q on $\mathbb{S}^1$ the geodesic distance becomes

$$d_{\mathbb{S}^1}(p_1, p_2) = d(x_1, x_2) = \min_{k\in\mathbb{Z}} |x_2 - x_1 + 2\pi k| = |(x_2 - x_1)_{2\pi}|. \tag{1}$$

Actually we need only $k \in \{0, \pm 1\}$ in the minimum. Clearly, this definition does not depend on the chosen center point q.
We want to determine general finite differences of $\mathbb{S}^1$-valued data. Let $w = (w_j)_{j=1}^d \in \mathbb{R}^d \backslash \{0\}$ with

$$\langle w, 1_d\rangle = \sum_{j=1}^{d} w_j = 0, \tag{2}$$

where 1_d denotes the vector with d components one. We define the *finite difference operator* $\Delta(\cdot; w) : \mathbb{R}^d \to \mathbb{R}$ by

$$\Delta(x; w) := \langle x, w\rangle \quad \text{for all } x \in \mathbb{R}^d.$$

By (2), we see that $\Delta(\cdot; w)$ vanishes for constant vectors and is therefore translation invariant, i.e.,

$$\Delta(x + \alpha 1_d; w) = \Delta(x; w) \quad \text{for all } \alpha \in \mathbb{R}.$$

Example 2.2. *For the binomial coefficients with alternating signs*

$$w = b_n := \left((-1)^{j+n-1}\binom{n}{j-1}\right)_{j=1}^{n+1}$$

we obtain the (forward) differences of order n:

$$\Delta(x;w) = \Delta_n(x) = \langle x, b_n\rangle = \sum_{j=1}^{n+1}(-1)^{j+n-1}\binom{n}{j-1}x_j.$$

Note that Δ_n does not only fulfill (2), but vanishes exactly for all 'discrete polynomials of order $n-1$', i.e., for all vectors from $\operatorname{span}\{(j^r)_{j=0}^n : r = 0,\dots,n-1\}$. Here we are interested in first and second order differences

$$\Delta_1(x_1,x_2) = \Delta(x;b_1) = x_2 - x_1,$$
$$\Delta_2(x_1,x_2,x_3) = \Delta(x;b_2) = x_1 - 2x_2 + x_3.$$

Moreover, we will apply the 'mixed second order' difference with $w = b_{1,1} := (-1,1,1,-1)^{\mathrm{T}}$ and use the notation

$$\Delta_{1,1}(x_1,x_2,x_3,x_4) = \Delta(x;b_{1,1}) = -x_1 + x_2 + x_3 - x_4.$$

We want to define differences for points $(p_j)_{j=1}^d \in (\mathbb{S}^1)^d$ using their representation $x := (x_j)_{j=1}^d \in [-\pi,\pi)^d$ with respect to an arbitrary center point of the exponential map. As the geodesic distance (1) the cyclic differences should be independent of the choice of the center point. This can be achieved if and only if the differences are shift invariant modulo 2π. Let $\mathbb{I}_d := \{1,\dots,d\}$. We define the *absolute cyclic difference* of $x \in [-\pi,\pi)^d$ (resp. $(p_j)_{j=1}^d \in (\mathbb{S}^1)^d$) with respect to w by

$$d(x;w) := \min_{\alpha\in\mathbb{R}}\big|\Delta\big([x+\alpha 1_d]_{2\pi};w\big)\big| = \min_{j\in\mathbb{I}_d}\big|\Delta\big([x-(x_j+\pi)1_d]_{2\pi};w\big)\big|, \tag{3}$$

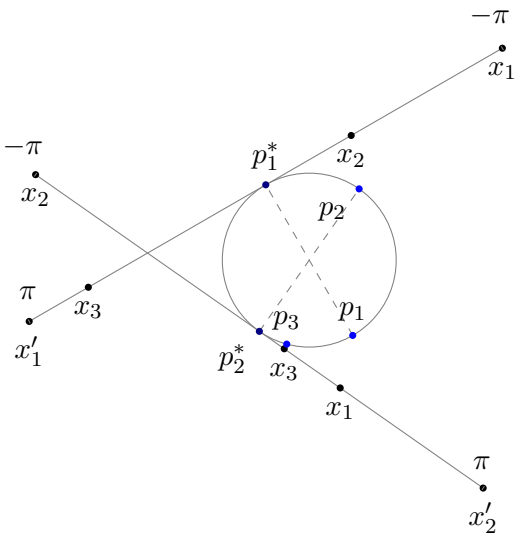
where $[x]_{2\pi}$ denotes the component-by-component application of $(t)_{2\pi}$ if $t \neq (2k+1)\pi$, $k\in\mathbb{Z}$ and $[(2k+1)\pi]_{2\pi} = \pm\pi$, $k\in\mathbb{Z}$. The definition allows that points having the same value are treated separately, cf. Figure 3. This ensures that $d(\cdot;w) : (\mathbb{S}^1)^d \to \mathbb{R}$ is a continuous map. For example we have $d\big((-\pi,0,-\pi)^{\mathrm{T}};b_2\big) = 0$. Figures 2 and 3 illustrate definition (3). The definition can also be seen as minimizing differences of points on a tangential line with respect to shifts by angles α. For each of these shifts all points having the value $-\pi$ are also placed at $+\pi$ by introducing $[\cdot]_{2\pi}$. Then, in fact, the various differences can be achieved for those cases, where one or more data points x_j lie at the boundary, i.e. have the value $x_j = \pm\pi$. This leads directly to the second formula for the absolute cyclic differences.
For the absolute cyclic differences related to the differences in Example 2.2 we will use the simpler notation

$$d_n(x) := d(x;b_n) \quad \text{and} \quad d_{1,1}(x) := d(x;b_{1,1}).$$

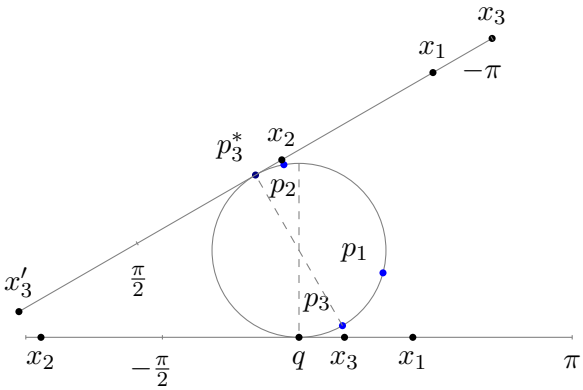
The following equivalent definition of absolute cyclic differences appears to be useful.

Lemma 2.3. *Let $x \in [-\pi,\pi)^d$ be sorted in ascending order as $-\pi \le x_{j_1} \le \ldots \le x_{j_d} < \pi$ and set $x^1 := (x_{j_i})_{i=1}^d$. Let P denote the corresponding permutation matrix, i.e., $Px = x^1$ and $x = P^{\mathrm{T}}x^1$. Consider the 2π shifted versions of x^1 given by*

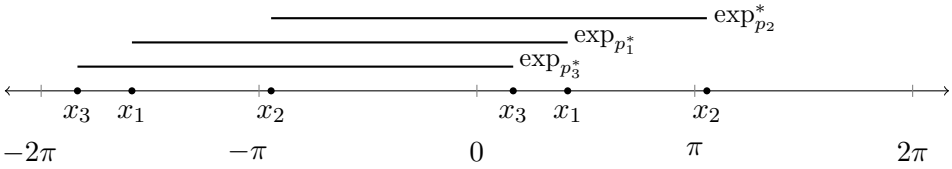
$$x^k = x^1 + 2\pi\sum_{j=1}^{k-1} e_j \quad k = 2,\dots,d,$$



(a) $\exp_{p_1^*}$ (top) and $\exp_{p_2^*}$ (bottom).



(b) $\exp_{p_3^*}$ (top) and $\exp_q$ (bottom).



(c) Settings from the tangential maps of p_j^*, $j = 1, 2, 3$, on $\mathbb{R}$ using the representation system according to $\exp_q$.

Figure 2. Three points p_j, $j = 1, 2, 3$, on the circle (blue) and their inverse exponential maps at p_j^*, $j = 1, 2, 3$, (dark blue), where p_j^* denotes the antipodal point of p_j. In other words, we cut the circle at the point p_j and unwind it with respect to the tangent line at the antipodal point p_j^*. The absolute cyclic differences take the three pairwise different positions of the points x_j, $j = 1, 2, 3$ to each other into account. These are shown in (c) with respect to the representation system from the arbitrary point q in (b).

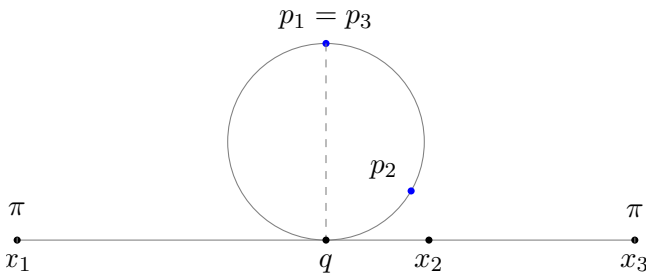


Figure 3. Three points p_j, $j = 1, 2, 3$ on the circle, where $p_1 = p_3$ and $\exp_q$, $q = p_1^*$. Though p_1, p_3 denote the same point on the circle they are treated separately in the definition of the absolute cyclic differences.

where $e_j \in \mathbb{R}^d$ denotes the j-th unit vector. Then it holds

$$d(x;w) = \min_{k\in\mathbb{I}_d} \big|\Delta(P^{\mathrm{T}}x^k; w)\big| = \min_{k\in\mathbb{I}_n} \Big|\Delta(x^k; w) + 2\pi\Big\langle \sum_{j=1}^{k-1} e_j, Pw\Big\rangle\Big|. \tag{4}$$

Proof. The first equality in (4) follows directly by definition (3). To see the second one, note that by linearity of the inner product we have

$$\langle P^{\mathrm{T}}x^k, w\rangle = \langle P^{\mathrm{T}}x^1, w\rangle + 2\pi\Big\langle \sum_{j=1}^{k-1} e_j, Pw\Big\rangle = \langle x, w\rangle + 2\pi\Big\langle \sum_{j=1}^{k-1} e_j, Pw\Big\rangle. \tag{5}$$

□

For the geodesic distance we obtain by (1) that $d_1(x) = \big|\big(\Delta_1(x))_{2\pi}\big)\big|$. In general the relation

$$d(x;w) = |(\langle x, w\rangle)_{2\pi}| \qquad \text{for all } x \in [-\pi, \pi)^d \tag{6}$$

does not hold true as the following example shows.

Example 2.4. *In general the n-th order absolute cyclic difference cannot be written as $d_n(x) = |(\langle x, b_n\rangle)_{2\pi}| = |(\Delta_n(x))_{2\pi})|$. Consider for example the absolute cyclic third order difference for $x := \frac{\pi}{16}(-15, -13, 12, 14)^{\mathrm{T}}$ given by (4) as*

$$d_3(x_1, x_2, x_3, x_4) = \min_{k=1,2,3,4} |\Delta_3(x^k)|, \quad \Delta_3(x) = -x_1 + 3x_2 - 3x_3 + x_4.$$

We obtain

$$\Delta_3(x^1) = \Delta_3(x) = \frac{-46\pi}{16},\ \Delta_3(x^2) = \Delta_3(x^4) = \frac{-78\pi}{16},\ \Delta_3(x^3) = \frac{18\pi}{16},$$

so that $d_3(x) = \frac{18\pi}{16} > \pi$.

For $w \in \{b_2, b_{1,1}\}$ relation (6) holds true by the next lemma.

Proposition 2.5. *For $w \in \{b_2, b_{1,1}\}$ the following relation holds true:*

$$d(x;w) = \min_{k\in\mathbb{Z}} |\Delta(x;w) + 2\pi k| = \big|\big(\Delta(x;w)\big)_{2\pi}\big|.$$

Note that we need only the minimum over $k \in \{0, \pm 1, \pm 2\}$ in Proposition 2.5 and more precisely

$$d(x;w) = \begin{cases} |\Delta(x;w)| & \text{if } |\Delta(x;w)| \in [0,\pi], \\ |\Delta(x;w) - 2\pi\sigma| = 2\pi - |\Delta(x;w)| & \text{if } |\Delta(x;w)| \in (\pi, 2\pi], \\ |\Delta(x;w)| - 2\pi & \text{if } |\Delta(x;w)| \in (2\pi, 3\pi], \\ |\Delta(x;w) - 4\pi\sigma| = 4\pi - |\Delta(x;w)| & \text{if } |\Delta(x;w)| \in (3\pi, 4\pi), \end{cases}$$

where $\sigma = \operatorname{sgn}(\Delta(x;w)) \in \{-1, 1\}$ and

$$\operatorname{sgn}(x) := \begin{cases} 1 & \text{if } x > 0, \\ 0 & \text{if } x = 0, \\ -1 & \text{otherwise.} \end{cases}$$

Proof. Since $|x_j - x_k| < 2\pi$ for $x_j, x_k \in [-\pi, \pi)$, we see that $|\Delta_2(x)| < 4\pi$ and $|\Delta_{1,1}(x)| < 4\pi$. First we consider d_2. By Lemma 2.3 we obtain

$$d_2(x) = \min_{k\in\mathbb{I}_3} \big|\Delta(P^{\mathrm{T}}x^k; b_2)\big| = \min_{k\in\mathbb{I}_3} \Big|\Delta_2(x) + 2\pi\big\langle \sum_{j=1}^{k} e_j, Pb_2\big\rangle\Big|, \tag{7}$$

where we can assume by the cyclic shift invariance of d_2 that $x_{j_1} = x_1$.
If $x^1 = (x_1, x_2, x_3)^{\mathrm{T}}$, then the corresponding permutation matrix P in Lemma 2.3 is the identity matrix. Further we obtain that $\Delta_2(x) = (x_1 - x_2) + (x_3 - x_2) \in (-2\pi, 2\pi)$ and by (7) we get

$$|\Delta_2(P^{\mathrm{T}}x^2)| = |\Delta_2(x^2)| = |\Delta_2(x) + 2\pi| \qquad \text{and} \qquad |\Delta_2(x^3)| = |\Delta_2(x) - 2\pi|.$$

If $x^1 = (x_1, x_3, x_2)^{\mathrm{T}}$, then $P = \left(\begin{smallmatrix} 1&0&0 \\ 0&0&1 \\ 0&1&0 \end{smallmatrix}\right)$ and $\Delta_2(x) \in (-4\pi, 0]$. In this case we get

$$|\Delta_2(P^{\mathrm{T}}x^2)| = |\Delta_2(x) + 2\pi| \qquad \text{and} \qquad |\Delta_2(P^{\mathrm{T}}x^3)| = |\Delta_2(x) + 4\pi|.$$

This proves the first assertion.
For $d_{1,1}$ we can again assume that $x_{j_1} = x_1$. Exploiting that

$$\Delta_{1,1}(x_1, x_2, x_3, x_4) = \Delta_{1,1}(x_1, x_3, x_2, x_4)$$

we have to consider the following three cases:
If $x^1 = (x_1, x_2, x_3, x_4)^{\mathrm{T}}$, then P is the identity matrix, $\Delta_{1,1}(x) = (x_2 - x_1) + (x_3 - x_4) \in (-2\pi, 2\pi)$ and

$$|\Delta_{1,1}(x^2)| = |\Delta_2(x) - 2\pi|, \quad |\Delta_{1,1}(x^3)| = |\Delta_2(x)|, \quad |\Delta_{1,1}(x^4)| = |\Delta_2(x) + 2\pi|.$$

If $x^1 = (x_1, x_2, x_4, x_3)^{\mathrm{T}}$, then $P = \left(\begin{smallmatrix} 1&0&0&0 \\ 0&1&0&0 \\ 0&0&0&1 \\ 0&0&1&0 \end{smallmatrix}\right)$ and $\Delta_2(x) \in [0, 2\pi)$. By (5) we have

$$|\Delta_{1,1}(x^2)| = |\Delta_2(x) - 2\pi|, \quad |\Delta_{1,1}(x^3)| = |\Delta_2(x)|, \quad |\Delta_{1,1}(x^4)| = |\Delta_2(x) - 2\pi|.$$

If $x^1 = (x_1, x_4, x_2, x_3)^{\mathrm{T}}$, then $P = \left(\begin{smallmatrix} 1&0&0&0 \\ 0&0&0&1 \\ 0&1&0&0 \\ 0&0&1&0 \end{smallmatrix}\right)$ and $\Delta_2(x) \in [0, 4\pi)$. Here we obtain

$$|\Delta_{1,1}(x^2)| = |\Delta_2(x) - 2\pi|, \quad |\Delta_{1,1}(x^3)| = |\Delta_2(x) - 4\pi|, \quad |\Delta_{1,1}(x^4)| = |\Delta_2(x) - 2\pi|.$$

This finishes the proof. □

For real-valued data, the finite differences considered in this paper (divided by the grid size) reduce to the standard finite differences and thus approach the corresponding higher order derivatives as the grid size gets smaller and the discrete energies converge to their corresponding continuous counterparts. Let us give some comments on the corresponding question for our higher order differences of cyclic data.

Remark 2.6. *It is quite common when dealing with manifold-valued functions, to define the corresponding function spaces by embedding the (Riemannian) manifold into Euclidean space and using the corresponding Euclidean definition assuming that the functions take their values in the manifold; see, e.g., [10]. Since our approach is based on covering spaces, we may use the universal covering $\mathbb{R}$ of $\mathbb{S}^1$ rather than the embedding space $\mathbb{R}^2$. Then, we may consider a function $f : \mathbb{R}^2 \to \mathbb{S}^1$ which is continuous near a point x, and the corresponding discretization (f_i) on sufficiently fine grids near x. Lifting both function and discretization to the covering space $\mathbb{R}$, we may use the convergence of finite differences to the corresponding derivatives on $\mathbb{R}$. Then we may go back to $\mathbb{S}^1$ with the canonical projection. Having in mind that our definition of $\mathbb{S}^1$-valued differences is compatible with this concept, the convergence of the corresponding discrete energies can be* locally *observed.*

3 Proximal mapping of absolute cyclic differences

In this section we give explicit expressions for the proximal mappings of our absolute cyclic second order differences. These analytic expressions are the ingredients for the proximal cyclic point algorithm described in the next section. The cyclic proximal mappings can be deduced from explicit expressions of the corresponding real valued proximal mappings. However, due to the ambiguity of the exponential map, a precise, sophisticated treatment of the involved functionals is necessary.

For a proper, closed, convex function $\varphi : \mathbb{R}^N \to (-\infty, +\infty]$ and $\lambda > 0$ the *proximal mapping* $\operatorname{prox}_{\lambda\varphi} : \mathbb{R}^N \to \mathbb{R}^N$ is defined by

$$\operatorname{prox}_{\lambda\varphi}(f) := \operatorname*{arg\,min}_{x\in\mathbb{R}^N} \frac{1}{2}\|f - x\|_2^2 + \lambda\varphi(x),$$

see [42]. The above minimizer exits and is uniquely determined. Many algorithms which were recently used in variational image processing reduce to the iterative computation of values of proximal mappings. An overview of applications of proximal mappings is given in [45]. Proximal mappings were generalized to functions on Riemannian manifolds in [21], replacing the norm by geodesic distances and introducing a proper notion of convexity. Note that for the manifold $\mathbb{S}^1$ there is no notion of convexity of a function, as used in [21], because it is not a Hadamard space.

In this section, we are interested in proximal mappings of absolute cyclic differences $d(\cdot; w)^p$, i.e., $\operatorname{prox}_{\lambda d(\cdot;w)^p} : (\mathbb{S}^1)^d \to (\mathbb{S}^1)^d$, for $w \in \mathbb{R}^d$. More precisely, we will determine for $\mathbb{S}^1$-valued vectors represented by $f \in [-\pi, \pi)^d$ the values

$$\operatorname{prox}_{\lambda d(\cdot;w)^p}(f) := \operatorname*{arg\,min}_{x\in[-\pi,\pi)^d} \frac{1}{2}\sum_{j=1}^{d} d(x_j, f_j)^2 + \lambda d(x; w)^p, \qquad \lambda > 0$$

for $p \in \{1, 2\}$ and first and second order absolute cyclic differences $d(\cdot; w)$, $w \in \{b_1, b_2, b_{1,1}\}$. Here $\arg\min_{x\in[-\pi,\pi)^d}$ means that we are looking for the representative of $x \in (\mathbb{S}^1)^d$ in $[-\pi, \pi)^d$.

In particular, we will see that these proximal mapping are single-valued for $f \in [-\pi, \pi)^d$ with $|(\langle f, w\rangle)_{2\pi}| < \pi$ and have two values for $|(\langle f, w\rangle)_{2\pi}| = \pi$.
We start by considering the proximal mappings of the appropriate differences in $\mathbb{R}^d$. Then we use the results to find the proximal functions of the absolute cyclic differences.

3.1 Proximity of differences on $\mathbb{R}^d$

First we give analytical expressions for $\operatorname{prox}_{\lambda|\langle\cdot,w\rangle - a|^p}$, where $p \in \{1, 2\}$ and $w \in \mathbb{R}^d$, $a \in \mathbb{R}$. Since we could not find a corresponding reference in the literature, the computation of the minimizer of

$$E(x; f, a, w) := \frac{1}{2}\|f - x\|_2^2 + \lambda|\langle x, w\rangle - a|^p, \qquad \lambda > 0$$

is described in the following lemmas. We start with $p = 1$.

Lemma 3.1. *For given $f \in \mathbb{R}^d$ and $0 \neq w \in \mathbb{R}^d$, $a \in \mathbb{R}$ set*

$$s := \operatorname{sgn}(\langle f, w\rangle - a) \quad \textit{and} \quad \mu := \frac{\langle f, w\rangle - a}{\|w\|_2^2}.$$

Then the minimizer $\hat{x}$ of

$$E(x; f, a, w) := \frac{1}{2}\|f - x\|_2^2 + \lambda|\langle x, w\rangle - a|, \qquad \lambda > 0$$

is given by

$$\hat{x} = f - s\, \min\{\lambda, |\mu|\}\, w \tag{8}$$

and the minimum by

$$E(\hat{x}; f, a, w) = \begin{cases} \|w\|_2^2\, \frac{1}{2}\mu^2 & \textit{if } |\mu| \leq \lambda, \\ \|w\|_2^2\big(\frac{1}{2}\lambda^2 + \lambda(|\mu| - \lambda)\big) & \textit{otherwise.} \end{cases} \tag{9}$$

Proof. Since $w \neq 0$, there exists a component $w_j \neq 0$ and we rewrite

$$E(x; f, a, w) = \frac{1}{2}\|f - x\|_2^2 + \lambda|w_j|\,\big|\big\langle \frac{w}{w_j}, x - \frac{a}{w_j}e_j\big\rangle\big|.$$

Substituting $y := x - \frac{a}{w_j}e_j$, $g = f - \frac{a}{w_j}e_j$ and $\nu := \lambda|w_j|$, $v := \frac{w}{w_j}$ we see that $\hat{x} = \hat{y}$, where $\hat{y}$ is the minimizer of

$$F(y; g, v) := \frac{1}{2}\|g - y\|_2^2 + \nu|\langle v, y\rangle|.$$

The (Fenchel) dual problem of $\arg\min_{y\in\mathbb{R}^d} F(y)$ reads

$$\hat{t} := \operatorname*{arg\,min}_{t\in\mathbb{R}} \big\{\|g - t\, w\|_2^2 \quad \text{subject to} \quad |t| \leq \nu\big\} \tag{10}$$

and the relation between the minimizers of the primal and dual problems is given by

$$\hat{y} = g - \hat{t}\, v. \tag{11}$$

Rewriting (10) we see that $\hat{t}$ is the minimizer of

$$(t - \tilde{\mu})^2 \quad \text{subject to} \quad |t| \leq \nu,$$

where $\tilde{\mu} := \frac{\langle v,g\rangle}{\|v\|^2}$. Hence we obtain

$$\hat{t} = \begin{cases} \tilde{\mu} & \text{if } |\tilde{\mu}| \le \nu, \\ \operatorname{sgn}(\tilde{\mu})\nu & \text{otherwise.} \end{cases}$$

and by (11) further

$$\hat{y} = g - \operatorname{sgn}(\tilde{\mu}) \min\{\nu, |\tilde{\mu}|\}\, v.$$

Substituting back results in (8) and plugging $\hat{x}$ into E we get (9). □

Example 3.2. *Let $p = 1$, $a = 0$, and $E(x; f, w) := E(x; f, 0, w)$.*

i) For $w = b_1 = (-1, 1)^{\mathrm{T}}$ and $f \in \mathbb{R}^2$ we get $\|w\|_2^2 = 2$ and $s = \operatorname{sgn}(f_2 - f_1)$ so that the minimizer of $E(x; f, b_1)$ follows by soft shrinkage of f with threshold λ:

$$\hat{x} = \begin{pmatrix} f_1 + s\, m \\ f_2 - s\, m \end{pmatrix}, \qquad m := \min\{\lambda, \frac{|f_2 - f_1|}{2}\}.$$

ii) For $w = b_2 = (1, -2, 1)^{\mathrm{T}}$ and $f \in \mathbb{R}^3$ we obtain $\|w\|_2^2 = 6$ and $s = \operatorname{sgn}(f_1 - 2f_2 + f_3)$. Consequently, the minimizer of $E(x; f, b_2)$ is given by

$$\hat{x} = \begin{pmatrix} f_1 - s\, m \\ f_2 + 2s\, m \\ f_3 - s\, m \end{pmatrix}, \qquad m := \min\left\{\lambda, \frac{|f_1 - 2f_2 + f_3|}{6}\right\}.$$

iii) For $w = b_{1,1} = (-1, 1, 1, -1)^{\mathrm{T}}$ and $f \in \mathbb{R}^4$ we obtain $\|w\|_2^2 = 4$ and $s = \operatorname{sgn}(f_2 - f_1 + f_3 - f_4)$, so that the minimizer of $E(x; f, b_{1,1})$ is given by

$$\hat{x} = \begin{pmatrix} f_1 + s\, m \\ f_2 - s\, m \\ f_3 - s\, m \\ f_4 + s\, m \end{pmatrix}, \qquad m := \min\left\{\lambda, \frac{|f_2 - f_1 + f_3 - f_4|}{4}\right\}.$$

We will apply the following corollary.

Corollary 3.3. *Let $0 \neq w \in \mathbb{R}^d$. Further, let $f, \tilde{f} \in \mathbb{R}^d$ and $a, \tilde{a} \in \mathbb{R}$ be given such that $|\langle f, w\rangle - a| < |\langle \tilde{f}, w\rangle - \tilde{a}|$. Then*

$$\min_{x \in \mathbb{R}^d} E(x; f, a, w) < \min_{x \in \mathbb{R}^d} E(x; \tilde{f}, \tilde{a}, w).$$

Proof. Set $\mu := \frac{\langle f,w\rangle - a}{\|w\|_2^2}$ and $\tilde{\mu} := \frac{\langle \tilde{f},w\rangle - \tilde{a}}{\|w\|_2^2}$. By assumption $|\mu| < |\tilde{\mu}|$ and according to (9) we have to consider three cases.

1. Let $|\tilde{\mu}| \le \lambda$. Then by assumption also $|\mu| < \lambda$ and we conclude by (9) that

$$\min_{x \in \mathbb{R}^d} E(x; f, a, w) = \frac{1}{2}\|w\|_2^2 \mu^2 < \frac{1}{2}\|w\|_2^2 \tilde{\mu}^2 = \min_{x \in \mathbb{R}^d} E(x; \tilde{f}, \tilde{a}, w).$$

2. Let $|\tilde{\mu}| > \lambda$ and $|\mu| \leq \lambda$. By (9) this implies

$$\min_{x\in\mathbb{R}^d} E(x; f, a, w) = \frac{1}{2}\|w\|_2^2\mu^2,$$
$$\min_{x\in\mathbb{R}^d} E(x; \tilde{f}, \tilde{a}, w) = \frac{1}{2}\|w\|_2^2\lambda^2 + \|w\|_2^2\lambda(|\tilde{\mu}| - \lambda).$$

Since $\|w\|^2\lambda(|\tilde{\mu}| - \lambda) > 0$ and $|\mu| \leq \lambda$ we obtain $\min_{x\in\mathbb{R}^d} E(x; f, a, w) < \min_{x\in\mathbb{R}^d} E(x; \tilde{f}, \tilde{a}, w)$.

3. Let $|\tilde{\mu}| > \lambda$ and $|\mu| > \lambda$. By (9) this implies

$$\begin{aligned}\min_{x\in\mathbb{R}^d} E(x; f, a, w) &= \frac{1}{2}\|w\|_2^2\lambda^2 + \|w\|_2^2\lambda(|\mu| - \lambda) \\ &< \frac{1}{2}\|w\|_2^2\lambda^2 + \|w\|_2^2\lambda(|\tilde{\mu}| - \lambda) = \min_{x\in\mathbb{R}^d} E(x; \tilde{f}, \tilde{a}, w)\end{aligned}$$

and we are done. □

Next we consider the case $p = 2$.

Lemma 3.4. *Let $0 \neq w \in \mathbb{R}^d$.*

i) Then, for $f \in \mathbb{R}^d$ and $a \in \mathbb{R}$, the minimizer $\hat{x}$ of

$$E(x; f, a, w) = \|f - x\|_2^2 + \lambda\big(\langle x, w\rangle - a\big)^2, \qquad \lambda > 0 \tag{12}$$

is given by

$$\hat{x} = f - \frac{\lambda(\langle f, w\rangle - a)}{1 + \lambda\|w\|_2^2}\, w$$

and the minimum by

$$E(\hat{x}; f, a, w) = \frac{\lambda}{1 + \lambda\|w\|_2^2}\big(\langle f, w\rangle - a\big)^2. \tag{13}$$

ii) If $\big(\langle f, w\rangle - a\big)^2 < \big(\langle \tilde{f}, w\rangle - \tilde{a}\big)^2$ for some $f, \tilde{f} \in \mathbb{R}^d$ and $a, \tilde{a} \in \mathbb{R}$, then

$$\min_{x\in\mathbb{R}^d} E(x; f, a, w) < \min_{x\in\mathbb{R}^d} E(x; \tilde{f}, \tilde{a}, w).$$

Proof. i) Setting the gradient of (12) to zero results in

$$\begin{aligned}2(x - f) + 2\lambda(\langle x, w\rangle - a)\, w &= 0, \\ (I + \lambda w w^{\mathrm{T}})x &= f + \lambda a w.\end{aligned}$$

Using the Sherman-Morrison formula [8, p. 129] it follows

$$\begin{aligned}\hat{x} &= \left(I - \frac{\lambda}{1 + \lambda\|w\|_2^2} w w^{\mathrm{T}}\right)(f + \lambda a w) \\ &= f - \frac{\lambda\langle f, w\rangle}{1 + \lambda\|w\|_2^2}\, w + \lambda a w - \frac{\lambda^2 a\|w\|_2^2}{1 + \lambda\|w\|_2^2}\, w \\ &= f - \frac{\lambda(\langle f, w\rangle - a)}{1 + \lambda\|w\|_2^2}\, w.\end{aligned}$$

For the corresponding energy we obtain by straightforward computation

$$\begin{aligned} E(\hat{x}; f, a, w) &= \|f - \hat{x}\|_2^2 + \lambda\big(\langle x, w\rangle - a\big)^2 \\ &= \frac{\lambda^2 \left(\langle f, w\rangle - a\right)^2}{\left(1 + \lambda \|w\|_2^2\right)^2} \|w\|_2^2 + \lambda \left[\langle f, w\rangle - \frac{\lambda \left(\langle f, w\rangle - a\right) \|w\|_2^2}{1 + \lambda \|w\|_2^2} - a\right]^2 \\ &= \frac{\lambda}{1 + \lambda \|w\|_2^2} \big(\langle f, w\rangle - a\big)^2. \end{aligned}$$

ii) follows directly from (13). □

3.2 Proximity of absolute cyclic differences of first and second order

Now we turn to $\mathbb{S}^1$-valued data represented by $f \in [-\pi, \pi)^d$. We are interested in the minimizers of

$$\mathcal{E}(x; f, w) := \frac{1}{2} \sum_{j=1}^{d} d(f_j, x_j)^2 + \lambda d(x; w)^p, \quad \lambda > 0 \tag{14}$$

on $[-\pi, \pi)^d$ for $p \in \{1, 2\}$ and $w \in \{b_1, b_2, b_{1,1}\}$. We start with the case $p = 1$.

Theorem 3.5. *For $w \in \{b_1, b_2, b_{1,1}\}$ set $s := \operatorname{sgn}(\langle f, w\rangle)_{2\pi}$. Let $p = 1$ and $f \in [-\pi, \pi)^d$, where d is adapted to the respective length of w.*

i) If $|(\langle f, w\rangle)_{2\pi}| < \pi$, then the unique minimizer of $\mathcal{E}(x; f, w)$ is given by

$$\hat{x} = (f - s\, m\, w)_{2\pi}, \qquad m := \min\left\{\lambda, \frac{|(\langle f, w\rangle)_{2\pi}|}{\|w\|_2^2}\right\}. \tag{15}$$

ii) If $|(\langle f, w\rangle)_{2\pi}| = \pi$, then $\mathcal{E}(x; f, w)$ has the two minimizers

$$\hat{x} = (f \mp s\, m\, w)_{2\pi}, \qquad m := \min\left\{\lambda, \frac{\pi}{\|w\|_2^2}\right\}.$$

Note that for $w = b_1$ case ii) appears exactly if f_1 and f_2 are antipodal points.

Proof. By (1) and Proposition 2.5 we can rewrite $\mathcal{E}$ in (14) as

$$\begin{aligned} \mathcal{E}(x; f, w) &= \frac{1}{2} \sum_{j=1}^{d} \min_{k_j \in \mathbb{Z}} |f_j - x_j - 2\pi k_j|^2 + \lambda \min_{\sigma \in \mathbb{Z}} |\langle x, w\rangle - 2\pi\sigma| \\ &= \min_{\substack{k \in \mathbb{Z}^d \\ \sigma \in \mathbb{Z}}} \frac{1}{2} \|f - x - 2\pi k\|_2^2 + \lambda |\langle x, w\rangle - 2\pi\sigma|, \end{aligned}$$

where $k = (k_j)_{j=1}^d$. Let

$$E_{k,\sigma}(x) := \frac{1}{2} \|f - x - 2\pi k\|_2^2 + \lambda |\langle x, w\rangle - 2\pi\sigma|.$$

We are looking for

$$\min_{x\in[-\pi,\pi)^d} \mathcal{E}(x;f,w) = \min_{x\in[-\pi,\pi)^d} \min_{\substack{k\in\mathbb{Z}^d\\ \sigma\in\mathbb{Z}}} E_{k,\sigma}(x) = \min_{\substack{k\in\mathbb{Z}^d\\ \sigma\in\mathbb{Z}}} \min_{x\in[-\pi,\pi]^d} E_{k,\sigma}(x), \tag{16}$$

where the last equality can be seen by the following argument: If for some k,σ the minimizer $\hat{x} := \arg\min_{x\in[-\pi,\pi]^d} E_{k,\sigma}(x)$ has components $\hat{x}_j = \pi$ for $j \in J \subseteq \mathbb{I}_d$, then we get using $\tilde{x} := \hat{x} - 2\pi \sum_{j\in J} e_j \in [-\pi,\pi)^d$, that

$$E_{k,\sigma}(\hat{x}) = \frac{1}{2}\|f - \tilde{x} - 2\pi \underbrace{(k - \sum_{j\in J} e_j)}_{\tilde{k}}\|_2^2 + \lambda|\langle \tilde{x}, w\rangle - 2\pi \underbrace{(\sigma - \langle \sum_{j\in J} e_j, w\rangle)}_{\tilde{\sigma}}| = E_{\tilde{k},\tilde{\sigma}}(\tilde{x}).$$

By Lemma 3.1 the minimizers over $\mathbb{R}^d$ of $E_{k,\sigma}(x)$ are given by

$$\hat{x}_{k,\sigma} = f - 2\pi k - s_{k,\sigma}\, m_{k,\sigma}\, w, \tag{17}$$

where

$$s_{k,\sigma} := \operatorname{sgn}(\nu_{k,\sigma}), \quad m_{k,\sigma} := \min\left\{\lambda, \frac{|\nu_{k,\sigma}|}{\|w\|_2^2}\right\} \quad \text{and} \quad \nu_{k,\sigma} := \langle f, w\rangle - 2\pi(\langle k, w\rangle + \sigma).$$

By Corollary 3.3 the minimum of $E_{k,\sigma}$ is determined by $|\nu_{k,\sigma}|$. Note that $|\langle f, w\rangle| < 2\pi$ for $w = b_1$ and $|\langle f, w\rangle| < 4\pi$ for $w \in \{b_2, b_{1,1}\}$. We distinguish two cases.

1. If $\langle f, w\rangle \in ((2r-1)\pi, (2r+1)\pi)$, $r \in \mathbb{Z}$ then $\nu_{k,\sigma}$ attains its smallest value exactly for $\langle k, w\rangle + \sigma = r$ and

$$\nu_{k,r-\langle k,w\rangle} = \langle f, w\rangle - 2\pi r = (\langle f, w\rangle)_{2\pi}.$$

By (17) we obtain

$$\hat{x}_{k,r-\langle k,w\rangle} = f - 2\pi k - s\, m\, w$$

with s, m as in (15). Corollary 3.3 implies that

$$E_{k,r-\langle k,w\rangle}(\hat{x}_{k,r-\langle k,w\rangle}) < E_{k,\sigma}(\hat{x}_{k,\sigma}) \le \min_{x\in[-\pi,\pi]^d} E_{k,\sigma}(x) \qquad \forall \sigma \in \mathbb{Z}\backslash\{r - \langle k, w\rangle\}.$$

Finally, there exists exactly one $k^* \in \mathbb{Z}^d$ such that $\hat{x}_{k^*,r-\langle k^*,w\rangle} \in [-\pi,\pi)^d$ and by (16) we conclude that

$$\hat{x} := \hat{x}_{k^*,r-\langle k^*,w\rangle} = f - 2\pi k^* - s\, m\, w = (f - s\, m\, w)_{2\pi}$$

is the unique minimizer of $\mathcal{E}(x;f,w)$ over $[-\pi,\pi)^d$.

2. If $\langle f, w\rangle = (2r-1)\pi$, $r \in \mathbb{Z}$, then $\nu_{k,\sigma}$ attains its smallest value exactly for $\langle k, w\rangle + \sigma \in \{r, r-1\}$ and by Corollary 3.3 the minimum of the corresponding functions $E_{k,\sigma}$ is smaller than those of the other functions in (16). We obtain

$$\nu_{k,r-\langle k,w\rangle} = -\pi, \quad \nu_{k,r-1-\langle k,w\rangle} = \pi$$

and

$$\hat{x}_{k,r-\langle k,w\rangle} = f - 2\pi k + m\, w, \quad \hat{x}_{k,r-1-\langle k,w\rangle} = f - 2\pi k - m\, w, \qquad m := \min\left\{\lambda, \frac{\pi}{\|w\|_2^2}\right\}.$$

As in part 1 of the proof we conclude that $\hat{x} = (f \pm m\, w)_{2\pi}$ are the minimizers of $\mathcal{E}(x; f, w)$ over $[-\pi, \pi)^d$. This finishes the proof. □

Next we focus on $p = 2$.

Theorem 3.6. *Let $p = 2$ in (14), $w \in \{b_1, b_2, b_{1,1}\}$ and $f \in [-\pi, \pi)^d$, where d is adapted to the respective length of w.*

i) If $|(\langle f, w\rangle)_{2\pi}| < \pi$, then the unique minimizer of $\mathcal{E}(x; f, w)$ is given by

$$\hat{x} = \left(f - \frac{\lambda(\langle f, w\rangle)_{2\pi}}{1 + \lambda \|w\|_2^2} w \right)_{2\pi}.$$

ii) If $|(\langle f, w\rangle)_{2\pi}| = \pi$, then $\mathcal{E}(x; f, w)$ has the two minimizers

$$\hat{x} = \left(f \mp \frac{\lambda\pi}{1 + \lambda \|w\|_2^2} w \right)_{2\pi}.$$

Proof. The proof follows the lines of the proof of Theorem 3.5 using Lemma 3.4. □

Finally, we need the proximal mapping $\operatorname{prox}_{\lambda d(f,\cdot)^2}$ for given $f \in (\mathbb{S}^1)^N$. The proximal mapping of the (squared) cyclic distance function was also computed (for more general manifolds) in [21]. Here we give an explicit expression for spherical data.

Proposition 3.7. *For $f, g \in [-\pi, \pi)^N$ let*

$$\mathcal{E}(x; g, f) := d(g, x)^2 + \lambda d(f, x)^2 = \sum_{j=1}^{N} d(g_j, x_j)^2 + \lambda d(f_j, x_j)^2.$$

Then the minimizer(s) of $\mathcal{E}(x; g, f)$ are given by

$$\hat{x} = \left(\frac{g + \lambda f}{1 + \lambda} + \frac{\lambda}{1 + \lambda} 2\pi\, v \right)_{2\pi},$$

where $v = (v_j)_{j=1}^N \in \mathbb{R}^N$ is defined by

$$v_j := \begin{cases} 0 & \text{if } |g_j - f_j| \le \pi, \\ \operatorname{sgn}(g_j - f_j) & \text{if } |g_j - f_j| > \pi \end{cases}$$

and the minimum is

$$\mathcal{E}(\hat{x}; g, f) = \frac{\lambda}{1 + \lambda} (g - f)_{2\pi}^2.$$

Proof. Obviously, the minimization of $\mathcal{E}$ can be done component wise so that we can restrict our attention to $N = 1$.

1. First we look at the minimization problem over $\mathbb{R}$ which reads

$$\min_{x \in \mathbb{R}} (g - x)^2 + \lambda (f - x)^2$$

and has the following minimizer and minimum:

$$\hat{x} = \frac{g + \lambda f}{1 + \lambda}, \qquad (g - \hat{x})^2 + \lambda (f - \hat{x})^2 = \frac{\lambda}{1 + \lambda} (g - f)^2.$$

2. For the original problem

$$\begin{aligned}\min_{x\in[-\pi,\pi)} \mathcal{E}(x;g,f) &= \min_{x\in[-\pi,\pi)} \Big\{ d(g,x)^2 + \lambda d(f,x)^2 \Big\} \\ &= \min_{x\in[-\pi,\pi)} \left\{ \min_{k\in\{0,\operatorname{sgn}(g)\}} (g-x-2\pi k)^2 + \min_{l\in\{0,\operatorname{sgn}(f)\}} \lambda (f-x-2\pi l)^2 \right\}\end{aligned}$$

we consider the related energy functionals on $\mathbb{R}$, namely

$$E_{k,l}(x;g,f) := (g-x-2\pi k)^2 + \lambda(f-x-2\pi l)^2, \quad k\in\{0,\operatorname{sgn} g\},\, l\in\{0,\operatorname{sgn} f\}.$$

By part 1 of the proof these functions have the minimizers

$$\hat{x}_{k,l} = \frac{(g-2\pi k)+\lambda(f-2\pi l)}{1+\lambda} = \frac{g+\lambda f - 2\pi(k+\lambda l)}{1+\lambda}$$

and

$$E_{k,l}(\hat{x}_{k,l};g,f) = \frac{\lambda}{1+\lambda}\left((g-2\pi k)-(f-2\pi l)\right)^2 = \frac{\lambda}{1+\lambda}\left(g-f-2\pi(k-l)\right)^2. \tag{18}$$

We distinguish three cases:

a) If $|g-f|<\pi$, then the minimum in (18) occurs exactly for $k=l$ and it holds

$$\hat{x}_{k,k} = \frac{g+\lambda f - 2\pi k(1+\lambda)}{1+\lambda} = \frac{g+\lambda f}{1+\lambda} - 2\pi k.$$

For $k=0$ we see that $\hat{x}_{0,0}\in[-\pi,\pi)$ and $\mathcal{E}(\hat{x};g,f) = \frac{\lambda}{1+\lambda}(g-f)^2$.

b) If $|g-f|>\pi$, then (18) has its minimum exactly for $k-l=\operatorname{sgn}(g-f)$ and

$$\hat{x}_{k,k-\operatorname{sgn}(g-f)} = \frac{g+\lambda f - 2\pi(k+\lambda(k-\operatorname{sgn}(g-f)))}{1+\lambda} = \frac{g+\lambda(f+\operatorname{sgn}(g-f)2\pi)}{1+\lambda} - 2\pi k$$

which is in $[-\pi,\pi)$ for $k=0$ or $k=\operatorname{sgn}(g)$ and

$$\mathcal{E}(\hat{x};g,f) = \frac{\lambda}{1+\lambda}(g-f-\operatorname{sgn}(g-f)2\pi)^2.$$

c) In the case $|g-f|=\pi$ the minimum in (18) is attained for $k-l=0,\pm1$ so that we have both solutions from i) and ii). This completes the proof. □

Remark 3.8. *The analytic expressions for the proximal mapping of the absolute cyclic differences in this subsection are based on the simple expressions for the corresponding real valued proximal mappings in the previous subsection. Such simple expression are only available for linear differences $|\langle f,w\rangle|$ and not, e.g., for $\sqrt{f_1^2+f_2^2}$. Therefore we cannot provide a cyclic equivalent of the isotropic TV regularizer, but can only deal with its "anisotropic" variant. Clearly, (weighted) diagonal directions can be involved to make the model "more rotationally invariant".*

4 Cyclic proximal point method

The proximal point algorithm (PPA) on the Euclidean space goes back to [49]. Recently this algorithm was extended to Riemannian manifolds of non-positive sectional curvature [21] and also to Hadamard spaces [1]. A cyclic version of the proximal point algorithm (CPPA) on the Euclidean space was given in [5], see also the survey [4]. A CPPA for Hadamard spaces can be found in [2]. In the CPPA the original function J is split into a sum $J = \sum_l J_l$ and, iteratively, the proximal mappings of the functions J_l are applied in a cyclic way. The great advantage of this method is that often the proximal mappings of the summands J_l are much easier to compute or can even be given in a closed form. In the following we develop a CPPA for functionals of $\mathbb{S}^1$-valued signals and images containing absolute cyclic first and second order differences.

4.1 One-dimensional data

First we have a look at the one-dimensional case, i.e., at signals. For given $\mathbb{S}^1$-valued signals represented by $f = (f_i)_{i=1}^N \in [-\pi,\pi)^N$, $N \in \mathbb{N}$, and regularization parameters $\alpha, \beta \geq 0$, $\max\{\alpha,\beta\} \neq 0$, we are interested in

$$\operatorname*{arg\,min}_{x\in[-\pi,\pi)^N} J(x), \quad J(x) = J(x,f) := F(x;f) + \alpha\,\mathrm{TV}_1(x) + \beta\,\mathrm{TV}_2(x), \tag{19}$$

where

$$F(x;f) := \frac{1}{2}\sum_{i=1}^{N} d(f_i,x_i)^2,$$
$$\mathrm{TV}_1(x) := \sum_{i=1}^{N-1} d(x_i,x_{i+1}), \quad \mathrm{TV}_2(x) := \sum_{i=2}^{N-1} d_2(x_{i-1},x_i,x_{i+1}).$$

To apply a CPPA we set $J_1(x) := F(x;f)$, split $\alpha\,\mathrm{TV}_1$ into an even and an odd part

$$\alpha\,\mathrm{TV}_1(x) = \sum_{\nu=0}^{1} \alpha \sum_{i=1}^{\left\lfloor \frac{N-1}{2} \right\rfloor} d(x_{2i-1+\nu}, x_{2i-\nu}) =: \sum_{\nu=0}^{1} J_{2+\nu}(x)$$

and $\beta\,\mathrm{TV}_2$ into three sums

$$\beta\,\mathrm{TV}_2(x) = \sum_{\nu=0}^{2} \beta \sum_{i=1}^{\left\lfloor \frac{N-1}{3} \right\rfloor} d_2(x_{3i-2+\nu}, x_{3i-1+\nu}, x_{3i+\nu}) =: \sum_{\nu=0}^{2} J_{4+\nu}(x)$$

Then the objective function decomposes as

$$J = \sum_{l=1}^{6} J_l.$$

We compute in the k-th cycle of the CPPA the signal

$$x^{(k)} := \operatorname{prox}_{\lambda_k J_6}\left(\operatorname{prox}_{\lambda_k J_5} \dots \left(\operatorname{prox}_{\lambda_k J_1}(x^{(k-1)})\right)\right).$$

The different proximal values can be obtained as follows:

Algorithm 1 CPPA for minimizing (19) or (21) for cyclic data

Input $\{\lambda_k\}_k$ fulfilling (20) and α, β or $\alpha = (\alpha_1, \alpha_2)$, $\beta = (\beta_1, \beta_2)$, γ
data $f \in [-\pi, \pi)^N$ or $f \in [-\pi, \pi)^{N \times M}$

function CPPA(α, β, λ_0, f)
  Initialize $x^{(0)} = f$, $k = 0$
  Initialize the cycle length as $c = 6$ (1D) or $c = 15$ (2D)
  repeat
    for $l \leftarrow 1$ **to** c **do**
      $x^{(k+\frac{l}{c})} \leftarrow \operatorname{prox}_{\lambda_k J_l}(x^{(k+\frac{l-1}{c})})$
    $k \leftarrow k + 1$
  until a convergence criterion are reached
  return $x^{(k)}$

i) By Proposition 3.7 with $x^{(k-1)}$ playing the role of g we get

$$x^{(k-1+\frac{1}{6})} := \operatorname{prox}_{\lambda_k J_1}(x^{(k-1)}).$$

ii) For $\nu = 0, 1$, we obtain the vectors

$$x^{(k-1+\frac{\nu+2}{6})} := \operatorname{prox}_{\lambda_k J_{2+\nu}}\left(x^{(k-1+\frac{\nu+1}{6})}\right)$$

by applying Theorem 3.5 with $w = b_1$ independently for the pairs $(x_{2i-1+\nu}, x_{2i+\nu})$, $i = 1, \ldots, \left\lfloor \frac{N-1}{2} \right\rfloor$.

iii) For $\nu = 0, 1, 2$, we compute

$$x^{(k-1+\frac{\nu+4}{6})} := \operatorname{prox}_{\lambda_k J_{4+\nu}}\left(x^{(k-1+\frac{\nu+3}{6})}\right)$$

by applying Theorem 3.5 with $w = b_2$ independently for the vectors $(x_{3i-2+\nu}, x_{3i-1+\nu}, x_{3i+\nu})$, $i = 1, \ldots, \left\lfloor \frac{N-1}{3} \right\rfloor$.

The parameter sequence $\{\lambda_k\}_k$ of the algorithm should fulfill

$$\sum_{k=0}^{\infty} \lambda_k = \infty, \quad \text{and} \quad \sum_{k=0}^{\infty} \lambda_k^2 < \infty. \tag{20}$$

This property is also essential for proving the convergence of the CPPA for real-valued data and data on a Hadamard manifold, see [2, 5]. In our numerical experiments we choose $\lambda_k := \lambda_0/k$ with some initial parameter $\lambda_0 > 0$ which clearly fulfill (20). The whole procedure is summarized in Algorithm 1.

4.2 Two-dimensional data

Next we consider two-dimensional data, i.e., images of the form $f := \left(f_{i,j}\right)_{i,j=1}^{N,M} \in [-\pi, \pi)^{N \times M}$, $N, M \in \mathbb{N}$. Our functional includes *h*orizontal and *v*ertical cyclic first and second order

differences d_1 and d_2 and $d$iagonal (mixed) differences $d_{1,1}$. For non-negative regularization parameters $\alpha := (\alpha_1, \alpha_2)$, $\beta := (\beta_1, \beta_2)$ and γ not all equal to zero we are looking for

$$\operatorname*{arg\,min}_{x\in[-\pi,\pi)^{N\times M}} J(x), \quad J(x) = J(x,f) := F(x;f) + \alpha\,\mathrm{TV}_1(x) + \beta\,\mathrm{TV}_2^{\mathrm{hv}}(x) + \gamma\,\mathrm{TV}_2^{\mathrm{d}}(x), \tag{21}$$

where

$$\begin{aligned}
F(x;f) &:= \frac{1}{2}\sum_{i,j=1}^{N,M} d(f_{i,j}, x_{i,j})^2,\\
\alpha\,\mathrm{TV}_1(x) &:= \alpha_1 \sum_{i,j=1}^{N-1,M} d(x_{i,j}, x_{i+1,j}) + \alpha_2 \sum_{i,j=1}^{N,M-1} d(x_{i,j}, x_{i,j+1}),\\
\beta\,\mathrm{TV}_2^{\mathrm{hv}}(x) &:= \beta_1 \sum_{i=1,j=2}^{N-1,M} d_2(x_{i-1,j}, x_{i,j}, x_{i+1,j}) + \beta_2 \sum_{i=2,j=1}^{N,M-1} d_2(x_{i,j-1}, x_{i,j}, x_{i,j+1}),\\
\gamma\,\mathrm{TV}_2^{\mathrm{d}}(x) &:= \gamma \sum_{i,j=1}^{N-1,M-1} d_{1,1}(x_{i,j}, x_{i+1,j}, x_{i,j+1}, x_{i+1,j+1}).
\end{aligned}$$

Here the objective function splits as

$$J = \sum_{l=1}^{15} J_l \tag{22}$$

with the following summands: Again we set $J_1 := F(x;f)$ and compute the proximal value of $\lambda_k J_1$ by Proposition 3.7. Each of the sums in TV_1 and $\mathrm{TV}_2^{\mathrm{hv}}$ can be split analogously as in the one-dimensional case, where we have to consider row and column vectors now. This results in $2(2+3) = 10$ functions $J_2, \dots, J_{11}$ whose proximal values can be computed by Theorem 3.5. Finally, we split $\mathrm{TV}_2^{\mathrm{d}}$. into the four sums

$$\gamma\,\mathrm{TV}_2^{\mathrm{d}}(x) = \sum_{\mu,\nu=0}^{1} \gamma \sum_{i,j=1}^{\left\lfloor\frac{N-1}{2}\right\rfloor,\left\lfloor\frac{M-1}{2}\right\rfloor} d_{1,1}(x_{2i-1+\mu,2j-1+\nu}, x_{2i+\mu,2j-1+\nu}, x_{2i-1+\mu,2j+\nu}, x_{2i+\mu,2j+\nu})$$

and denote the inner sums by $J_{12}, \dots, J_{15}$. Clearly, the proximal values of the functions $\lambda_k J_l$, $l = 12, \dots, 15$ can be computed separately for the vectors

$$(x_{2i-1+k,2j-1+l}, x_{2i+k,2j-1+l}, x_{2i-1+k,2j+l}, x_{2i+k,2j+l}),\ i = 1, \dots \left\lfloor\frac{N-1}{2}\right\rfloor,\ j = 1, \dots, \left\lfloor\frac{M-1}{2}\right\rfloor$$

by Theorem 3.5 with $w = b_{1,1}$. In summary, the computation can be done by Algorithm 1. Note that the presented approach immediately generalizes to arbitrary dimensions.

4.3 Convergence

Since $\mathbb{S}^1$ is not a Hadamard space, the convergence analysis of the CPPA in [2] cannot be applied. We show the convergence of the CPPA for the 2D $\mathbb{S}^1$-valued function (21) under certain conditions. The 1D setting in (19) can then be considered as a special case. In the following, let $\mathbb{I} := \{1, \dots, N\} \times \{1, \dots, M\}$.

Our first condition is that the data $f \in (\mathbb{S}^1)^{N\times M}$ is dense enough, this means that the distance between neighboring pixels

$$d_\infty(f) := \max_{(i,j)\in\mathbb{I}} \max_{(k,l)\in N_{i,j}} d(f_{i,j}, f_{k,l}), \qquad \mathcal{N}_{i,j} := \big\{(k,l) \in \mathbb{I} : \, |i-k| + |l-j| = 1\big\}$$

is sufficiently small. Similar conditions also appear in the convergence analysis of nonlinear subdivision schemes for manifold-valued data in [59, 61]. In the context of nonlinear subdivision schemes, even more severe restrictions such as 'almost equally spaced data' are frequently required [31]. This imposes additional conditions on the second order differences to make the data almost lie on a 'line'. Our analysis requires only bounds on the first, but not on the second order differences.

Our next requirement is that the regularization parameters α, β, γ in (21) are sufficiently small. For large parameters any solution tends to become almost constant. In this case, if the data is for example equidistantly distributed on the circle, e.g., $f_i = 2\pi i/N$ in 1D, any $2\pi j/N$ shift is again a solution. In this situation the model loses its interpretation which is an inherent problem due to the cyclic structure of the data.

Finally, the parameter sequence $\{\lambda_k\}_k$ of the CPPA has to fulfill (20) with a small ℓ^2 norm. The later can be achieved by rescaling.

Our convergence analysis is based on a convergence result in [2] and an unwrapping procedure. We start by reformulating the convergence result for the CPPA of real-valued data, which is a special case of [2] and can also be derived from [4].

Theorem 4.1. *Let $E = \sum_{l=1}^{c} E_l$, where E_l, $l = 1, \dots, c$, are proper, closed, convex functionals on $\mathbb{R}^{N\times M}$. Let E have a global minimizer. Assume that there exists $L > 0$ such that the iterates $\{x^{(k+\frac{l}{c})}\}$ of the CPPA (see Algorithm 1) satisfy*

$$E_l(x^{(k)}) - E_l(x^{(k+\frac{l}{c})}) \le L\|x^{(k)} - x^{(k+\frac{l}{c})}\|_2, \quad l = 1, \dots, c,$$

for all $k \in \mathbb{N}_0$. Then the sequence $\{x^{(k)}\}_k$ converges to a minimizer of E. Moreover the iterates fulfill

$$\begin{aligned}
&\|x^{(k+\frac{l-1}{c})} - x^{(k+\frac{l}{c})}\|_2 \le 2\lambda_k L,\\
&\|x^{(k+1)} - x\|_2^2 \le \|x^{(k)} - x\|_2^2 - 2\lambda_k[E(x^{(k)}) - E(x)] + 2\lambda_k^2 L^2 c(c+1) \quad \text{for all } x \in \mathbb{R}^{N\times M}.
\end{aligned}$$

The next lemma states a discrete analogue of a well-known result on unwrapping or lifting from algebraic topology. We supply a short proof since we did not found it in the literature.

Lemma 4.2. *Let $x \in (\mathbb{S}^1)^{N\times M}$ with $d_\infty(x) < \frac{\pi}{2}$. For $q \in \mathbb{S}^1$ not antipodal to $x_{1,1}$ fix an $\tilde{x}_{1,1} \in \mathbb{R}$ such that $\exp_q(\tilde{x}_{1,1}) = x_{1,1}$. Then there exists a unique $\tilde{x} \in \mathbb{R}^{N\times M}$ such that for all $(i,j) \in \mathbb{I}$ the following relations are fulfilled:*

i) $\exp_q(\tilde{x}_{i,j}) = x_{i,j}$,

ii) $d(x_{i,j}, x_{k,l}) = |\tilde{x}_{i,j} - \tilde{x}_{k,l}|, \quad (k,l) \in \mathcal{N}_{i,j}$.

We call $\tilde{x}$ the lifted *or* unwrapped *image of x (w.r.t. a fixed $\tilde{x}_{1,1}$).*

Proof. For $x_{k,l}$, $(k,l) \in \mathcal{N}_{1,1}$, it holds by assumption on $d_\infty(x)$ that $d(x_{1,1}, x_{k,l}) < \frac{\pi}{2}$. Hence we have $s_{k,l} := (x_{k,l} - x_{1,1})_{2\pi} \in \big(-\frac{\pi}{2}, \frac{\pi}{2}\big)$, where with an abuse of notation $x_{k,l}$ stands for an

arbitrary representative in $T_q\mathbb{S}^1$ of $x_{k,l}$. Then obviously $\tilde{x}_{k,l} := \tilde{x}_{1,1} + \operatorname{sgn}(s_{k,l})d(x_{1,1}, x_{k,l})$, $(k,l) \in N_{1,1}$ are the unique values satisfying i) and ii).
For $x_{2,2} \in \mathcal{N}_{2,1} \cap \mathcal{N}_{1,2}$ consider

$$\begin{aligned}
\tilde{x}_{2,2} &:= \tilde{x}_{1,2} + \operatorname{sgn}\big((x_{2,2} - x_{1,2})_{2\pi}\big)d(x_{1,2}, x_{2,2}) \\
&= \tilde{x}_{1,1} + \operatorname{sgn}(s_{1,2})d(x_{1,1}, x_{1,2}) + \operatorname{sgn}\big((x_{2,2} - x_{1,2})_{2\pi}\big)d(x_{1,2}, x_{2,2}), \\
\tilde{y}_{2,2} &:= \tilde{x}_{2,1} + \operatorname{sgn}\big((x_{2,2} - x_{2,1})_{2\pi}\big)d(x_{2,1}, x_{2,2}) \\
&= \tilde{x}_{1,1} + \operatorname{sgn}(s_{2,1})d(x_{1,1}, x_{2,1}) + \operatorname{sgn}\big((x_{2,2} - x_{2,1})_{2\pi}\big)d(x_{2,1}, x_{2,2}).
\end{aligned}$$

By assumption on $d_\infty(x)$ we see that $|\tilde{x}_{2,2} - \tilde{y}_{2,2}| < 2\pi$ so that $\tilde{x}_{2,2} = \tilde{y}_{2,2}$. Thus $\tilde{x}_{2,2}$ is the unique value with properties i) and ii).
Proceeding this scheme successively, we obtain the whole unique image $\tilde{x}$ fulfilling i) and ii). □

For $\delta \in (0, \pi)$ we define

$$\mathcal{S}(f, \delta) := \big\{x \in (\mathbb{S}^1)^{N\times M} : d_\infty(x, f) \le \delta\big\},$$

where

$$d_\infty(x, f) := \max_{(i,j)\in\mathbb{I}} d(x_{i,j}, f_{i,j}),$$

to measure how 'near' the images f and x are to each other.

Lemma 4.3. *Let $f \in (\mathbb{S}^1)^{N\times M}$ with $d_\infty(f) < \frac{\pi}{8}$ and $q \in \mathbb{S}^1$ be not antipodal to $f_{1,1}$. Fix $\tilde{f}_{1,1}$ with $\exp_q(\tilde{f}_{1,1}) = f_{1,1}$ and let $\tilde{f}$ be the corresponding lifting of f. Let $\delta \in (0, \frac{\pi}{8}]$.*

i) Then every $x \in \mathcal{S}(f, \delta)$ has a unique lifting $\tilde{x}$ w.r.t. to the base point q with $|\tilde{x}_{1,1} - \tilde{f}_{1,1}| \le \frac{\pi}{8}$.

ii) For J defined by (21), let $\tilde{J}$ denote its analog for real-valued data, i.e.,

$$\tilde{J}(x) = \tilde{J}(x, \tilde{f}) := \widetilde{F}(x; \tilde{f}) + \alpha\widetilde{\mathrm{TV}}_1(x) + \beta\widetilde{\mathrm{TV}}_2^{\mathrm{hv}}(x) + \gamma\widetilde{\mathrm{TV}}_2^{\mathrm{d}}(x), \tag{23}$$

where the cyclic distances in F and in the TV *terms are replaced by absolute differences in $\tilde{F}$ and $\widetilde{\mathrm{TV}}$. Then it holds*

$$J(x) = \tilde{J}(\tilde{x}) \quad \textit{for all} \quad x \in \mathcal{S}(f, \delta). \tag{24}$$

Proof. By definition of $\mathcal{S}(f, \delta)$ and assumption on f we have for any $x \in \mathcal{S}(f, \delta)$ that

$$d(x_{i,j}, x_{k,l}) \le d(x_{i,j}, f_{i,j}) + d(f_{i,j}, f_{k,l}) + d(f_{k,l}, x_{k,l}) < \frac{3\pi}{8}, \quad (k,l) \in \mathcal{N}_{i,j},$$

and hence $d_\infty(x) < \frac{3\pi}{8}$. Further it holds $d(x_{1,1}, f_{1,1}) < \frac{\pi}{8}$. Consequently, every $x \in \mathcal{S}(f, \delta)$ has a unique lifting $\tilde{x}$ by Lemma 4.2 w.r.t. to the base point q fulfilling $|\tilde{x}_{1,1} - \tilde{f}_{1,1}| \le \frac{\pi}{8}$.
To see (24) we show the equality for the involved summands in J and $\tilde{J}$ separately.
First we consider TV_1. By properties of the lifting in Lemma 4.2 we have $d(x_{i,j}, x_{i,j+1}) = |\tilde{x}_{i,j} - \tilde{x}_{i,j+1}|$ and $d(x_{i,j}, x_{i+1,j}) = |\tilde{x}_{i,j} - \tilde{x}_{i+1,j}|$. By the definition of TV_1 and $\widetilde{\mathrm{TV}}_1$, this implies $\mathrm{TV}_1(x) = \widetilde{\mathrm{TV}}_1(\tilde{x})$.
Next we consider $\mathrm{TV}_2^{\mathrm{hv}}$. The corresponding second order differences are given by the expressions $d_2(x_{i-1,j}, x_{i,j}, x_{i+1,j})$ and $d_2(x_{i,j-1}, x_{i,j}, x_{i,j+1})$, respectively. We exemplarily consider the first term. Since

$$d(x_{i-1,j}, x_{i+1,j}) \le d(x_{i-1,j}, f_{i-1,j}) + d(f_{i-1,j}, f_{i+1,j}) + d(f_{i+1,j}, x_{i+1,j}) < \frac{\pi}{8} + \frac{\pi}{4} + \frac{\pi}{8} = \frac{\pi}{2}$$

the distance between any two members of the triple is smaller than $\frac{\pi}{2}$. Due to the properties of the lifting $\tilde{x}$ this implies $|\Delta(\tilde{x}_{i-1,i}, \tilde{x}_{i,j}, \tilde{x}_{i+1,j}; b_2)| < \pi$. Then we conclude by Proposition 2.5 that $\mathrm{TV}_2^{\mathrm{hv}}(x) = \widetilde{\mathrm{TV}}_2^{\mathrm{hv}}(\tilde{x})$. Similarly it follows that $\mathrm{TV}_2^{\mathrm{d}}(x) = \widetilde{\mathrm{TV}}_2^{\mathrm{d}}(\tilde{x})$.
Concerning the data term $F(x; f)$ we consider $e_{i,j} := d(x_{i,j}, f_{i,j})$ and $\tilde{e}_{i,j} := |\tilde{x}_{i,j} - \tilde{f}_{i,j}|$. By definition of $\mathcal{S}(f, \delta)$ we have $e_{i,j} \leq \delta = \frac{\pi}{8}$ and by construction of $\tilde{f}$ and $\tilde{x}$ that $\tilde{e}_{i,j} = e_{i,j} + 2\pi k_{i,j}$, $k_{i,j} \in \mathbb{N}$ and $k_{1,1} = 0$. Furthermore it holds $|\tilde{e}_{i,j+1} - \tilde{e}_{i,j}| = \big||\tilde{x}_{i,j+1} - \tilde{f}_{i,j+1}| - |\tilde{x}_{i,j} - \tilde{f}_{i,j}|\big| \leq 2\delta$. If $k_{i,j} \neq k_{i,j+1}$, then there exists $k \in \mathbb{Z}\backslash\{0\}$ such that

$$|\tilde{e}_{i,j+1} - \tilde{e}_{i,j}| = |\tilde{e}_{i,j+1} - \tilde{e}_{i,j} + 2\pi k| \geq 2\pi - 2\delta > 2\delta$$

which is a contradiction. Thus $k_{i,j} = k_{i,j+1}$. Similarly we conclude $k_{i,j} = k_{i+1,j}$. In summary we obtain $k_{i,j} = k_{1,1} = 0$ for all $(i, j) \in \mathbb{I}$ which implies $e_{i,j} = \tilde{e}_{i,j}$. This finishes the proof. □

Remark 4.4. *The set $\mathcal{S}(f, \delta)$ is a convex subset of $(\mathbb{S}^1)^{N\times M}$ which means that for $x, y \in \mathcal{S}(f, \delta)$ and $t \in [0, 1]$ we have $[x, y]_t \in \mathcal{S}(f, \delta)$. Here $[x, y]_t$ denotes the point reached after time t on the unit speed geodesic starting at x in direction of y. Recall that a function φ is convex on $\mathcal{S}(f, \delta)$ if for all $x, y \in \mathcal{S}(f, \delta)$ and all $\lambda \in [0, 1]$ the relation $\varphi([x, y]_t) \leq t\varphi(x) + (1-t)\varphi(y)$ holds true. Let $f \in (\mathbb{S}^1)^{N\times M}$ with $d_\infty(f) < \frac{\pi}{8}$ and $\delta \in (0, \frac{\pi}{8}]$. Then we conclude by Lemma 4.3, since $\tilde{J}$ is convex, that J is convex on $\mathcal{S}(f, \delta)$.*

Lemma 4.5. *Let $f \in (\mathbb{S}^1)^{N\times M}$ and $m := \max\{\alpha_1, \alpha_2, \beta_1, \beta_2, \gamma\} > 0$. Let $\varepsilon > 0$ such that*

$$\mathrm{TV}_1(f) + \mathrm{TV}_2^{\mathrm{hv}}(f) + \mathrm{TV}_2^{\mathrm{d}}(f) \leq \frac{\varepsilon^2}{m}. \tag{25}$$

Then any minimizer x^* of J in (21) fulfills

$$d_\infty(x^*, f) \leq \left(\sum_{i,j=1}^{N,M} d(x^*_{i,j}, f_{i,j})^2\right)^{\frac{1}{2}} \leq \varepsilon.$$

Proof. Any minimizer x^* of (21) satisfies

$$J(x^*) \leq J(f) \leq m\left(\mathrm{TV}_1(f) + \mathrm{TV}_2^{\mathrm{hv}}(f) + \mathrm{TV}_2^{\mathrm{d}}(f)\right).$$

As a consequence we obtain

$$d_\infty(x^*, f)^2 \leq \sum_{i,j=1}^{N,M} d(x^*_{i,j}, f_{i,j})^2 \leq m\big(\mathrm{TV}_1(f) + \mathrm{TV}_2^{\mathrm{hv}}(f) + \mathrm{TV}_2^{\mathrm{d}}(f)\big) \leq \varepsilon^2. \qquad \square$$

Remark 4.6. *Lemma 4.5 holds also true for real-valued data and $\tilde{J}$ in (23).*

Now we combine Lemma 4.5 and 4.3 to locate the minimizers of J and $\tilde{J}$.

Lemma 4.7. *Let $f \in (\mathbb{S}^1)^{N\times M}$ with $d_\infty(f) < \frac{\pi}{8}$ and $0 < \varepsilon < \delta \leq \frac{\pi}{8}$ be given. Choose the parameters α, β, γ of J in (21) such that (25) with ε holds true. Then any minimizer x^* of J lies in $\mathcal{S}(f, \delta)$. Furthermore, if $\tilde{f}$ is the unique lifting of f w.r.t. a base point q and fixed $\tilde{f}_{1,1}$ with $\exp_q(\tilde{f}_{1,1}) = f_{1,1}$, then each minimizer y^* of $\tilde{J}$ defines a minimizer $x^* := \exp_q(y^*)$ of J. Conversely, the uniquely defined lifting $\tilde{x}^*$ of a minimizer x^* of J is a minimizer of $\tilde{J}$.*

Proof. By Lemma 4.5 we obtain $d_\infty(x^*, f) \le \varepsilon < \frac{\pi}{8}$ so that $x^* \in \mathcal{S}(f, \delta)$.
In order to show the second statement note that the mapping $x \mapsto \tilde{x}$ is a bijection from $\mathcal{S}(f, \delta)$ to the set $\widetilde{\mathcal{S}}(f, \delta)$ defined by

$$\widetilde{\mathcal{S}}(f, \delta) := \bigtimes_{(i,j)\in\mathbb{I}} \big[\tilde{f}_{i,j} - \delta, \tilde{f}_{i,j} + \delta\big].$$

If y^* minimizes $\tilde{J}$, then it lies in $\widetilde{\mathcal{S}}(f, \delta)$ which follows by Remark 4.6. By (24) and the minimizing property of y^* we obtain for any $x \in \mathcal{S}(f, \delta)$ that

$$J(\exp_q(y^*)) = \tilde{J}(y^*) \le \tilde{J}(\tilde{x}) = J(x).$$

As a consequence, $\exp_q(y^*)$ is a minimizer of J on $\mathcal{S}(f, \delta)$. By Lemma 4.5 all the minimizers of J are contained in $\mathcal{S}(f, \delta)$ so that $\exp_q(y^*)$ is a minimizer of J on $(\mathbb{S}^1)^{N\times M}$.
We proceed with the last statement. Let x^* be a minimizer of J with lifting $\tilde{x}^*$. Then we get for any $\tilde{y} \in \widetilde{S}(f, \delta)$ that

$$\tilde{J}(\tilde{x}^*) = J(x^*) \le J(\exp_q(\tilde{y})) = \tilde{J}(\tilde{y}).$$

This shows that $\tilde{x}^*$ is a minimizer of $\tilde{J}$ on $\widetilde{\mathcal{S}}(f, \delta)$. Since by Remark 4.6 all minimizers of $\tilde{J}$ lie in $\widetilde{\mathcal{S}}(f, \delta)$, the last assertion follows. □

Next we locate the iterates of the CPPA for real-valued data on a ball whose radius can be controlled.

Lemma 4.8. *For $f \in \mathbb{R}^{N\times M}$ and $\lambda := \{\lambda_k\}_k$ with property (20), let $\{x^{(k+\frac{l}{c})}\}$ be the sequence produced by Algorithm 1 for $\tilde{J}$. Assume that $\|f - x^{(k+\frac{l}{c})}\|_\infty \le \pi$. Let $x^* \in \mathbb{R}^{N\times M}$ be the minimizer of $\tilde{J}$. Then, for $k \in \mathbb{N}_0$ and $l \in \{1, \dots, c\}$, it holds*

$$\|x^{(k+\frac{l}{c})} - x^*\|_2 \le R := \sqrt{\|f - x^*\|_2^2 + 2\|\lambda\|_2^2 L^2 c(c+1)} + 2\|\lambda\|_\infty cL, \tag{26}$$

where $c = 15$ denotes the number of inner iterations and $L = 4$.

The assumption on the distances $|f_{i,j} - x_{i,j}^{(k+\frac{l}{c})}|$, $(i, j) \in \mathbb{I}$, to be smaller than π is automatically fulfilled for any unwrapping of $\mathbb{S}^1$-valued data.

Proof. By Theorem 4.1 we know that

$$\|x^{(k+1)} - x\|_2^2 \le \|x^{(k)} - x\|_2^2 - 2\lambda_k[\tilde{J}(x^{(k)}) - \tilde{J}(x)] + 2\lambda_k^2 L^2 c(c+1). \tag{27}$$

As a constant L we can choose the maximum of the Lipschitz constants of the involved summands. For $\widetilde{\mathrm{TV}}_1$, $\widetilde{\mathrm{TV}}_2^{\mathrm{hv}}$ and $\widetilde{\mathrm{TV}}_2^{\mathrm{d}}$ the Lipschitz constants are 1, 4, and 4, respectively. For the quadratic data term we have

$$\frac{1}{2}\big||f_{i,j} - x_{i,j}|^2 - |f_{i,j} - y_{i,j}|^2\big| \le \frac{1}{2}|2f_{i,j} - x_{i,j} - y_{i,j}||x_{i,j} - y_{i,j}| \le \pi|x_{i,j} - y_{i,j}|.$$

Therefore, we can set $L = 4$. Plugging in the minimizer $x = x^*$ into (27) and using $x^{(0)} = f$ yields

$$\begin{aligned}
\|x^{(k+1)} - x^*\|_2^2 &\le \|x^{(k)} - x^*\|_2^2 + 2\lambda_k^2 L^2 c(c+1) \\
&\le \|x^{(0)} - x^*\|_2^2 + 2\sum_{j=0}^{k} \lambda_j^2 L^2 c(c+1) \\
&\le \|f - x^*\|_2^2 + 2\|\lambda\|_2^2 L^2 c(c+1).
\end{aligned} \tag{28}$$

By Theorem 4.1 it holds

$$\|x^{(k+\frac{l}{c})} - x^{(k+\frac{l-1}{c})}\|_2 \leq 2\lambda_k L. \tag{29}$$

Using the triangle inequality we obtain

$$\|x^{(k+\frac{l}{c})} - x^*\|_2 \leq \|x^{(k+\frac{l}{c})} - x^{(k+\frac{l-1}{c})}\|_2 + \ldots + \|x^{(k+\frac{1}{c})} - x^{(k)}\|_2 + \|x^{(k)} - x^*\|_2^2,$$

which implies the assertion by (28) and (29). □

Now we compare the proximal mappings acting on data with values in $\mathbb{S}^1$ and $\mathbb{R}$.

Lemma 4.9. *For $f \in (\mathbb{S}^1)^{N\times M}$ with $d_\infty(f) < \frac{\pi}{8}$, let J be defined by (21) with the splitting (22). Let $\tilde{f}$ be the unique lifting of f w.r.t. a base point q not antipodal to $f_{1,1}$ and fixed $\tilde{f}_{1,1}$ with $\exp_q(\tilde{f}_{1,1}) = f_{1,1}$. Further, denote by $\tilde{J}$ the functional (23) corresponding to J. Then, for any $x \in \mathcal{S}(f,\delta)$, $\delta \in (0, \frac{\pi}{8}]$ and its lifting $\tilde{x}$ w.r.t. q, we have*

$$\operatorname{prox}_{\lambda J_l}(x) = \exp_q(\operatorname{prox}_{\lambda \tilde{J}_l}(\tilde{x})), \quad l \in \{1,\ldots,15\}, \tag{30}$$

i.e., the canonical projection $\exp_q$ commutes with the proximal mappings.

Proof. The function J_1 is based on the distance to the data f. Since $x \in \mathcal{S}(f,\delta)$, we have $d(x_{i,j}, f_{i,j}) \leq \frac{\pi}{8}$ for all $(i,j) \in \mathbb{I}$. The components of the proximal mapping $\operatorname{prox}_{\lambda J_1}$ are given by Proposition 3.7 from which we conclude (30) for $l = 1$.
The proximal mappings of J_l, $l = 2,\ldots,15$, are given via proximal mappings of the first and second order cyclic differences. We consider the first order difference $d_1 = d$. By the triangle inequality, we have $d(x_{i,j}, x_{i,j+1}) \leq \frac{3\pi}{8}$ as well as $d(x_{i,j}, x_{i+1,j}) \leq \frac{3\pi}{8}$. By the explicit form of the proximal mapping in Theorem 3.5 we obtain (30) for J_l, $l = 2,\ldots,5$.
Next we consider the horizontal and vertical second order differences $d_2(x_{i-1,j}, x_{i,j}, x_{i+1,j})$ and $d_2(x_{i,j-1}, x_{i,j}, x_{i,j+1})$. We have that $d(x_{i,j-1}, x_{i,j}) < \frac{3\pi}{8}$, $d(x_{i,j}, x_{i,j+1}) < \frac{3\pi}{8}$ as well as $d(x_{i,j-1}, x_{i,j}) < \frac{\pi}{2}$. Hence all contributing values of x lie on a quarter of the circle. Applying the proximal mapping in Theorem 3.5 the resulting data lie on one half of the circle. An analogous statement holds true for the horizontal part. Hence the proximal mappings of the ordinary second differences agree with the cyclic version (under identification via $\exp_q$). This implies (30) for J_l, $l = 6,\ldots,11$.
Finally, we consider the mixed second order differences $d_{1,1}(x_{i,j}, x_{i+1,j}, x_{i,j+1}, x_{i+1,j+1})$. As above, we have for neighboring data items that the distance is smaller than $\frac{3\pi}{8}$. For all four contributing values of x we have that the pairwise distance is smaller by $\frac{\pi}{2}$. Thus again they lie on a quarter of the circle. Hence, the proximal mapping for the ordinary mixed second differences agree with the cyclic version (under identification via $\exp_q$). This implies (30) for J_l, $l = 12,\ldots,15$. □

We note that Lemma 4.9 does not guarantee that $\operatorname{prox}_{\lambda J_l}(x)$ remains in $\mathcal{S}(f,\delta)$. Therefore it does not allow for an iterated application. In the following main theorem we combine the preceding lemmas to establish this property.

Theorem 4.10. *Let $f \in (\mathbb{S}^1)^{N\times M}$ with $d_\infty(f) < \frac{\pi}{8}$. Let $\lambda := \{\lambda_k\}_k$ fulfill property (20) and*

$$\sqrt{\varepsilon^2 + 2\|\lambda\|_2^2 L^2 c(c+1)} + 2\|\lambda\|_\infty cL < \frac{\pi}{16},$$

for some $\varepsilon > 0$, where $c = 15$ and $L = 4$. Further, assume that the parameters α, β, γ of the functional J in (21) and ε satisfy (25). Then the sequence $\{x^{(k)}\}_k$ generated by the CPPA in Algorithm 1 converges to a global minimizer of J.

Proof. Let $\tilde{f}$ be the lifting of of f with respect to a base point q not antipodal to $f_{1,1}$ and fixed $\tilde{f}_{1,1}$ with $\exp_q(\tilde{f}_{1,1}) = f_{1,1}$. Further, let $\tilde{J}$ denote the real analog of J. By Lemma 4.3 we have $\mathrm{TV}_1(f) = \widetilde{\mathrm{TV}}_1(\tilde{f})$ and $\mathrm{TV}_2^\bullet(f) = \widetilde{\mathrm{TV}}_2^\bullet(\tilde{f})$ for $\bullet \in \{\mathrm{hv}, \mathrm{d}\}$ such that (25) is also fulfilled for the real-valued setting. Then we can apply Remark 4.5 and conclude that the minimizer y^* of $\tilde{J}$ fulfills $\|y^* - \tilde{f}\|_2 \leq \varepsilon < \frac{\pi}{16}$. By (26) we obtain

$$\begin{aligned} R &= \sqrt{\|y^* - \tilde{f}\|_2^2 + 2\|\lambda\|_2^2 L^2 c(c+1)} + 2\|\lambda\|_\infty cL \\ &\leq \sqrt{\varepsilon^2 + 2\|\lambda\|_2^2 L^2 c(c+1)} + 2\|\lambda\|_\infty cL < \frac{\pi}{16}. \end{aligned}$$

By Lemma 4.8 the iterates $y^{(k+\frac{l}{c})}$ of the real-valued CPPA fulfill

$$\|y^{(k+\frac{l}{c})} - y^*\|_2 \leq R < \frac{\pi}{16}.$$

Hence $\|y^{(k+\frac{l}{c})} - \tilde{f}\|_\infty < \frac{\pi}{8}$ which means that all iterates $y^{(k+\frac{l}{c})}$ stay within $\tilde{\mathcal{S}}(\tilde{f}, \frac{\pi}{8})$.
Next, we consider the sequence $\{x^{(l+\frac{k}{c})}\}$ of the CPPA for the $\mathbb{S}^1$-valued data f. We use induction to verify $x^{(k+\frac{l}{c})} = \exp_q(y^{(k+\frac{l}{c})})$. By definition we have $x^{(0)} = f = \exp_q(\tilde{f}) = \exp_q(y^{(0)})$. Assume that $x^{(k+\frac{l-1}{c})} = \exp_q(y^{(k+\frac{l-1}{c})})$. By bijectivity of the lifting, cf. Lemma 4.3, and since $y^{(k+\frac{l-1}{c})} \in \tilde{\mathcal{S}}(\tilde{f}, \delta)$, we conclude $x^{(k+\frac{l-1}{c})} \in \mathcal{S}(f, \delta)$. By Lemma 4.9 we obtain

$$\exp_q(y^{(k+\frac{l}{c})}) = \exp_q\big(\mathrm{prox}_{\lambda_k \tilde{J}_l}(y^{(k+\frac{l-1}{c})})\big) = \mathrm{prox}_{\lambda_k J_l}(x^{(k+\frac{l-1}{c})}) = x^{(k+\frac{l}{c})}.$$

By the same argument as above we have again $x^{(k+\frac{l}{c})} \in \mathcal{S}(f, \delta)$.
Finally, we know by Theorem 4.1 that

$$x^{(k)} = \exp_q(y^{(k)}) \to \exp_q(y^*) \quad \text{as} \quad k \to \infty$$

and by Lemma 4.7 that $x^* := \exp_q(y^*)$ is a global minimizer of J. This completes the proof. □

Remark 4.11. *The assumptions of Theorem 4.10 are more restrictive than what is actually observed in practice. As it is also known e.g., from subdivision schemes for manifold-valued data [59] the actually observed domains of convergence are significantly larger than theoretically proved. When dealing with global convergence of manifold-valued algorithms, the first issue typically is, that the objects under consideration are not even globally defined, e.g.,* the shortest path *joining two points is not globally defined. However, such a path is also not a purely local object, e.g., on $\mathbb{S}^1$ it is well defined as long as the points are not antipodal. In this sense, the algorithm proposed in this paper is not only local in nature.*
On the other hand, if neighboring data are too far apart, algorithms for nonlinear problems might even be not well-defined. Thus addressing global convergence fails at the point where well-definiteness is questionable. For example, in our case, if neighboring points are antipodal, the proximal mapping of the distance function is two-valued. Even if we apply a rule to force one alternative, the interpretation of the obtained result is still open. Of course such artificial examples can be constructed. However, in all experiments we did so far we have observed the convergence of our algorithm.

5 Numerical results

In this section we demonstrate the performance of our algorithms by numerical examples. First we consider one-dimensional signals to demonstrate the influence of first and/or second order differences on the denoising process. Furthermore, we compare our regularizers based on first and second order differences with those involving a cyclic Huber function.
For two-dimensional data, we provide three groups of denoising experiments:

i) InSAR imaging, where we denoise a synthetic image as well as a real world acquisition;

ii) hue channel reconstruction, which is an artificial, but illuminating experiment;

iii) phase valued image restoration occurring in event-related potentials (ERPs) appearing in habituation processes.

For the numerical computations of the following examples, the algorithms presented in Section 4 were implemented in MatLab. The computations were performed on a MacBook Pro with an Intel Core i5, 2.6 Ghz and 8 GB of RAM using MatLab 2013, Version 2013a (8.1.0.604) on Mac OS 10.9.2.

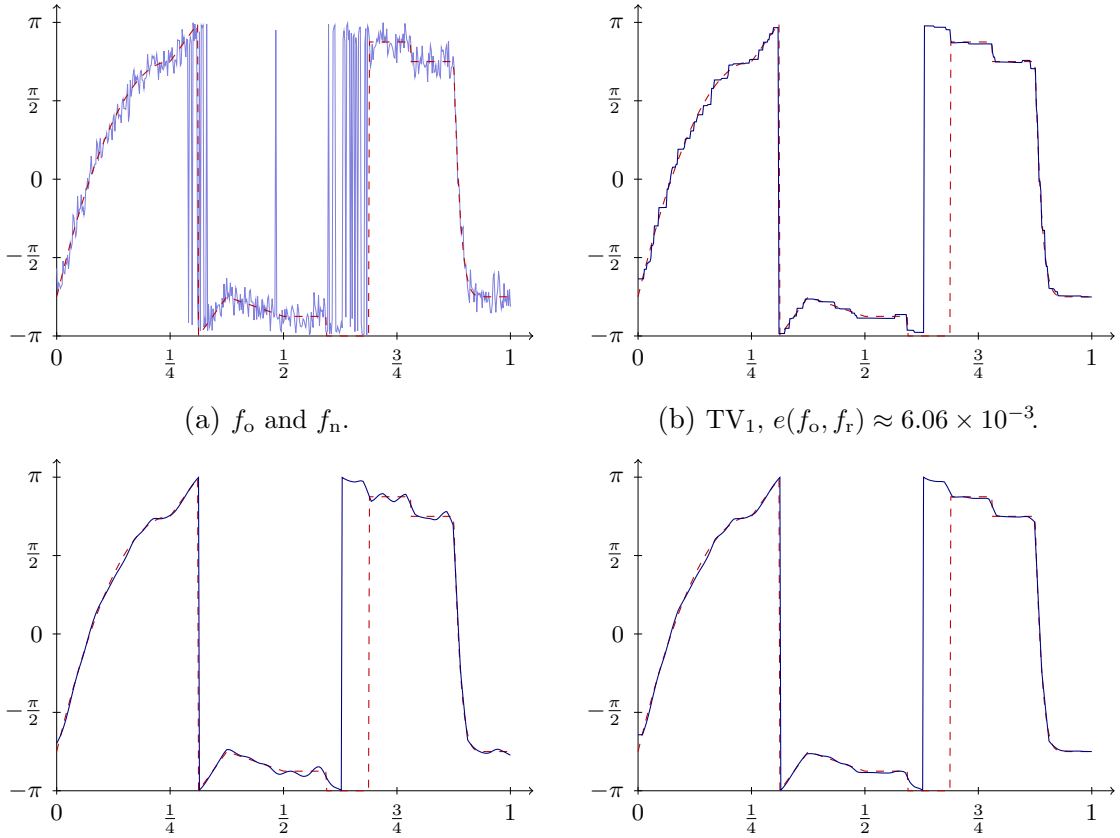
5.1 Signal denoising of synthetic data

Influence of first and/or second order differences. The first example of a synthetic one-dimensional signal demonstrates the effect of different models containing absolute cyclic first order differences, second order differences or both combined. The function $f : [0,1] \to [-\pi, \pi)$ given by

$$f(x) := \begin{cases} -24\pi x^2 + \frac{3}{4}\pi & \text{for } 0 \le x \le \frac{1}{4}, \\ 4\pi x - \frac{\pi}{4} & \text{for } \frac{1}{4} < x \le \frac{3}{8}, \\ \left(-\pi x - \frac{3}{8}\right)_{2\pi} & \text{for } \frac{3}{8} < x \le \frac{1}{2}, \\ \left(-\frac{j+7}{8}\pi\right)_{2\pi} & \text{for } \frac{3j+16}{32} < x \le \frac{3j+19}{32},\ j = 0,1,2,3, \\ \frac{3}{2}\pi \exp\left(-\frac{35}{7} - \frac{1}{1-x}\right) - \frac{3}{4}\pi & \text{for } \frac{7}{8} < x \le 1, \end{cases}$$

is sampled equidistantly to obtain the original signal $f_{\mathrm{o}} = \left(f\left(\frac{i-1}{N-1}\right)\right)_{i=1}^{N}$ at $N = 500$ samples. This function is distorted by wrapped Gaussian noise η of standard deviation $\sigma = \frac{1}{5}$ to get $f_{\mathrm{n}} := \left(f_{\mathrm{o}} + (\eta)_{2\pi}\right)_{2\pi} = (f_{\mathrm{o}} + \eta)_{2\pi}$, see also Remark 2.1. The functions f_{o} and f_{n} are depicted in Figure 4 (a). Note the following effects due to the cyclic data representation on $[-\pi, \pi)$: The linear increase on $\left[\frac{1}{4}, \frac{3}{8}\right]$ of f is continuous and the change from π to $-\pi$ at $\frac{5}{16}$ is just due to the chosen representation system. Similarly the two constant parts with the values $-\pi$ and $\frac{7}{8}\pi$ differ only by a jump size of $-\frac{\pi}{8}$. For the noise around these two areas, we have the same situation.
We apply Algorithm 1 with different model parameters α and β to f_{n} which yields the restored signals f_{r}. The restoration error is measured by the 'cyclic' mean squared error (cMSE) with respect to the arc length distance

$$e(f_{\mathrm{o}}, f_{\mathrm{r}}) := \frac{1}{N} \sum_{i=1}^{M} d(f_{\mathrm{o},i}, f_{\mathrm{r},i})^2.$$



(a) f_{o} and f_{n}.

(b) TV_1, $e(f_{\mathrm{o}}, f_{\mathrm{r}}) \approx 6.06 \times 10^{-3}$.

(c) TV_2, $e(f_{\mathrm{o}}, f_{\mathrm{r}}) \approx 4.34 \times 10^{-3}$.

(d) $\mathrm{TV}_1\&\mathrm{TV}_2$, , $e(f_{\mathrm{o}}, f_{\mathrm{r}}) \approx 3.53 \times 10^{-3}$.

Figure 4. Denoising of an one-dimensional signal by minimizing (19) with CPPA. (a) Original signal f_{o} (dashed red) and disturbed signal by wrapped Gaussian noise f_{n} (solid black). (b)—(d) Reconstructed signals f_{r} using (b) only the TV_1 regularizer ($\alpha = \frac{3}{4}$), (c) only the TV_2 regularizer ($\beta = \frac{3}{2}$), and (d) both of them ($\alpha = \frac{1}{2}$, $\beta = 1$). While (b) suffers from the staircasing effect, (c) shows weak results at constant areas. The combination of both regularizers in (d) yields the best image.

We use $\lambda_0 = \pi$ and $k = 4000$ iterations as stopping criterion. For any choice of parameters α, β the computation time is about 6 seconds.
The result f_r in Figure 4 (b) is obtained using only the TV_1 regularization ($\alpha = \frac{3}{4}, \beta = 0$). The restoration of constant areas is favored by this regularization term, but linear, quadratic and exponential parts suffer from the well-known 'staircasing' effect. Utilizing only the TV_2 regularization ($\alpha = 0, \beta = \frac{2}{3}$) , cf. Figure 4 (c), the restored function becomes worse in flat areas, but shows a better quality in the linear parts. By combining the regularization terms ($\alpha = \frac{1}{2}$, $\beta = 1$) as illustrated in Figure 4 (d) both the linear and the constant parts are reconstructed quite well and the cMSE is smaller than for the other choices of parameters. Note that α and β were chosen in $\frac{1}{4}\mathbb{N}$ with respect to an optimal cMSE.

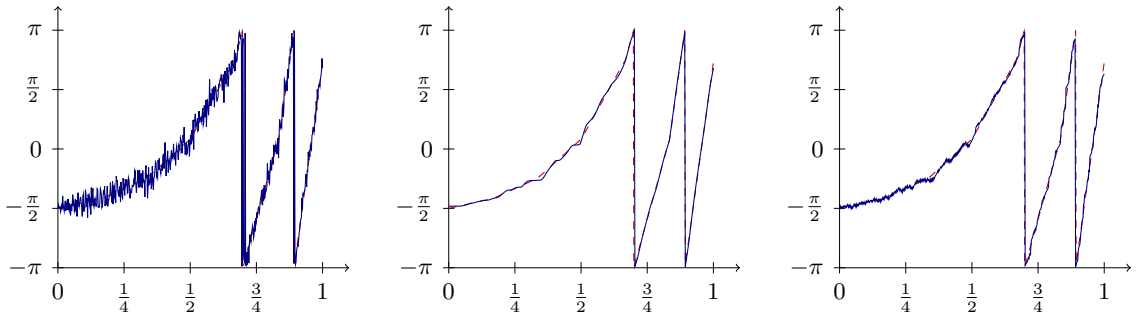
Comparison to Huber regularization. The second example compares our regularizers with those involving the Huber functional for cyclic data. Following the definition in [63], the Huber regularizer is given by $h(d(x, y))$, where h denotes the *Huber function* (Moreau envelope of the absolute value function)

$$h(s) := \begin{cases} \tau^2 s^2 & \text{for } s < \frac{\omega}{\sqrt{2}\tau}, \\ \omega\sqrt{2}\tau s - \omega^2/2 & \text{otherwise.} \end{cases}$$

The parameter $\omega > 0$ corresponds to the width of the interval around 0, where the absolute value function is replaced by a quadratic function, while $\tau > 0$ characterizes the parabola. The corresponding proximal mapping was derived in [63] and is used within the cyclic proximal point method.
Choosing a continuous piecewise linear function with increasing slope $a\frac{\pi}{2}, a = 1, 2, 4, 8, 16$ as in Figure 5, the method related to the Huber regularizer does not remove the noise in all of the signal whereas our model removes the noise almost perfectly. Note that we have tested a wide range of parameters τ, ω and refer to the best result with respect to the error $e(f_0, f_r)$ achieved for $\tau = 4\sqrt{2}$ and $\omega = \frac{\pi}{16}$. This is due to the fact that the Huber model balances between Tikhonov-like smoothing artifacts if ω is large and staircasing if ω is small.

5.2 Image denoising of InSAR data

The complex-valued synthetic aperture radar (SAR) data is obtained emitting specific radar signals at equidistant points and measuring the amplitude and phase of their reflections by the earth's surface. The amplitude provides information about the reflectivity of the surface. The phase encodes both the change of the elevation of the surface's elements within the measured area and their reflection properties and is therefore rather arbitrary. When taking two SAR images of the target area at the same time but from different angles or locations. The phase difference of these images encodes the elevation, but it is restricted to one wavelength and also includes noise. The result is the so called interferometric synthetic aperture radar (InSAR) data and consists of the 'wrapped phase' or the 'principal phase', a value in $[-\pi, \pi)$ representing the surface elevation. For more details see, e.g., [11, 40].
After a suitable preregistration the same approach can be applied to two images from the same area taken at different points in time to measure surface displacements, e.g., before and after an earthquake or the movement of glaciers.
The main challenge in order to unwrap the phase is the presence of noise. Ideally, if the surface would be smooth enough and no noise would be present, unwrapping is uniquely determined,



(a) Noisy signal $\sigma = 0.3$.

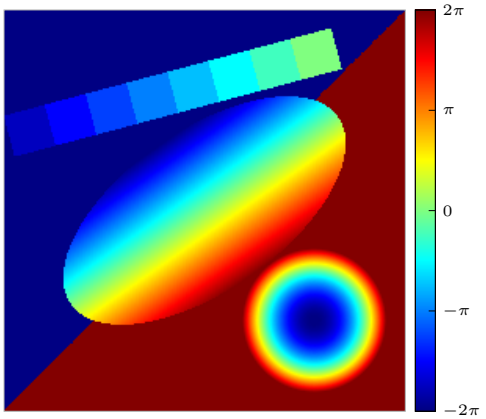
(b) Denoised with TV_1&TV_2 $e(f_o, f_r) \approx 2.02 \times 10^{-3}$.

(c) Denoised with Huber. $e(f_o, f_r) \approx 5.15 \times 10^{-2}$.

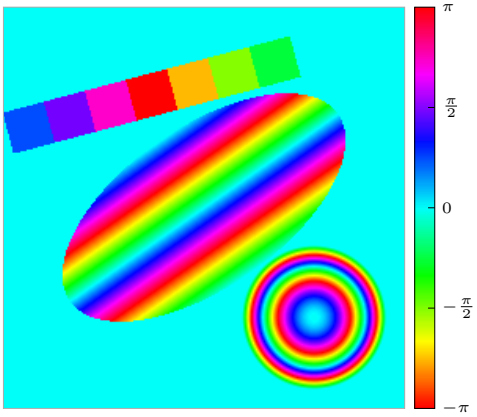
Figure 5. Comparison of denoising with cyclic first and second order differences, and an approach using the Huber functional. (a) Noisy signal $\sigma = 0.3$. (b) Denoised signal by our method ($\alpha = (\frac{1}{2}, \frac{1}{2})$, $\beta = (1, 1)$, $\gamma = 0$). (c) Denoised signal by Huber's model ($\alpha = \frac{1}{4}$). The later introduces artifacts in regions with smaller slope.

i.e., differences between two pixels larger than π are regarded as a wrapping result and hence become unwrapped.
There are several algorithms to unwrap, even combining the denoising and the unwrapping, see for example [6, 7]. For denoising, Deledalle et al. [18] use both SAR images and apply a non-local means algorithm jointly to their reflection, the interferometric phase and the coherence.

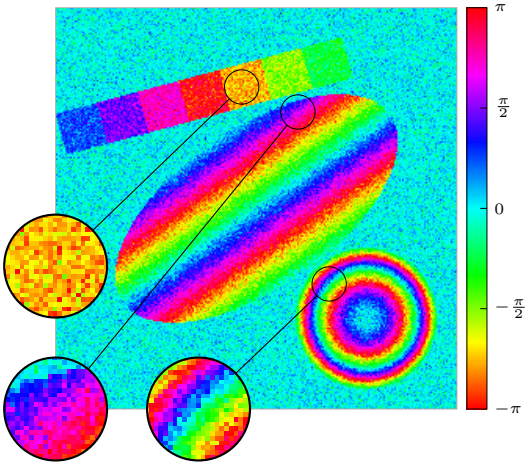
Application to synthetic data. In order to get a better understanding in the two-dimensional case, let us first take a look at a synthetic surface given on $[0, 1]^2$ with the profile shown in Figure 6 (a). This surface consists of two plates of height $\pm 2\pi$ divided at the diagonal, a set of stairs in the upper left corner in direction $\frac{\pi}{3}$, a linear increasing area connecting both plateaus having the shape of an ellipse with major axis at the angle $\frac{\pi}{6}$, and a half ellipsoid forming a dent in the lower right of the image with circular diameter of size $\frac{9}{25}$ and depth 4π. The initial data is given by sampling the described surface at $M = N = 256$ sampling points. The usual InSAR measurement would ideally result in data as given in Figure 6 (b), i.e., the data is wrapped with respect to 2π. In the figure the resulting ideal phase is represented using the hue component of the HSV color space. Again, the data is perturbed by wrapped Gaussian noise, standard deviation $\sigma = 0.3$, see Figure 6 (c).
For an application of Algorithm 1 to the minimization problem (21), we have to fix five parameters $\alpha_1, \alpha_2, \beta_1, \beta_2, \gamma$ which were chosen on $\frac{1}{8}\mathbb{N}$ such that they minimize the cMSE. Using only the cyclic first order differences with $\alpha = \frac{1}{8}(3, 2)$, see Figure 6 (d), the reconstructed image f_r reproduces the piecewise constant parts of the stairs in the upper left part and the background, but introduces a staircasing in both linear increasing areas inside the ellipse and in the half ellipsoid. This is highlighted in the three magnifications in Figure 6 (d). Applying only cyclic second order differences with $\beta_1 = \beta_2 = \gamma = \frac{1}{8}$ manages to reconstruct the linear increasing part and the circular structure of the ellipsoid, but compared to the first case it even increases the cMSE due to the approximation of the stairs and the background, see especially the magnification of the stairs in Figure 6 (e). Combining first and second order cyclic differences by setting $\alpha_1 = \alpha_2 = \frac{1}{8}(2, 1)$ and $\beta_1 = \beta_2 = \frac{1}{8}$, $\gamma = 0$, these disadvantages can be reduced, cf. Figure 6 (f). Note especially the three magnified regions and the cMSE.



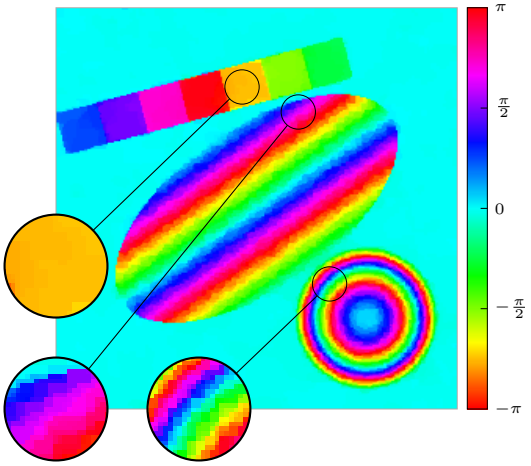
(a) Original data.



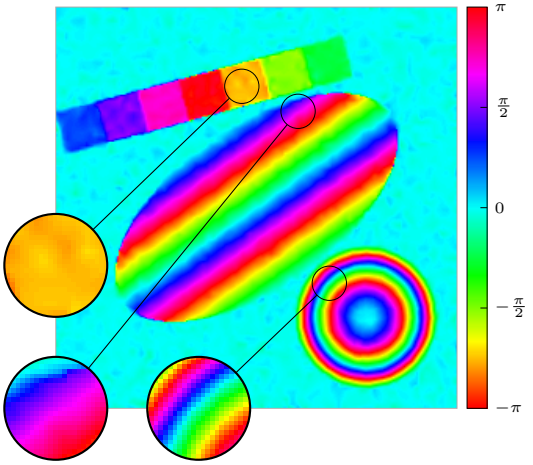
(b) Wrapped data f_{o}.



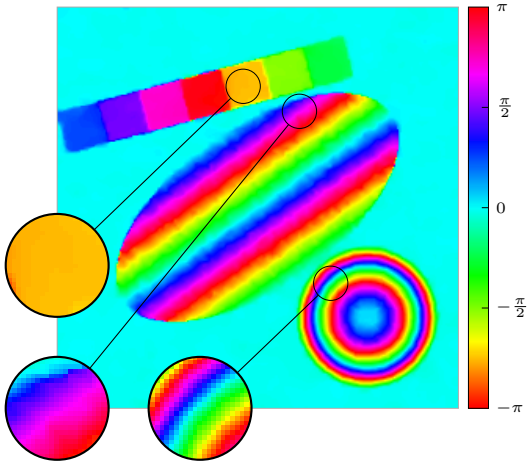
(c) Noisy data f_{n}.



(d) TV_1, $e(f_{\mathrm{o}}, f_{\mathrm{r}}) = 7.09 \times 10^{-3}$.

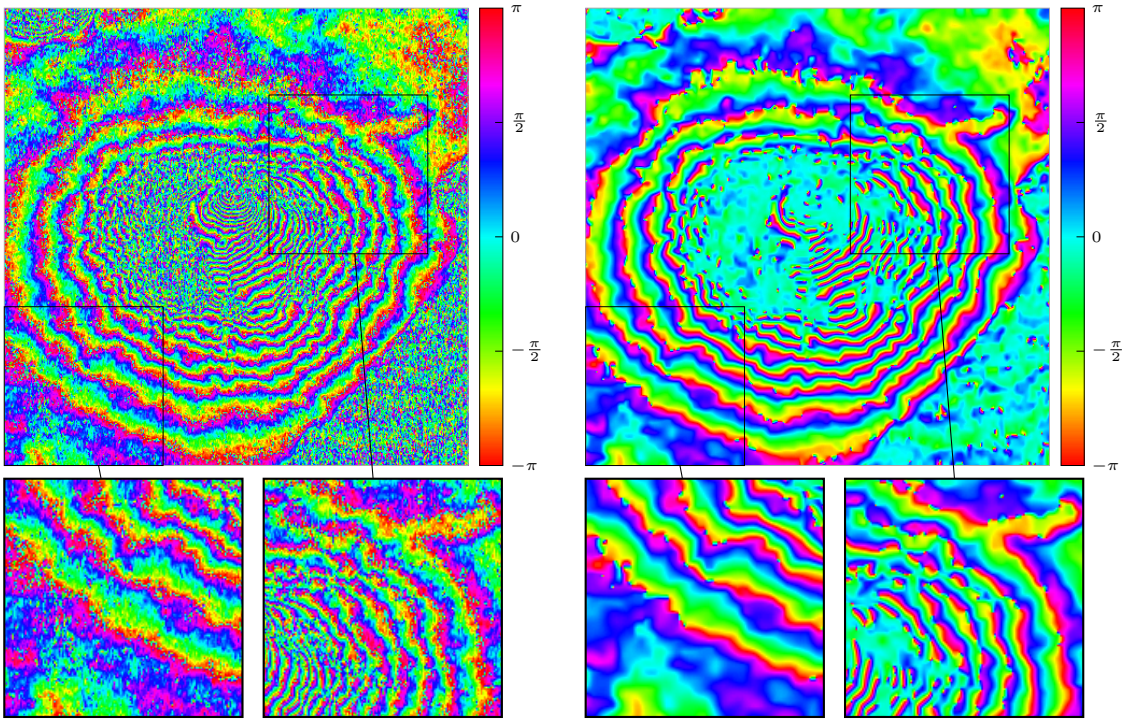


(e) $\mathrm{TV}_2^{\mathrm{hv}}$, $e(f_{\mathrm{o}}, f_{\mathrm{r}}) = 6.70 \times 10^{-3}$.



(f) TV_1 & $\mathrm{TV}_2^{\mathrm{hv}}$, $e(f_{\mathrm{o}}, f_{\mathrm{r}}) = 5.37 \times 10^{-3}$.

Figure 6. Denoising of two-dimensional artificial data by minimizing (21) with CPPA. (a) Artificial surface, (b) its wrapped variant, and (c) wrapped image corrupted by wrapped Gaussian noise. (d)–(f) Reconstructed images f_{r} using (d) only the TV_1 regularizer ($\alpha = (\frac{3}{8}, \frac{1}{4})$), (e) only the $\mathrm{TV}_2^{\mathrm{hv}}$ & $\mathrm{TV}_2^{\mathrm{d}}$ regularizer ($\beta = (\frac{1}{8}, \frac{1}{8})$, $\gamma = \frac{1}{8}$), and (f) both of them ($\alpha = (\frac{1}{4}, \frac{1}{8})$, $\beta = (\frac{1}{8}, \frac{1}{8})$, $\gamma = 0$).



(a) Complete Vesuvuis data. (b) Denoised Vesuvius data.

Figure 7. (a) Noisy InSAR data set of the Vesuvius taken by the ERS-1 satellite [48]. (b) Denoised image by minimizing (21) with CPPA ($\alpha = \left(\frac{1}{4}, \frac{1}{4}\right)$, $\beta = \left(\frac{3}{4}, \frac{3}{4}\right)$ and $\gamma = \frac{3}{4}$).

Application to real-world data. Next we examine a real-world example. The data from [48] is a set of InSAR data recorded in 1991 by the ERS-1 satellite capturing topographical information from the Mount Vesuvius. The data is available online[1] and a part of it was also used as an example in [63] for TV based denoising of manifold-valued data. In Figure 7 the phase is represented by the hue component of the HSV color space. We apply Algorithm 1 to the image of size 426×432, cf. Figure 7 (a), with $\alpha_1 = \alpha_2 = \frac{1}{4}$ and $\beta_1 = \beta_2 = \gamma = \frac{3}{4}$. This reduces the noise while keeping all significant plateaus, ascents and descents, cf. Figure 7 (b). The left zoom illustrates how the plateau in the bottom left of the data is smoothened but kept in its main elevation shown in blue. In the zoom on the right all major parts except the noise are kept. We notice just a little smoothening due to the linearization introduced by TV_2. In the bottom left of this detail some of the fringes are eliminated, and a small plateau is build instead, shown in cyan. The computation time for the whole image using $k = 600$ iterations as stopping criterion was 86.6 sec and 11.1 sec for each of the details of size 150×150.

5.3 Image hue value denoising.

The next example demonstrates the role of an appropriate denoising which respects the cyclic distance. Consider the test image "sailboat on a lake"[2] in Figure 8 (a). The hue channel of the HSV model contains cyclic data. Corrupting this channel by wrapped Gaussian noise results in the noisy image in Figure 8 (b), where especially the sky and lawn are pertained. Note that this experiment is artificial and is at this time not related to noise appearing in practice.
When denoising the hue channel with our first and second order cyclic difference model the resulting image in Figure 8 (c) differs only in a few spots, see for example the color of the path in the background which has lost a small amount of its red component. In comparison, applying a real valued first and second order difference model for denoising the hue introduces a change in the tone of the complete image as shown in Figure 8 (d).

5.4 Habituation data

Event-related potentials (ERPs) provide a noninvasive method to measure the neural activity during processing certain cognitive stimuli. The ERPs are obtained by segmenting an electroencephalogram (EEG) according to a repetitive stimulus. The following data were acquired at the Saarland University Hospital and we refer to [39] for a circumvent description of the experimental setup. Briefly speaking, a pure tone of 1 kHz with 50 dB sound pressure level (SPL) loudness is presented to a subject on a right headphone speaker for 40 ms. This is repeated $N = 931$ times with a break of 1 sec. After a pause of 3 minutes a second series of tones is presented using 100dB SPL. The acquired data was segmented into trials of 800 ms. By the sampling rate of 512 Hz we obtain $M = 410$ values per segment/trial. The segments $s_j \in \mathbb{R}^M$ form an $N \times M$ ERP image $(s_1, \ldots, s_N)^{\mathrm{T}}$. As 50 dB SPL is a low sound intensity, the subject can easily habituate to the sound contrary to 100 dB SPL which is a loud and aversive stimulus and hence habituation is not possible.
Long-term habituation was analyzed in the time domain, e.g., in [60] using partial averages of trials and in [55] based on nonlocal means denoising. Recently it became evident that more promising results can be obtained by working with the instantaneous phase of the data,

[1] at `https://earth.esa.int/workshops/ers97/program-details/speeches/rocca-et-al/`
[2] Taken from the USC-SIPI Image Database, see `http://sipi.usc.edu/database/database.php?volume=misc&image=14`

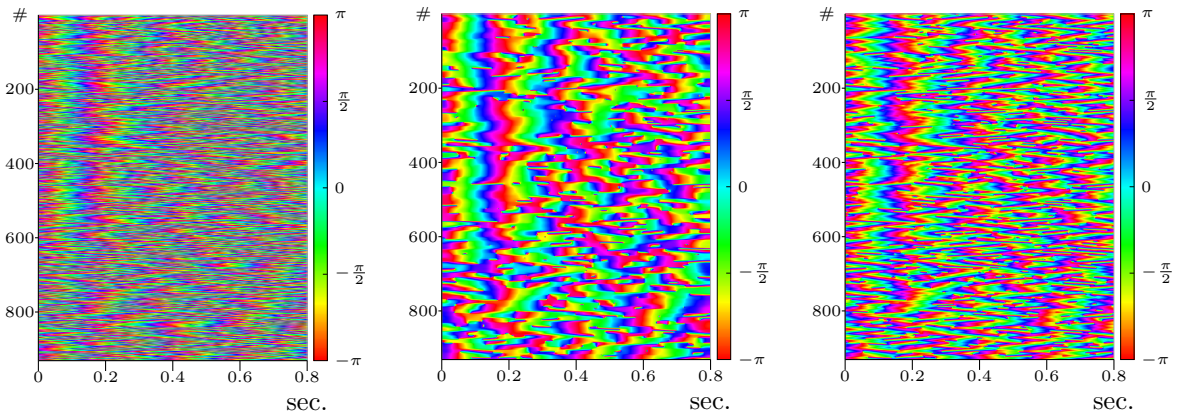
(a) Original image.

(b) Image with noisy hue.

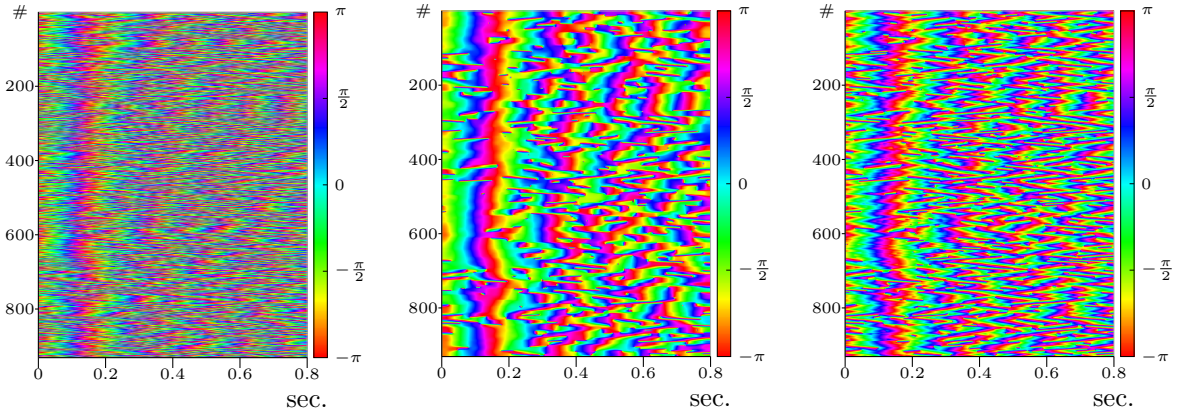
(c) Our cyclic method.

(d) Real valued method.

Figure 8. Denoising the image "sailboat on lake" with noisy hue channel: (a) Original image. (b) Image with hue channel corrupted by wrapped Gaussian noise ($\sigma = 0.4$). (c) Image with denoised hue channel by our cyclic first and second order difference model. (d) Image with denoised hue channel by real valued first and second order difference model (both $\alpha = (\frac{1}{4}, \frac{1}{4})$, $\beta = (\frac{1}{8}, \frac{1}{8})$, $\gamma = 0$).



(a) 50 dB SPL sweeps. (b) 50 dB SPL, $\mathrm{TV}_1 \,\&\, \mathrm{TV}_2^{\mathrm{hv}}$. (c) 50 dB SPL, $\mathrm{TV}_1^2 \,\&(\mathrm{TV}_2^{\mathrm{hv}})^2$.



(d) 100 dB sweeps (e) 100 dB SPL, $\mathrm{TV}_1 \,\&\, \mathrm{TV}_2^{\mathrm{hv}}$. (f) 100 dB SPL, $\mathrm{TV}_1^2 \,\&(\mathrm{TV}_2^{\mathrm{hv}})^2$.

Figure 9. Denoising of the instantaneous phase of ERP data: (a) and (d): Original phase data for the experiments with 50 dB SPL and 100 dB SPL, (b) and (e): Denoised with $\mathrm{TV}_1 \,\&\, \mathrm{TV}_2^{\mathrm{hv}}$, cf. (21), (c) and (f): Denoised with $\mathrm{TV}_1^2 \,\&(\mathrm{TV}_2^{\mathrm{hv}})^2$, cf. (31), where for all cases $\alpha = (2, 4)$, $\beta = (2.5, 1.25)$, $\gamma = 0$.

see [36, 43, 54]. In particular such approaches can cope with certain amplitude fluctuations which depend on the experimental setup. To this end, all single trials are convolved with the fourth order derivative of the complex Gaussian wavelet ψ^a of scale $a = 40$ from the MatLab Wavelet-Toolbox [41] and from the resulting complex valued signals only the phases $p_j \in [-\pi, \pi)^M$, $j = 1, \ldots, N$, are kept for further usage. The phase data is arranged in an image $(p_1, \ldots, p_N)^{\mathrm{T}}$ for the experimental series with 50 dB SPL and 100 dB SPL, respectively, cf. Figures 9 (a) and (d). .

We denoise both images by minimizing (21) with parameters $\alpha = (2, 4)$, $\beta = (2.5, 1.25)$ and $\gamma = 0$ by the CPPA. The results are shown in Figures 9 (b) and (e). Additionally, we compare this result with the denoising result obtained by minimizing

$$J(x) := F(x; f) + \alpha\, \mathrm{TV}_1^2(x) + \beta(\mathrm{TV}_2^{\mathrm{hv}})^2(x), \tag{31}$$

where TV_1^2 and $(\mathrm{TV}_2^{\mathrm{hv}})^2$ are defined as in (21) but with *squared* cyclic first and second order differences. The CPPA can be applied similarly as for (21) by using Theorem 3.6 instead

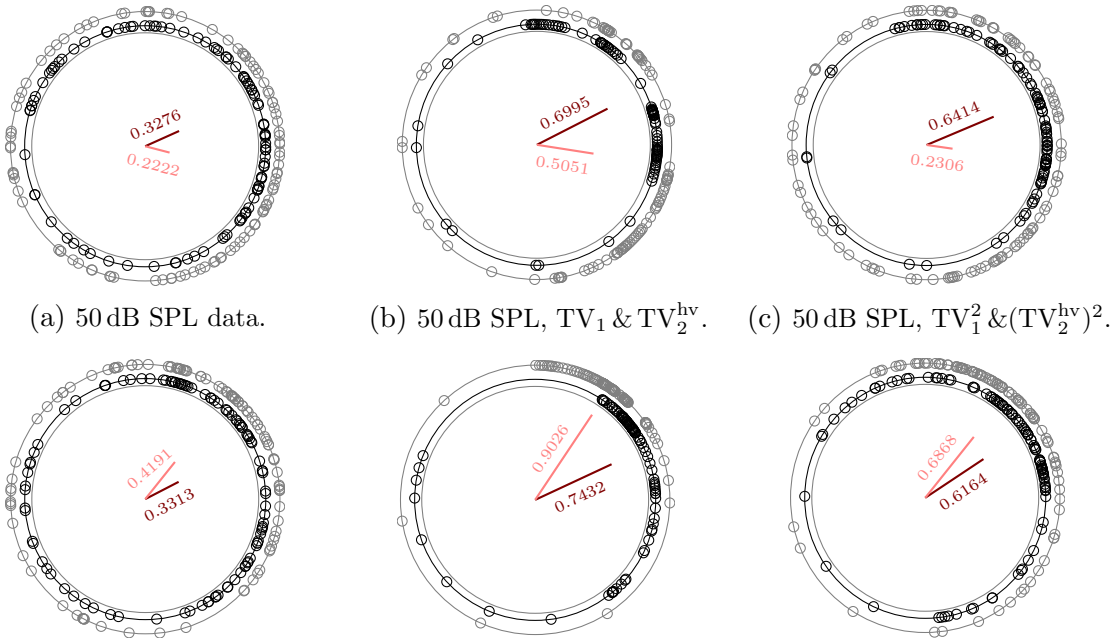
of Theorem 3.5 for computing the proximal values. The denoising results are depicted in Figures 9 (c) and (f).
For comparison of both minimizations we measure the habituation effect using a statistical approach by *circular means* which was also suggested in [55]. Taking a sequence of $Q = 100$ trials starting at $b \in \{1, \dots, N-Q\}$ of the sample 97, which corresponds to the time of 0.1893 secs, we compute

$$R_b = \frac{1}{Q}\left(\sum_{i=1}^{Q} \cos p_{b+i,97},\ \sum_{i=1}^{Q} \sin p_{b+i,97}\right).$$

In polar coordinates $R_b = |R_b|\, \mathrm{e}^{i\varphi_b}$, where the angle φ_b denotes the circular mean and $|R_b| \in [0,1]$ indicates the data concentration, i.e., a large value $|R_b|$ denotes a small variation, see [33].
In Figure 10 the first ($b = 1$) and last ($b = 832$) single trials are compared. The points $(\cos p_{b+i,97}, \sin p_{b+i,97})$, $i = 1, \dots, 100$ are shown on the inner circle for $b = 1$ in black and on the outer circle for $b = 832$ in gray.
To compare the corresponding values of R_b, $b = 1, 832$, they are normalized with respect to the thin most inner circle. The resulting complex numbers R_1 and R_{832} are drawn in dark and light red, respectively. For each of them, the value $|R_b|$ is inscribed at each line. The presence of noise in the original data results in a small value $|R_b|$ for both the 50 dB trials in Figure 10 (a) and 100 dB in (d). While there seems a little indication, that the patient is habituated to the 50 dB tone, there is only little difference comparing with the two 100 dB sweeps.
By denoising with the TV functional from (21) one can clearly see the alignment of the phase in both sweeps of the 100 dB data in 10 (e) are well concentrated, while for the 50 dB data, the second is spread, see Figure 10 (b) on the outer circle.
When using the functional (31), the denoising emphasizes this habituation even more, as can be seen when comparing the absolute values of $|R_b|$, $b = 1, 832$, in Figure 10 (c) and (f): the value decreases for the 50 dB data by about by two thirds while decreasing by only about one tenth for the loud 100 dB sweeps.

6 Conclusions

In this paper we considered functionals having regularizers with second order absolute cyclic differences for $\mathbb{S}^1$-valued data. Their definition required a proper notion of higher order differences of cyclic data generalizing the corresponding concept in Euclidian spaces. We derived a CPPA for the minimization of our functionals and gave the explicit expressions for the appearing proximal mappings. We proved convergence of the CPPA under certain conditions. To the best of our knowledge this is the first algorithm dealing with higher order TV-type minimization for $\mathbb{S}^1$-valued data. We demonstrated the denoising capabilities of our model on synthetic as well as on real-world data.
Future work includes the application of our higher order methods for cyclic data to other imaging tasks such as segmentation, inpainting or deblurring. For deblurring, the usually underlying linear convolution kernel has to be replaced by a nonlinear construction based on intrinsic (also called Karcher) means. This leads to the task of solving the new associated inverse problem.
Further, we intend to investigate other couplings of first and second order derivatives similar to infimal convolutions or GTV for Euclidean data. Finally, we want to set up higher order



(a) 50 dB SPL data. (b) 50 dB SPL, TV_1 & $\mathrm{TV}_2^{\mathrm{hv}}$. (c) 50 dB SPL, TV_1^2 &$(\mathrm{TV}_2^{\mathrm{hv}})^2$.

(d) 100 dB SPL data. (e) 100 dB SPL, TV_1 & $\mathrm{TV}_2^{\mathrm{hv}}$. (f) 100 dB SPL, TV_1^2 &$(\mathrm{TV}_2^{\mathrm{hv}})^2$.

Figure 10. The first (inner circle, dark) and last (outer circle, light) 100 sweeps from Figure 9 for the original data (first column), denoising the instantaneous phase with $p = 1$ and $p = 2$ (second and third column) for both measurements of 50 dB SPL (first row) and 100 dB SPL (second row). The thin circle denotes the $|R| = 1$ line, the two inner lines denote R_b, $b = 1, 832$, in dark and light red, respectively.

TV-like methods for more general manifolds, e.g. higher dimensional spheres. Here, we do not believe that it is possible to derive explicit expressions for the involved proximal mappings – at least not for Riemannian manifolds of nonzero sectional curvature. Instead, we plan to resort to iterative techniques.

Acknowledgement. We like to thank D. J. Strauss from the Systems Neuroscience & Neurotechnology Unit of the Saarland University for providing us the long-term habituation data. Furthermore we acknowledge the financial support by DFG Grant STE571/11-1.

References

[1] M. Bačák. The proximal point algorithm in metric spaces. *Isr. J. Math.*, 194(2):689–701, 2013.

[2] M. Bačák. Computing medians and means in Hadamard spaces. *to appear in SIAM J. Optim.*, 2014.

[3] J. M. Bardsley. Wavefront reconstruction methods for adaptive optics systems on ground-based telescopes. *SIAM Journal Matrix Analysis and Applications*, (30):67–83, 2008.

[4] D. P. Bertsekas. Incremental gradient, subgradient, and proximal methods for convex optimization: a survey. Technical Report LIDS-P-2848, Laboratory for Information and Decision Systems, MIT, Cambridge, MA, 2010.

[5] D. P. Bertsekas. Incremental proximal methods for large scale convex optimization. *Math. Program., Ser. B*, 129(2):163–195, 2011.

[6] J. Bioucas-Dias, V. Katkovnik, J. Astola, and K. Egiazarian. Absolute phase estimation: adaptive local denoising and global unwrapping. *Appl. Optics*, 47(29):5358–5369, 2008.

[7] J. Bioucas-Dias and G. Valadão. Phase unwrapping via graph cuts. *IEEE Trans. on Image Process.*, 16(3):698–709, 2007.

[8] A. Björck. *Numerical Methods for Least Squares Problems*. SIAM, Philadelphia, 1996.

[9] K. Bredies, K. Kunisch, and T. Pock. Total generalized variation. *SIAM J. Imaging Sci.*, 3(3):1–42, 2009.

[10] H. Brezis and Y. Li. Topology and Sobolev spaces. *Journal of Functional Analysis*, 183(2):321–369, 2001.

[11] R. Bürgmann, P. A. Rosen, and E. J. Fielding. Synthetic aperture radar interferometry to measure earth's surface topography and its deformation. *Annu. Rev. Earth Planet. Sci.*, 28(1):169–209, 2000.

[12] A. Chambolle and P.-L. Lions. Image recovery via total variation minimization and related problems. *Numer. Math.*, 76(2):167–188, 1997.

[13] R. H. Chan, X. Yuan, and W. Zhang. A phase model for point spread function estimation in ground-based astronomy. *Preprint Chinese University of Hongkong*, 2014.

[14] T. F. Chan, S. Esedoglu, and F. E. Park. Image decomposition combining staircase reduction and texture extraction. *J. Vis. Commun. Image R.*, 18(6):464–486, 2007.

[15] T. F. Chan, A. Marquina, and P. Mulet. High-order total variation-based image restoration. *SIAM J. Sci. Comput.*, 22(2):503–516, 2000.

[16] C. Chefd'Hotel, D. Tschumperlé, R. Deriche, and O. Faugeras. Regularizing flows for constrained matrix-valued images. *J. Math. Imaging Vis.*, 20(1-2):147–162, 2004.

[17] Q. Chu, S. Jefferies, and J. G. Nagy. Iterative wavefront reconstruction for astronomical imaging. *SIAM Journal on Scientific Computing*, to appear.

[18] C.-A. Deledalle, L. Denis, and F. Tupin. NL-InSAR: Nonlocal interferogram estimation. *IEEE Trans. Geosci. Remote Sensing*, 49(4):1441–1452, 2011.

[19] S. Didas, G. Steidl, and S. Setzer. Combined ℓ_2 data and gradient fitting in conjunction with ℓ_1 regularization. *Adv. Comput. Math.*, 30(1):79–99, 2009.

[20] S. Didas, J. Weickert, and B. Burgeth. Properties of higher order nonlinear diffusion filtering. *J. Math. Imaging Vis.*, 35:208–226, 2009.

[21] O. P. Ferreira and P. R. Oliveira. Proximal point algorithm on Riemannian manifolds. *Optimization*, 51(2):257–270, 2002.

[22] N. I. Fisher. Statistical Analysis of Circular Data. Cambridge University Press, 1995.

[23] P. Fletcher. Geodesic regression and the theory of least squares on Riemannian manifolds. *Int. J. Comput. Vision*, 105(2):171–185, 2013.

[24] P. Fletcher and S. Joshi. Riemannian geometry for the statistical analysis of diffusion tensor data. *Signal Process.*, 87(2):250–262, 2007.

[25] M. Giaquinta, G. Modica, and J. Souček. Variational problems for maps of bounded variation with values in S^1. *Calc. Var.*, 1(1):87–121, 1993.

[26] M. Giaquinta and D. Mucci. The BV-energy of maps into a manifold: relaxation and density results. *Ann. Sc. Norm. Super. Pisa Cl. Sci.*, 5(4):483–548, 2006.

[27] M. Giaquinta and D. Mucci. Maps of bounded variation with values into a manifold: total variation and relaxed energy. *Pure Appl. Math. Q.*, 3(2):513–538, 2007.

[28] J. W. Goodman. *Introduction to Fourier Optics*. McGraw-Hill, 1996.

[29] P. Grohs, H. Hardering, and O. Sander. Optimal a priori discretization error bounds for geodesic finite elements. Technical Report 2013-16, Seminar for Applied Mathematics, ETH Zürich, Switzerland, 2013.

[30] P. Grohs and J. Wallner. Interpolatory wavelets for manifold-valued data. *Appl. Comput. Harmon. Anal.*, 27(3):325–333, 2009.

[31] S. Harizanov, P. Oswald, and T. Shingel. Normal multi-scale transforms for curves. *Found. Comput. Math.*, 11(6):617–656, 2011.

[32] W. Hinterberger and O. Scherzer. Variational methods on the space of functions of bounded Hessian for convexification and denoising. *Computing*, 76(1):109–133, 2006.

[33] S. R. Jammalamadaka and A. SenGupta. *Topics in Circular Statistics*. World Scientific Publishing Company, 2001.

[34] S. Lefkimmiatis, A. Bourquard, and M. Unser. Hessian-based norm regularization for image restoration with biomedical applications. *IEEE Trans. on Image Process.*, 21(3):983–995, 2012.

[35] J. Lellmann, E. Strekalovskiy, S. Koetter, and D. Cremers. Total variation regularization for functions with values in a manifold. In *IEEE ICCV 2013*, pages 2944–2951, 2013.

[36] Y. F. Low and D. J. Strauss. A performance study of the wavelet-phase stability (wps) in auditory selective attention. *Brain Res. Bull.*, 2011.

[37] M. Lysaker, A. Lundervold, and X.-C. Tai. Noise removal using fourth-order partial differential equations with applications to medical magnetic resonance images in space and time. *IEEE Trans. on Image Process.*, 12(12):1579–1590, 2003.

[38] M. Lysaker and X.-C. Tai. Iterative image restoration combining total variation minimization and a second-order functional. *Int. J. Comput. Vis.*, 66(1):5–18, 2006.

[39] M. Mariam, W. Delb, F. I. Corona-Strauss, M. Bloching, and D. J. Strauss. Comparing the habituation of late auditory evoked potentials to loud and soft sound. *Physiol. Meas.*, 30(2):141–153, 2009.

[40] D. Massonnet and K. L. Feigl. Radar interferometry and its application to changes in the Earth's surface. *Rev. Geophys.*, 36(4):441–500, 1998.

[41] M. Misiti, Y. Misiti, G. Oppenheim, and J.-M. Poggi. Matlab-wavelet toolbox, version 2.0. *The Mathworks, Inc., Natick, MA*, 2000.

[42] J. J. Moreau. Fonctions convexes duales et points proximaux dans un espace hilbertien. *C. R. Acad. Sci. Paris Ser. A Math.*, 255:2897–2899, 1962.

[43] Z. Mortezapouraghdam, L. Haab, F. I. Corona-Strauss, G. Steidl, and D. J. Strauss. Assessment of long-term habituation correlates in event-related potentials using a von Mises model. *Preprint*, 2013.

[44] K. Papafitsoros and C. B. Schönlieb. A combined first and second order variational approach for image reconstruction. *J. Math. Imaging Vis.*, 2(48):308–338, 2014.

[45] N. Parikh and S. Boyd. Proximity algorithms. *Foundations and Trends in Optimization*, 1(3):123–231, 2013.

[46] X. Pennec. Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements. *J. Math. Imaging Vis.*, 25(1):127–154, 2006.

[47] I. U. Rahman, I. Drori, V. C. Stodden, and D. L. Donoho. Multiscale representations for manifold-valued data. *Multiscale Model. Simul.*, 4(4):1201–1232, 2005.

[48] F. Rocca, C. Prati, and A. M. Guarnieri. Possibilities and limits of SAR interferometry. *ESA SP*, pages 15–26, 1997.

[49] R. T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM J. Control Optim.*, 14(5):877–898, 1976.

[50] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60(1):259–268, 1992.

[51] O. Scherzer. Denoising with higher order derivatives of bounded variation and an application to parameter estimation. *Computing*, 60:1–27, 1998.

[52] S. Setzer and G. Steidl. Variational methods with higher order derivatives in image processing. In *Approximation XII: San Antonio 2007*, pages 360–385, 2008.

[53] S. Setzer, G. Steidl, and T. Teuber. Infimal convolution regularizations with discrete l1-type functionals. *Commun. Math. Sci.*, 9(3):797–872, 2011.

[54] D. J. Strauss, W. Delb, R. D'Amelio, Y. F. Low, and P. Falkai. Objective quantification of the tinnitus decompensation by synchronization measures of auditory evoked single sweeps. *IEEE Trans. Neural Syst. Rehab. Eng.*, 16(1):74–81, 2008.

[55] D. J. Strauss, T. Teuber, G. Steidl, and F. I. Corona-Strauss. Exploiting the self-similarity in ERP images by nonlocal means for single-trial denoising. *IEEE Trans. Neural Syst. Rehabil. Eng.*, 21(4):576–583, 2013.

[56] E. Strekalovskiy and D. Cremers. Total variation for cyclic structures: Convex relaxation and efficient minimization. In *IEEE CVPR 2011*, pages 1905–1911. IEEE, 2011.

[57] E. Strekalovskiy and D. Cremers. Total cyclic variation and generalizations. *J. Math. Imaging Vis.*, 47(3):258–277, 2013.

[58] T. Valkonen, K. Bredies, and F. Knoll. Total generalized variation in diffusion tensor imaging. *SIAM J. Imag. Sci.*, 6(1):487–525, 2013.

[59] J. Wallner and N. Dyn. Convergence and C^1 analysis of subdivision schemes on manifolds by proximity. *Comput. Aided Geom. D.*, 22:593–622, 2005.

[60] V. Walpurger, G. Hebing-Lennartz, H. Denecke, and R. Pietrowsky. Habituation deficit in auditory event-related potentials in tinnitus complainers. *Hearing Res.*, 181(1):57–64, 2003.

[61] A. Weinmann. Nonlinear subdivision schemes on irregular meshes. *Constr. Approx.*, 31(3):395–415, 2010.

[62] A. Weinmann. Interpolatory multiscale representation for functions between manifolds. *SIAM J. Math. Anal.*, 44(1):162–191, 2012.

[63] A. Weinmann, L. Demaret, and M. Storath. Total variation regularization for manifold-valued data. *Preprint, ArXiv*, 2013.