ETH
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Seminar for Applied Mathematics SAM

# Total Variation Regularization by Iteratively Reweighted Least Squares on Hadamard Spaces and the Sphere

P. Grohs and M. Sprecher

Research Report No. 2014-39
December 2014

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

Funding SNF: 140635

# Total Variation Regularization by Iteratively Reweighted Least Squares on Hadamard Spaces and the Sphere *

Philipp Grohs Markus Sprecher

December 19, 2014

**Abstract**

We consider the problem of reconstructing an image from noisy and/or incomplete data, where the image/data take values in a metric space $X$ (e.g. $\mathbb{R}$ for grayscale, $S^2$ for the chromaticity component of RGB-images or $SPD(3)$, the set of positive definite $3\times 3$-Matrices, for Diffusion Tensor Magnetic Resonance Imaging (DT-MRI)). We use the common technique of minimizing a total variation (TV) functional $J$. After having defined $J$ for arbitrary metric spaces $X$ we will propose an adaption of the Iteratively Reweighted Least Squares (IRLS) algorithm to minimize $J$. For the case of $X$ being a Hadamard space, such as $SPD(n)$, we prove existence and uniqueness of a minimizer of a regularized functional $J^\varepsilon$ where $\varepsilon > 0$ and show that these minimizers convergence to a minimizer of $J$ when the regularization parameter $\varepsilon$ tends to zero. We show that IRLS can also be applied for $X$ being a half-sphere. For the case of $X$ being a Riemannian manifold we propose to use Newton's method on Manifolds to numerically compute the minimizer of $J^\varepsilon$. To demonstrate our algorithm we present some numerical experiments where we denoise and/or inpaint sphere-valued and SPD-valued images.

**Keywords:** Iteratively reweighted least squares, total variation, regularization, manifold-valued data

## 1 Introduction

In many applications images are corrupted with some kind of noise. The common tasks of denoising is to remove this noise while keeping the main features of the image unchanged. Furthermore the image might not be known everywhere. The task of inpainting is to restore the image at the unknown regions.

### 1.1 Real-valued Images

In 1992, Rudin, Osher and Fatemi revolutionized the field of real-valued image denoising and restoration with their celebrated paper [27]. Their approach is based on minimizing the total variation $TV(u) = \int_\Omega |\nabla u|$ of an image $u\colon \Omega \to \mathbb{R}$ subject to the constraints $\int_\Omega u = \int_\Omega a$ and $\int_\Omega (u-a)^2 = \sigma^2$. Here $a$ denotes the noisy image and $\sigma$ is a given constant. It was the first method which was able to remove noise while still preserving edges. Chambolle and Lions [8] showed that the problem has a unique solution and is equivalent to minimizing the TV-functional

$$J(u) = \frac{1}{2}\|u-a\|_2^2 + \lambda TV(u), \tag{1}$$

*The research of the authors was supported by the Swiss National Fund under grant SNF 140635.

where $\lambda > 0$ depends on $\sigma$. The functional $J$ is especially useful to remove Gaussian noise. For noise types with heavier tail such as Laplacian noise, the functional $u \mapsto \|u-a\|_1 + \lambda TV(u)$ is more appropriate. A discrete version of (1) can be found by using a finite difference discretization of $\nabla u$. For 2D-images we can find two different ways to discretize the total variation norm in the literature, the anisotropic version

$$TV_{aniso}(u) = \sum_{i,j} |u_{i+1,j} - u_{i,j}| + |u_{i,j+1} - u_{i,j}| \tag{2}$$

and the isotropic version

$$TV_{iso}(u) = \sum_{i,j} \sqrt{(u_{i+1,j} - u_{i,j})^2 + (u_{i,j+1} - u_{i,j})^2}. \tag{3}$$

Several methods to find the minimizer of the discrete Versions of (1) have been proposed. For example by using Fenchel duals [7], the alternating direction method of multipliers [33], and Split-Bregman methods [32].

## 1.2 Color Images

After Rudin et al. published their impressive denoising result on grayscale images, a logical next step was to extend the TV method for color images. Various TV functionals ([4, 10, 17, 26] to mention but a few) have been proposed with promising results. Initially, the two linear color models (the channel-by-channel and the vectorial model) were examined before also nonlinear models for color image were studied ([10, 23, 28, 30]). One of them, the chromaticity-brightness model, turned out to be closer to human perception than others. Chang and Kang [10] were able to present persuasive numerical results. Nevertheless, linear models were in favor for quite a while, presumably due to their simplicity in comparison to the nonlinear models.

## 1.3 Manifold-valued Images

Images taking values in a manifold appear naturally in various signal and image processing applications. The most prominent example is DT-MRI. There are several proposals how to regularize DT-MRI images [11, 12, 21]. Recently there has also been made progress in extending TV regularization to arbitrary Riemannian manifolds. In [19] Lellmann, Strekalovskiy, Kötter and Cremers presented a first framework and an algorithmic solution for TV regularization for arbitrary Riemannian manifolds. Their idea is to reformulate the variational problem as a multilabel optimization problem which then can be solved approximately by convex relaxation techniques. They mentioned that "Many properties of minimizers of total variation-regularized models in manifolds are still not fully understood". In [31] Weinmann, Demaret and Storath considered a generalization of the anisotropic (see Equation (2)) version of the discrete total variation to Riemannian manifolds. They propose a proximal point algorithm to minimize their TV-functional and prove convergence for data taking values in Hadamard spaces. Convergence results for spaces which are not Hadamard is still an open problem.

## 1.4 Total Variation on Arbitrary Metric Spaces

We define the TV-functional for any metric space $(X,d)$. Let $V$ be an index set of all the pixels of our image (usually a two-dimensional grid). Our generalizations of the the total variation are

$$TV_{aniso}(u) := \sum_{(i,j)\in E} d(u_i,u_j) \quad \text{and} \quad TV_{iso}(u) := \sum_{i\in V} \sqrt{\sum_{j\in n(i)} d^2(u_i,u_j)}, \tag{4}$$

where $E \subset V \times V$ is a set of edges and $n(i) := \{j \in V | (i,j) \in E\}$. Let $a\colon V_k \subset V \to X$ be the given noisy image. Whenever possible we will drop the subscripts *aniso* and *iso* to indicate that the statement holds for both cases. We can now define the generalized TV-functional $J\colon X^V \to \mathbb{R}$ by

$$J(u) := \frac{1}{2}\sum_{i\in V_k} d^2(u_i,a_i) + \lambda TV(u). \tag{5}$$

Throughout this document we will assume that $(V,E)$ is connected and $V_k$ is non-empty. However if $(V,E)$ is not connected and $V_k$ contains at least one vertex of every connected component of $V$ we can write $J$ as a sum of independent subfunctionals and the statements remain true.

## 1.5 Our Contribution

We present an adaption of the Iteratively Reweighted Least Squares (IRLS) algorithm to minimize the TV-functional. IRLS has been proven to be very successful for recovering sparse signals [13] and was already applied to scalar TV-minimization problems in [25]. We will prove convergence to a global minimizer in the case of Hadamard spaces. Using new arguments and a result from differential topology, we show that our algorithm is also applicable for the important case of $X$ being a half-sphere. This result is of independent interest and can also be seen as a first step towards a theory of convergence for spaces which are not Hadamard. The fact that the IRLS algorithm can be interpreted as an alternating minimization algorithm [13] will be very useful to obtain our results.

## 1.6 Outline

We have structured the article as follows. In Chapter 2 we define our adaption of the IRLS algorithm. It generates a sequence for which the value of the TV-functional is decreasing. Under some conditions on $J$ we prove that this sequence converges to a minimizer of a regularized functional $J^\varepsilon$. In Chapter 3 we present the necessary theory of Hadamard spaces and prove that if $X$ is a Hadamard space then $J$ is convex and satisfies the conditions from Chapter 2. We also prove that the regularized functional $J^\varepsilon$ has a unique minimizer and these minimizers converge to a minimizer of $J$ when $\varepsilon$ tends to zero. Our algorithm for TV-minimization for $X$ being a Riemannian manifold is presented in Chapter 4. In Chapter 5 we show that if $X$ is a half sphere the functional of the optimization problem occurring in the IRLS algorithm has a unique critical point. Finally in Chapter 6 we present details on the implementation and some numerical experiments.

# 2 Iteratively Reweighted Least Squares

## 2.1 Definition of IRLS

The IRLS algorithm [13] alternates between reweighting and minimization steps. The generalization of the algorithm is straightforward. Let $\varepsilon > 0$, the minimization and reweighting steps of the anisotropic case are

$$u^{new} = \operatorname*{argmin}_{u \in X^V} \sum_{i \in V_k} d^2(a_i, u_i) + \lambda \sum_{(i,j) \in E} w_{i,j} d^2(u_i, u_j) \quad \text{and} \tag{6}$$

$$w_{i,j}^{new} = W_{aniso}^{\varepsilon}(u)_{i,j} := (d^2(u_i, u_j) + \varepsilon^2)^{-\frac{1}{2}} \quad \text{for all } (i,j) \in E. \tag{7}$$

Analogously the steps of the isotropic case are

$$u^{new} = \operatorname*{argmin}_{u \in X^V} \sum_{i \in V_k} d^2(a_i, u_i) + \lambda \sum_{i \in V} w_i \sum_{j \in n(i)} d^2(u_i, u_j) \quad \text{and} \tag{8}$$

$$w_i^{new} = W_{iso}^{\varepsilon}(u)_i := \left( \sum_{j \in n(i)} d^2(u_i, u_j) + \varepsilon^2 \right)^{-\frac{1}{2}} \quad \text{for all } i \in V. \tag{9}$$

The parameter $\varepsilon > 0$ is necessary to avoid division by zero and is typically chosen very small. In the literature [15] we also find the alternative reweighting defined by

$$w_{i,j}^{new} = \max\left(\varepsilon, \max(d(u_i, u_j), \varepsilon)^{-1}\right)$$

in the anisortropic case. However this reweighting seems to be more difficult to analyze and we will therefore stick to the reweightings given in Equation (7) and (9). Note that for Euclidean spaces we have to minimize a quadratic function in the minimization step. This can be done by solving a linear system which is where the term least squares of IRLS comes from.

## 2.2 Analysis of IRLS

To analyze IRLS we use ideas of [9]. We define the regularized functionals $J_{aniso}^{\varepsilon}$ and $J_{iso}^{\varepsilon}$ by

$$J_{aniso}^{\varepsilon}(u) := \frac{1}{2} \sum_{i \in V_k} d^2(u_i, a_i) + \lambda \sum_{(i,j) \in E} \sqrt{d^2(u_i, u_j) + \varepsilon^2}, \text{ and} \tag{10}$$

$$J_{iso}^{\varepsilon}(u) := \frac{1}{2} \sum_{i \in V_k} d^2(u_i, a_i) + \lambda \sum_{i \in V} \sqrt{\sum_{j \in n(i)} d^2(u_i, u_j) + \varepsilon^2}. \tag{11}$$

**Remark 1.** *Although the functional $J^{\varepsilon}$ is smooth when $X$ is a smooth Riemannian manifold the performance of standard methods to find its minimizer (e.g. Newton) is poor. A possible explanation for this observation is that $J^{\varepsilon}$ is highly nonlinear for $\varepsilon$ small.*

To view IRLS as an alternating minimization technique we define the two variable functionals

$\tilde{J}^\varepsilon_{aniso}\colon \mathbb{R}^E_{>0}\times X^V\to\mathbb{R}$ and $\tilde{J}^\varepsilon_{iso}\colon \mathbb{R}^V_{>0}\times X^V\to\mathbb{R}$ by

$$\tilde{J}^\varepsilon_{aniso}(w,u) := \frac{1}{2}\sum_{i\in V_k} d^2(u_i,a_i)+\frac{1}{2}\lambda\sum_{(i,j)\in E} w_{i,j}(d^2(u_i,u_j)+\varepsilon^2)+w_{i,j}^{-1}, \text{ and} \tag{12}$$

$$\tilde{J}^\varepsilon_{iso}(w,u) := \frac{1}{2}\sum_{i\in V_k} d^2(u_i,a_i)+\frac{1}{2}\lambda\sum_{i\in V} w_i\left(\sum_{j\in n(i)} d^2(u_i,u_j)+\varepsilon^2\right)+w_i^{-1}. \tag{13}$$

We now show a connection between the critical points of $\tilde{J}^\varepsilon$ and $J^\varepsilon$.

**Lemma 2.** *The pair $(w,u)$ is a critical point of $\tilde{J}^\varepsilon$ if and only if $u$ is a critical point of $J^\varepsilon$ and $w=W^\varepsilon(u)$ where $W^\varepsilon$ is defined in* (7) *for the anisotropic and in* (9) *for the isotropic case.*

*Proof.* Note that

$$\tilde{J}^\varepsilon(w,u) = \frac{1}{2}\sum_{i\in V_k} d^2(u_i,a_i)+\frac{1}{2}\lambda\sum_i w_i W^\varepsilon(u)_i^{-2}+w_i^{-1} \tag{14}$$

$$J^\varepsilon(u) = \frac{1}{2}\sum_{i\in V_k} d^2(u_i,a_i)+\lambda\sum_i W^\varepsilon(u)_i^{-1}. \tag{15}$$

Hence we have

$$\begin{aligned}\tfrac{\partial\tilde{J}^\varepsilon}{\partial w_i}(w,u) &= \tfrac{1}{2}\lambda(W^\varepsilon(u)_i^{-2}-w_i^{-2}),\\ \tfrac{\partial\tilde{J}^\varepsilon}{\partial u}(w,u)\Big|_{w=W^\varepsilon(u)} &= \tfrac{1}{2}\textstyle\sum_{i\in V_k}\tfrac{\partial d^2(u_i,a_i)}{\partial u}-\lambda\sum_i W^\varepsilon(u)_i^{-2}\tfrac{\partial W^\varepsilon(u)_i}{\partial u} \quad = \tfrac{dJ^\varepsilon}{du}(u)\end{aligned}$$

and

$$\frac{\partial\tilde{J}^\varepsilon}{\partial w}(w,u)=0 \text{ and } \frac{\partial\tilde{J}^\varepsilon}{\partial u}(w,u)=0 \quad\Leftrightarrow\quad W^\varepsilon(u)=w \text{ and } \frac{dJ^\varepsilon}{du}(u). \qquad \square$$

The idea of alternating minimization is to fix one variable and to minimize the functional w.r.t. the other variable and vice versa. Consider the sequences $(w^{(j)})_{j\in\mathbb{N}},(u^{(j)})_{j\in\mathbb{N}}$ recursively defined by

$$w^{(j+1)} := \operatorname*{argmin}_w \tilde{J}^\varepsilon(w,u^{(j)}), \text{ and} \tag{16}$$

$$u^{(j+1)} := \operatorname*{argmin}_{u\in X^V} \tilde{J}^\varepsilon(w^{(j+1)},u) \tag{17}$$

Note that Equations (16) and (17) are equivalent to Equation (6) and (7) respectively Equation (8) and (9). Hence alternating minimization of $\tilde{J}^\varepsilon$ is equivalent to IRLS. An immediate consequence is that the value of $J^\varepsilon$ is non-increasing.

**Theorem 3.** *The sequence $\left(J^\varepsilon(u^{(j)})\right)_{j\in\mathbb{N}}$ is non-increasing.*

*Proof.* Note that $J^\varepsilon(u)=\tilde{J}^\varepsilon(W^\varepsilon(u),u)$. Hence the statement follows by the interpretation of the IRLS-algorithm as alternating minimization of $\tilde{J}^\varepsilon$. $\square$

As explained in [24] alternating minimization can fail to converge to a minimizer. However under the assumption that $J^\varepsilon$ has a unique critical point we can prove convergence. We will need the following elementary topological Lemma.

**Lemma 4.** *Let $A$ be a compact space, $B$ a topological space, $f\colon A \to B$ a continuous function, $a \in A$ such that $f(x) = f(a)$ if and only if $x = a$ and $(x^{(i)})_{i\in\mathbb{N}} \subset A$ a sequence with $\lim_{i\to\infty} f(x^{(i)}) = f(a)$. Then $\lim_{i\to\infty} x^{(i)} = a$.*

*Proof.* Assume there exist an open neighborhood $N(a)$ of $a$ such that for infinitely many $i \in \mathbb{N}$ we have $x^{(i)} \notin N(a)$. As $A\backslash N(a)$ is compact there exist a subsequence $\left(x^{(n_i)}\right)_{i\in\mathbb{N}}$ which converges to some $\bar{a} \in U\backslash N(a)$. We now have

$$f(\bar{a}) = \lim_{i\to\infty} f(x^{(n_i)}) = f(a).$$

which contradicts the assumption. □

We are now able to prove the main result of this chapter.

**Theorem 5.** *Assume that*

*i) $J^\varepsilon$ has a unique minimizer.*

*ii) $u \mapsto \tilde{J}^\varepsilon(w,u)$ has a unique minimizer for all $w \in \mathbb{R}^E_{>0}$ $\left(resp. w \in \mathbb{R}^V_{>0}\right)$.*

*Let $u^{(0)} \in X^V$, $w^{(0)} \in \mathbb{R}^E_{>0}$ (resp. $w^{(0)} \in \mathbb{R}^V_{>0}$) and $(w^{(j)}, u^{(j)})$ be defined by Equations* (16) *and* (17)*. Then the sequence $u^{(j)}$ converges to the unique minimizer of $J^\varepsilon$.*

*Proof.* Note that $\tilde{J}^\varepsilon(w^{(j)}, u^{(j)})$ is bounded from below and monotonically decreasing with $j$. Therefore $\tilde{J}^\varepsilon(w^{(j)}, u^{(j)})$ converges to some value $c \in \mathbb{R}$. Note that $\|w^{(j)}\|_{\ell^\infty} \leqslant \varepsilon^{-1}$ and the sequence $(w^{(j)}, u^{(j)})$ is bounded. Therefore it has a subsequence $(w^{(n_j)}, x^{(n_j)})$ converging to some $(\bar{w}, \bar{u})$. Let

$$w' := \operatorname*{argmin}_{w\in\mathbb{R}^V} \tilde{J}^\varepsilon(w, \bar{u}), \quad \text{and} \quad u' := \operatorname*{argmin}_{u\in M^V} \tilde{J}^\varepsilon(w', u).$$

We have

$$c = \lim_{j\to\infty} \tilde{J}^\varepsilon(w^{(n_j+1)}, u^{(n_j+1)}) = \tilde{J}^\varepsilon(w', u') \leqslant \tilde{J}^\varepsilon(w', \bar{u}) \leqslant \tilde{J}^\varepsilon(\bar{w}, \bar{u}) = \lim_{j\to\infty} \tilde{J}^\varepsilon(w^{(n_j)}, u^{(n_j)}) = c.$$

As we have equality and the functions $u \mapsto \tilde{J}^\varepsilon(w', u)$ and $w \mapsto \tilde{J}^\varepsilon(w, \bar{u})$ have unique minimizers we have $w' = \bar{w}$ and $u' = \bar{u}$. It follows that $\bar{w} = W(\bar{u})$ and $(\bar{u}, \bar{w})$ is a critical point of $\tilde{J}^\varepsilon$. Therefore by Lemma 2 $\bar{u}$ is the unique minimizer of $J^\varepsilon$ and $(\bar{w}, \bar{u})$ is the unique minimizer of $\tilde{J}^\varepsilon$. Finally by Lemma 4 we get that $\lim_{j\to\infty}(w^{(j)}, u^{(j)}) = (\bar{w}, \bar{u})$. □

# 3 TV on Hadamard Spaces

We start this chapter by the definition of Hadamard spaces, also known as spaces of non-positive curvature.

**Definition 1.** *A complete metric space $X$ is called a Hadamard space if for all $x_0, x_1 \in X$ there is $y \in X$ such that for all $z \in X$ we have*

$$d^2(z,y) \leqslant \frac{1}{2}d^2(z,x_0) + \frac{1}{2}d^2(z,x_1) - \frac{1}{4}d^2(x_0,x_1). \tag{18}$$

A comprehensive introduction to the nowadays well-established theory of Hadamard spaces is [3]. The space of positive definite matrices is a Hadamard space (see [5], Page 314). An important property of Hadamard spacess is that for any two points there exist a unique geodesic connecting them.

**Definition 2.** *A curve $\gamma\colon [0,1] \to X$, where $X$ is a metric space is called a geodesic if*

$$d(\gamma(t_1),\gamma(t_2)) = |t_1 - t_2| d(\gamma(0),\gamma(1)) \quad \text{for all } t_1, t_2 \in [0,1].$$

To prove uniqueness of a minimizer of $J^\varepsilon$ we will need the notion of convexity.

**Definition 3.** *A function $f\colon X \to \mathbb{R}$, where $X$ is a metric space is called convex (resp. strictly convex) if for every non-constant geodesic $\gamma\colon [0,1] \to X$ the function $f \circ \gamma\colon [0,1] \to \mathbb{R}$ is convex (resp. strictly convex).*

Convexity will be important to get information on the set of minimizers.

**Definition 4.** *A set $C \subset X$ is called geodesically convex if for any geodesic $\gamma\colon [0,1] \to X$ with $\gamma(0), \gamma(1) \in C$ we have $\gamma(t) \in C$ for all $t \in [0,1]$.*

**Lemma 6.** *Let $X$ be a Hadamard space. Then the set of minimizers of a convex function $f\colon X \to \mathbb{R}$ is geodesically convex.*

The proof follows immediately from the theory of convex functions on $\mathbb{R}$ and the definition of geodesically convex sets. As in our TV-functional distances, respectively squared distances, occur, it is of interest if $d$, respectively $d^2$, is convex. For Hadamard spaces this is well-known and stated below.

**Theorem 7** (Sturm [29])**.** *Let $X$ be a Hadamard space, $d\colon X \times X \to \mathbb{R}$ its distance function and $y \in X$. Then the function $x \mapsto d^2(x,y)$ is strictly convex.*

The prove follows by repeatedly using Equation (18). Note that if $X$ is a Hadamard space then $X^n$ with the metric

$$d(x,y) := \sqrt{\sum_{i=1}^n d^2(x_i,y_i)} \quad \text{for all} \quad x = (x_i)_{i=1}^n,\ y = (y_i)_{i=1}^n \in X^n,$$

is also a Hadamard space.

**Theorem 8** (Sturm [29], Ballmann (Prop. 5.4. in [3]))**.** *For a Hadamard space $X$ the distance function $d\colon X^2 \to \mathbb{R}$ is convex.*

**Corollary 9.** *If $X$ is a Hadarmard space the functional $J_{aniso}\colon X^V \to \mathbb{R}$ defined in* (5) *is convex.*

*Proof.* This follows from Theorem 7 and Theorem 8. □

To prove convexity of $J_{iso}$ we need the following lemma.

**Lemma 10.** *If $f_1, \dots, f_n\colon [0,1] \to \mathbb{R}_{\geqslant 0}$ are convex functions then the function $\sqrt{f_1^2 + \dots + f_n^2}$ is also convex.*

*Proof.* Note that by induction it suffices to prove the statement for $n = 2$. By convexity of $f_1$ and $f_2$ we have

$$\begin{aligned} f_1(t) &\leqslant t f_1(1) + (1-t) f_1(0) \\ f_2(t) &\leqslant t f_2(1) + (1-t) f_2(0). \end{aligned}$$

Squaring and adding the inequalities yields

$$(f_1^2 + f_2^2)(t) \leqslant t^2 (f_1^2 + f_2^2)(1) + (1-t)^2 (f_1^2 + f_2^2)(0) + 2t(1-t)(f_1(0)f_1(1) + f_2(0)f_2(1)).$$

By Cauchy-Schwarz we get

$$\begin{aligned}(f_1^2+f_2^2)(t) &\leqslant t^2(f_1^2+f_2^2)(1)+(1-t)^2(f_1^2+f_2^2)(0)+2t(1-t)(f_1(0)f_1(1)+f_2(0)f_2(1))\\ &\leqslant t^2(f_1^2+f_2^2)(1)+(1-t)^2(f_1^2+f_2^2)(0)+2t(1-t)\sqrt{f_1^2+f_2^2}(1)\sqrt{f_1^2+f_2^2}(0)\\ &= \left(t\sqrt{f_1^2+f_2^2}(1)+(1-t)\sqrt{f_1^2+f_2^2}(0)\right)^2.\end{aligned}$$

Taking the square root shows that the function $\sqrt{f_1^2+f_2^2}$ is convex. □

**Corollary 11.** *If $X$ is a Hadarmard space the functional $J_{iso}\colon X^V \to \mathbb{R}$ defined in* (5) *is convex.*

*Proof.* This follows from Theorem 7, Theorem 8 and Lemma 10. □

**Corollary 12.** *If $X$ is a Hadamard space the set of minimizers of $J$ is geodesically convex.*

*Proof.* This is a consequence of the convexity of $J$ and Lemma 6. □

To prove that the functionals of the optimization problems (6) and (8) are convex we need to prove that the function $(x,y)\mapsto d^2(x,y)$ is convex. To do so we use the following lemma.

**Lemma 13.** *Let $f\colon [0,1]\to\mathbb{R}_{\geqslant 0}$ be convex. Then $f^2$ is also convex.*

*Proof.* By convexity we have

$$f(t)\leqslant (1-t)f(0)+tf(1) \text{ for all } t\in[0,1]$$

As $f(t)\geqslant 0$ we can square this inequality to get for every $t\in[0,1]$

$$\begin{aligned}f(t)^2 &\leqslant (1-t)^2f(0)^2+t^2f(1)^2+2t(1-t)f(0)f(1) &(19)\\ &= (1-t)f(0)^2+tf(1)^2-t(1-t)(f(0)-f(1))^2 &(20)\\ &\leqslant (1-t)f(0)^2+tf(1)^2. \quad \square &(21)\end{aligned}$$

**Lemma 14.** *If $X$ is a Hadamard space and $\varepsilon>0$ the functionals $J^\varepsilon$ and $u\mapsto \tilde{J}^\varepsilon(w,u)$ are strictly convex for all $w\in\mathbb{R}^E_{>0}$ respectively $w\in\mathbb{R}^V_{>0}$.*

*Proof.* By Lemma 13 the functionals are convex. To show strict convexity it is enough to prove that for any $u^1,u^2\in X^{V_k}$ with $u^1\neq u^2$ there is one term of the functional for which we have strict inequality. If there exist $i\in V_k$ with $u_i^1\neq u_i^2$ we have a strict inequality by Theorem 7. If there is no $i\in V_k$ with $u_i^1\neq u_i^2$ we can find $(i,j)\in E$ such that $u_i^1=u_i^2$ and $u_j^1\neq u_j^2$. Then by the same argument as before we have strict inequality. □

**Lemma 15.** *Let $u^\varepsilon$ be the unique minimizer of $J^\varepsilon$. Then $\lim_{\varepsilon\to 0}u^\varepsilon$ is well-defined and a minimizer of $J$.*

*Proof.* We prove the statement for $J=J_{aniso}$, the isotropic case can be done similarly. Let $C$ be the set of minimizers of $J$, $E_+=\{(i,j)\in E|\exists u=(u_i)_{i\in V}\in C \text{ s.t } d(u_i,u_j)>0\}$ and $E_0=E\backslash E_+$. By Corollary 12 $C$ is geodesically convex. Note that the function $K\colon X^V\to\mathbb{R}$ defined by

$$K(u):=\sum_{(i,j)\in E_+}\frac{1}{d(u_i,u_j)}$$

is strictly convex on $C$. Hence there exist a unique minimizer $u^0 \in C$. We define

$$K^\varepsilon(u) := \sum_{(i,j)\in E_+} \frac{1}{d(u_i,u_j)+\varepsilon}.$$

We have the inequalities

$$x + \frac{\varepsilon^2}{2(x+\varepsilon)} \leqslant \sqrt{x^2+\varepsilon^2} \leqslant x + \frac{\varepsilon^2}{2x}.$$

Hence

$$J(u^\varepsilon) \leqslant J(u^\varepsilon) + \lambda\varepsilon|E_0| + \frac{\lambda\varepsilon^2}{2}K^\varepsilon(u^\varepsilon) \leqslant J^\varepsilon(u^\varepsilon) \leqslant J^\varepsilon(u^0) \leqslant J(u^0) + \lambda\varepsilon|E_0| + \frac{\lambda\varepsilon^2}{2}K(u^0).$$

It follows that $\lim_{\varepsilon\to 0} J(u^\varepsilon) = J(u^0)$, $K^\varepsilon(u^\varepsilon) \leqslant K(u^0)$ and $d(u_i,u_j) \geqslant (K(u^0))^{-1} - \varepsilon$ for all $(i,j) \in E_+$. As

$$K(u^0) \geqslant K^\varepsilon(u^\varepsilon) = K(u^\varepsilon) - \varepsilon \sum_{(i,j)\in E_+} \frac{1}{d(u_i^\varepsilon,u_j^\varepsilon)(d(u_i^\varepsilon,u_j^\varepsilon)+\varepsilon)} \geqslant K(u^\varepsilon) - \varepsilon|E_+|(K(u^0))^{-2}$$

we have $\lim_{\varepsilon\to 0} K(u^\varepsilon) = K(u^0)$. Convergence of $(u^\varepsilon)_{\varepsilon>0}$ to $u^0$ for $\varepsilon \to 0$ now follows from Lemma 4. □

# 4 Algorithm for Riemannian Manifolds

If $X$ is a Riemannian manifold we propose to use the Riemannian Newton method to solve the optimization problems (6) and (8). In the last decade several opimization algorithms, such as Newton's method, have been generalized to optimization problems on manifolds. More details on optimization on Manifolds can be found in [1].

## 4.1 Gradient, Hessian and Newton's Method on Manifolds

For a Riemannian manifold $M$ we denote by $T_xM$ the tangential space at $x \in M$ and by $\exp_x \colon T_xM \to M$ the exponential mapping at $x \in M$, i.e. $\exp_x(v) := \gamma(1)$ where $\gamma$ is the unique geodesic with $\gamma(0) = x$ and $\dot\gamma(0) = v$. If $f \colon M \to \mathbb{R}$ is 2-times differentiable at $x \in M$ there exists $G \in T_xM$ and a self-adjoint operator $H \colon T_xM \to T_xM$ such that

$$f(\exp_x(v)) = f(x) + \langle G, v\rangle + \frac{1}{2}\langle v, H(v)\rangle + \mathcal{O}(|v|^3). \tag{22}$$

We call $G$ and $H$ the Riemannian gradient and Hessian and denote it by $\operatorname{grad} f(x)$ and $\operatorname{Hess} f(x)$. Equation (22) can be seen as a Taylor expansion of the manifold valued function $f$ at $x \in M$.

**Remark 16.** *We also mention that $f$ is convex at $x \in M$ if and only if $H$ is positive semidefinite, i.e. $\langle v, H(v)\rangle \geqslant 0$ for all $v \in T_xM$.*

Riemmanian Newton's method for a function $f \colon M \to \mathbb{R}$ creates a sequence $(x_i)_{i\in\mathbb{N}} \subset M$ defined by $x_{i+1} := \phi(x_i)$ where

$$\phi(x) := \exp_x\left(-(\operatorname{Hess} f(x))^{-1}(\operatorname{grad} f(x))\right).$$

In [1] it was shown that Riemannian Newtons's method converges, under similar assumptions as in the linear theory, locally quadratic to a critical point.

### 4.2 The Algorithm

If our graph $(V,E)$ is sparse the Hessian of the functional in the optimization problem (6) resp. (8) will be sparse as well, which will alllow us to solve a Newtonstep in moderate time. Our algorithm is shown below.

**Algorithm 1** TV-minimization

**Input:** Graph $(V,E)$, noisy image $a \in M^{V_k}$ where $V_k \subset V$ and parameters $\lambda, \varepsilon, tol > 0$.
**Output:** Approximation for the minimizer $u \in M^V$ of $J^\varepsilon(Inois, \lambda, E)$.

```
Choose a first guess u^(0) for u (e.g. u^(0) = a)
Set i = 0
repeat
  W = W^ε(u^(i))
  u^(i,0) = u^(i)
  k = 0
  repeat
    u^(i,k+1) = newtonstep(W, u^(i,k), λ, a, V, E)
    k = k + 1
  until d(u^(i,k-1), u^(i,k)) < tol
  u^(i+1) = u^(i,k)
  i = i + 1
until d(u^(i-1), u^(i)) < tol
return u^(i)
```

It is observed in practice that the number of newtoniterations necessary in one IRLS-step is very small ($< 10$). However for large data sets just one newtonstep can be expensive and we would like to perform as few newtonsteps as possible. For Euclidean data we can restrict the number of newtoniterations in each IRLS-step to one. A natural improvement of our algorithm for non-Euclidean data is to do a reweighting after each newtoniteration. Even though convergence is no longer guaranteed it is observed that we still have convergence in practice. Another option is to do newtoniterations until the value of $J^\varepsilon$ is smaller than before the first newtoniteration (this usually happens after one newtonstep). This way we can still guarantee convergence while reducing the computational cost.

## 5 TV on the Sphere

An important application of TV-regularization are *RGB*-images $u\colon V \to [0,1]^3$, where each component contains the information on how much of the corresponding color the pixels contain. A common way to deal with *RGB*-images is to separate them into the brightness ($V \to \mathbb{R}, i \mapsto \|u_i\|_2$) and the chromaticity component ($V \to S^2, i \mapsto u_i/\|u_i\|_2$) and deal with them independently. The values of the chromaticity component are on $S^2$. Unfortunately, this is not a Hadamard manifold and we can not apply the theory of Chapter 3. In fact the TV-functional does in general not have a unique solution as the following example shows

**Example 1.** *Consider the TV-functional $J$ associated to $V = V_k = \{1,2\}$, $E = \{(1,2)\}$, $a_1 = (1,0,0)$, $a_2 = (-1,0,0)$ and $\lambda > 0$. Note that any minimizer $(u_1,u_2)$ of $J$ satisfies $u_i \neq a_j$ for any $i,j \in \{1,2\}$. By rotational symmetry of the sphere there can not be a unique minimizer.*

The example above is a special case where the points lie opposite to each other and this raises the

question if we have uniqueness if all the points lie 'close' to each other. For RGB-images, we can restrict our space to

$$S^2_{\geqslant 0} := \{x = (x_i)_{i=1}^3 \in S^2 | x_i \geqslant 0 \text{ for } i \in \{1,2,3\}\} \subset \{x \in S^2 | \sum_{i=1}^3 x_i > 0\}.$$

Using new arguments and the Poincare-Hopf Theorem we will show in Section 5.1 that for data on a half-sphere the optimization problems (6) and (8) have a unique critical point and we can therefore guarantee to find the minimizer using Riemannian Newton's method. Hence we can apply IRLS also for $S^2_{\geqslant 0}$-valued images. By Theorem 3 the sequence of functional values are non-increasing. However there are still some open questions: It is unclear, if Assumption i) in Theorem 5 holds true, i.e. if $J^\varepsilon$ has a unique minimizer. Furthermore it is not clear if these minimizers convergence to a minimizer of $J$ if $\varepsilon \to 0$. If we replace the squared distances with distances in the functional $J$ defined in (5) the optimization problem becomes a multifacility location problem. This problem has been studied in [2, 14] however also in this setting existence and uniqueness of a minimizer is an open problem.

## 5.1 Existence and Uniqueness of the Critical Point

To show that the optimization problems (6) and (8) for $a = (a_i)_{i \in V_k} \in (HS^m)^{V_k}$, where $HS^m := \{x = (x_i)_{i=1}^{m+1} \in S^m | x_{m+1} > 0\}$ is the open half sphere, have a unique critical point in $(HS^m)^V$ we first observe that these functionals are of the form

$$J(u) = \sum_{i \in V_k} d^2(a_i, u_i) + \sum_{(i,j) \in E} W_{i,j} d^2(u_i, u_j), \tag{23}$$

with weights $W_{i,j} \in \mathbb{R}^E_{>0}$. Here $d$ denotes the spherical distance. Even though $J$ itself is not convex we will prove in Lemma 19 that $J$ is locally convex at every critical point of $J$. By Remark 16 this is equivalent to the Hessian being positive definite. In Lemma 20 we will show that the gradient grad$J$ is pointing outward at the boundary of $(HS^m)^V$. Then we can use the Poincare-Hopf Theorem well-known in differential topology to show that there is a unique critical point which is also the unique minimizer.
By the cosine formula for the dot product we have $d(x,y) = \arccos(x^T y)$ for all $x, y \in S^m$. The derivatives of $\arccos^2$ are

$$(\arccos^2)'(x) = \begin{cases} \frac{-2\arccos(x)}{\sqrt{1-x^2}} & x \in (-1,1) \\ -2 & x = 1 \end{cases} \quad \text{and} \quad (\arccos^2)''(x) = \begin{cases} \frac{2+(\arccos^2)'(x)x}{1-x^2} & x \in (-1,1) \\ \frac{2}{3} & x = 1 \end{cases}. \tag{24}$$

Let $\alpha \colon [-1,\infty) \to \mathbb{R}$ be a $C^2$ extension of $\arccos^2$ and $f \colon \{(x,y) \in (\mathbb{R}^{m+1})^2 | x^T y > -1\}^2 \to \mathbb{R}$ be defined by $f(x,y) := \alpha(x^T y)$. The Taylor expansion of $f$ of order 2 is

$$f(x+dx, y+dy) = f(x,y) + \begin{pmatrix} dx^T & dy^T \end{pmatrix} \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix} + \frac{1}{2} \begin{pmatrix} dx^T & dy^T \end{pmatrix} \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{pmatrix} \begin{pmatrix} dx \\ dy \end{pmatrix} + \mathcal{O}(|dx|^3 + |dy|^3) \tag{25}$$

with

$$\frac{\partial f}{\partial x} = \alpha' y,\ \frac{\partial f}{\partial y} = \alpha' x,\ \frac{\partial^2 f}{\partial x^2} = \alpha'' y y^T,\ \frac{\partial^2 f}{\partial x \partial y} = \alpha'' x y^T,\ \frac{\partial^2 f}{\partial y \partial x} = \alpha'' y x^T + \alpha' I_{m+1},\ \frac{\partial^2 f}{\partial y^2} = \alpha'' x x^T$$

where $I_{m+1}$ denotes the $(m+1)\times(m+1)$ identity matrix and we have written $\alpha'$ and $\alpha''$ for $\alpha'(x^T y)$ respectively $\alpha''(x^T y)$. The Taylor expansion of the exponential map exp of order 2 yields

$$\exp(x,r) = \cos(|r|)x + \frac{\sin(|r|)}{|r|}r = x + r - \frac{r^T r}{2}x + \mathscr{O}(|r|^3). \tag{26}$$

Let $r = \log_x(x+dx)$ and $s = \log_y(y+dy)$. Replacing $dy$ and $dz$ by $\exp(y,r)-y$ respectively $\exp(z,s)-z$ in Equation (25) and using Equation (26) yields

$$\begin{aligned} f(\exp(x,r),\exp(y,s)) &= f(x,y) + \begin{pmatrix} r^T & s^T \end{pmatrix} \begin{pmatrix} \frac{\partial f}{\partial x} \\ \frac{\partial f}{\partial y} \end{pmatrix} \\ &\quad + \frac{1}{2}\begin{pmatrix} r^T & s^T \end{pmatrix} \begin{pmatrix} \frac{\partial^2 f}{\partial x^2} - x^T \frac{\partial f}{\partial x} I_{m+1} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} - y^T \frac{\partial f}{\partial y} I_{m+1} \end{pmatrix} \begin{pmatrix} r \\ s \end{pmatrix} \\ &\quad + \mathscr{O}(|r|^3 + |s|^3). \end{aligned}$$

The Taylor expansion of $d^2(x,y)$ is therefore

$$\begin{aligned} d^2(\exp_x(r),\exp_y(s)) &= d^2(x,y) + \alpha'(x^T y)(y^T r + x^T s) && (27)\\ &\quad + \frac{1}{2}\alpha''(x^T y)(y^T r + x^T s)^2 && (28)\\ &\quad + \frac{1}{2}\begin{pmatrix} r^T & s^T \end{pmatrix} \begin{pmatrix} -\beta(x^T y)I & \alpha'(x^T y)I \\ \alpha'(x^T y)I & -\beta(x^T y)I \end{pmatrix} \begin{pmatrix} r \\ s \end{pmatrix} && (29)\\ &\quad + \mathscr{O}(|r|^3 + |s|^3), && (30) \end{aligned}$$

for any $x, y \in S^m$, $r \in T_x M$ and $s \in T_y M$ where $\beta(x) := x\alpha'(x)$. For $u \in (HS^m)^V$ we define the matrix $T(u) \in \mathbb{R}^{V\times V}$ by

$$T(u)_{ij} := \begin{cases} -1_{V_k}(i)\beta(a_i^T u_i) - \sum_{k\in n(i)} W_{i,k}\beta(u_i^T u_k) & i = j \\ W_{i,j}\alpha'(u_i^T u_j) & (i,j) \in E \\ 0 & \text{otherwise} \end{cases},$$

where $1_{V_k}$ denotes the indicator function of $V_k$.

**Lemma 17.** *If the matrix $T(u)$ is positive definite the functional $J$ is strictly convex at $u$.*

*Proof.* The term (28) is non-negative. Adding up the second part (29) of the Hessian for all the terms in the functional $J$ yields the matrix $T(u)$. □

**Lemma 18.** *Let $u \in (HS^m)^V$ be a critical point of $J$. Then we have*

$$-1_{V_k}(i)\alpha'(u_i^T a_i)a_i = (T(u)u)_i \quad \textit{for all} \quad i \in V.$$

*Proof.* Let $u$ be a critical point of $J$ and $\bar{J}$ the natural extension of $J$. Then for all $i \in V$ there exist $\mu_i \in \mathbb{R}$ s.t.

$$\mu_i u_i = \frac{d\bar{J}}{du_i} = 1_{V_k}(i)\alpha'(u_i^T a_i)a_i + \sum_{j\in n(i)} W_{i,j}\alpha'(u_i^T u_j)u_j. \tag{31}$$

Multiplying Equation (31) with $u_i$ yields

$$\mu_i = 1_{V_k}(i)\beta(u_i^T a_i) + \sum_{j\in n(i)} W_{i,j}\beta(u_i^T u_j).$$

Therefore for all $i \in V$ we have

$$\left(1_{V_k}(i)\beta(u_i^T a_i) + \sum_{j\in n(i)} W_{i,j}\beta(u_i^T u_j)\right) u_i = 1_{V_k}(i)\alpha'(u_i^T a_i)a_i + \sum_{j\in n(i)} W_{i,j}\alpha'(u_i^T u_j)u_j,$$

which can be rewritten as the desired equation. □

**Lemma 19.** *Every critical point of $J$ is a local minimizer of $J$.*

*Proof.* Let $e_{m+1} := (\underbrace{0 \ \dots \ 0}_{m \text{ times}} \ 1)$ and $r := (e_{m+1}^T u_i)_{i\in V} = (u_i)_{m+1} \in \mathbb{R}^V_{>0}$. By Lemma 18 we have

$$(T(u)r)_i = e_{m+1}^T (T(u)u)_i = -1_{V_k}(i)\alpha'(u_i^T a_i)(a_i)_{m+1} \geqslant 0 \quad \text{for all} \quad i \in V. \tag{32}$$

It follows that $M := diag(r)T(u)diag(r)$ is diagonally dominant and therefore positive semidefinite. We now prove that $M$ is positive definite. Assume that $v \in \mathbb{R}^V$ is an eigenvector of $M$ with eigenvalue 0 and let $i \in V$ such that $|v_i| \geqslant |v_j|$ for all $j \in V$. As $(Mv)_i = 0$ it follows that $v_j = v_i$ for all $j \in n(i)$ and recursively that $v_j = v_i$ for all $j \in V$. Let $j \in V_k$ then

$$(Mv)_j = r_j(T(u)r)_j v_j = -r_j\alpha'(u_j^T a_j)(a_j)_{m+1}v_j \neq 0,$$

a contradiction. Hence $M$ is positive definite and therefore $T(u)$ is positive definite as well. □

We now prove that the gradient of $J$ at the boundary

$$\delta\left(HS^m\right)^V = \{u \in \left(S^m\right)^V \,|\, (u_i)_{m+1} \geqslant 0 \text{ for all } i \in V,\ \exists i \in V\ (u_i)_{m+1} = 0\},$$

of $\left(HS^m\right)^V$ is pointing outwards.

**Lemma 20.** *Let $u = (u_i)_{i\in V} \in \delta\left(HS^m\right)^V$. Then we have $(grad_{u_i}J(u))_{m+1} \leqslant 0$ for all $i \in V$ and there exists $i \in V$ with $(grad_{u_i}J(u))_{m+1} < 0$.*

*Proof.* We have

$$\begin{aligned}
(\mathrm{grad}_{u_i}J(u))_{m+1} &= e_{m+1}^T \mathrm{grad}_{u_i}J(u) && (33)\\
&= e_{m+1}^T P_{T_{u_i}M}\mathrm{grad}_{u_i}\bar{J}(u) && (34)\\
&= e_{m+1}^T \mathrm{grad}_{u_i}\bar{J}(u) && (35)\\
&= 1_{V_k}(i)\alpha'(u_i^T a_i)e_{m+1}^T a_i + \sum_{j\in n(i)} W_{i,j}\alpha'(u_i^T u_j)e_{m+1}^T u_j && (36)\\
&\leqslant 0. && (37)
\end{aligned}$$

Let $j \in V$ such that $u_j \in \delta HS^m$. Consider a path $j = j_0, j_1, \dots, j_l \in V$ from $j$ to a $j_l \in V_k$. Let $i$ be the largest index of this path such that $u_i \in \delta HS^m$. Then we have strict inequality above. □

To prove that $J$ has a unique critical point we will use the classical Poincare-Hopf theorem.

**Theorem 21** (Poincare-Hopf [20])**.** *Let $M$ be a compact manifold with boundary $\delta M$ and $U\colon \bar{M} \to TM$ a vector field on $\bar{M}$ such that $U$ is pointing outward on $\delta M$. Assume that $U$ has a continuous derivative $DU$, all zeros of $U$ are isolated and $DU(z)$ is invertible for all zeros $z \in M$ of $U$. Then $U$ has finitely many zeros $z_1, \dots, z_n \in M$ and*

$$\sum_{i=1}^{n} sign(\det(DU(z_i))) = \chi(M),$$

*where $\chi(M)$ is the euler characteristic of $M$.*

We are now able to proof the main result of this Chapter.

**Theorem 22.** *The functional $J$ defined in 23 has a unique critical point.*

*Proof.* Consider the gradient vector field grad$J$ of $J$. By Lemma 19 all zeros $z$ are isolated and $\det(D\text{grad}J(z)) > 0$. By Lemma 20 the vector field grad$J$ is pointing outward at $\delta\,(HS^m)^V$. Hence by Theorem 21 the number of zeros of grad$J$ is 1. □

# 6 Numerical Experiments

In Section 6.1 and 6.2 we will first describe how to compute the second derivative of the squared distance function on $S^m$ and $SPD(n)$. The application of Algorithm 1 is then straight forward. We used the anisotropic version of the TV-functional. It would also be possible to implement the isotropic version. The numerical experiments were conducted on a laptop using a single 2.9 GHz Intel Core i7 processor.

## 6.1 Sphere-Valued Functions

From Section 5.1 we have

$$\begin{aligned} d^2(\exp_x(r), \exp_y(s)) &= d^2(x,y) + \alpha'(y^T r + x^T s) && (38)\\ &\quad + \frac{1}{2}\begin{pmatrix} r^T & s^T \end{pmatrix}\left(\alpha''\begin{pmatrix} yy^T & yx^T \\ xy^T & xx^T \end{pmatrix} + \begin{pmatrix} -\beta I & \alpha' I \\ \alpha' I & -\beta I \end{pmatrix}\right)\begin{pmatrix} r \\ s \end{pmatrix} && (39)\\ &\quad + \mathcal{O}(|r|^3 + |s|^3), && (40) \end{aligned}$$

for any $x, y \in S^m$, $r \in T_xM$ and $s \in T_yM$ where $\alpha = \alpha(x^T y) = \arccos^2(x^T y)$ and $\beta = \beta(x^T y) := x^T y \alpha'(x^T y)$. The sum of the two matrices given in (39) defines not only a linear map $T_xS^m \times T_yS^m \to T_xS^m \times T_yS^m$ but a linear map $\mathbb{R}^{m+1} \times \mathbb{R}^{m+1} \to \mathbb{R}^{m+1} \times \mathbb{R}^{m+1}$. To restrict ourselves to the tangent space at $(x,y) \in S^m \times S^m$ an orthogonal basis of $T_xS^m$ and $T_yS^m$ using $QR$-decomposition are constructed. By transformation of basis one can compute the gradient and Hessian of $d^2$ w.r.t. to the new basis. Hence we can compute second derivatives of squared distance functions and therefore also of $J^\varepsilon(w,u)$ given by Equation (12) and (13). An algorithm for TV-regularization on spheres following the explanations above was implemented by Manuel Cavegn as part of his Master thesis [6].

In Figure 1 on Page 15 the result of this TV-regularization algorithm to a $359 \times 361$ color-image can be seen. The noisy image was generated from the original image with gaussian noise of variance 0.05. The noisy image $a\colon V \to \mathbb{R}^3$ was first decomposed into its chromaticity ($V \to S^2, i \mapsto a_i/\|a_i\|_2$) and brightness

$(V \to \mathbb{R}, i \mapsto \|a_i\|_2)$ components. Then the two components were denoised independently with $\lambda_{br} = 0.3$, $\lambda_{cr} = 0.5$ and 5 IRLS-steps. Finally the two denoised components were combined to the denoised image. The computation took about 30 seconds. The majority of the computational time is used to solve the linear systems of the newtonsteps.

To compare our algorithm with existing algorithms we did the same denoising experiment with a variation of the proximal point algorithm proposed in [31]. However, instead of applying the steps sequentially we apply them at the same time for all pixels. This version is faster (at least in Matlab) and allowed us to denoise also large images in reasonable time. However it is observed that in this adaption it is necessary to choose smaller $\lambda$'s, we chose $\lambda_{br} = .01$ and $\lambda_{cr} = .02$. Note that the proximal point algorithm minimizes the anisotropic total variation functional whereas our implementation minimizes the isotropic total variation functional. We computed 138 Iterations. The computation took again about 30 seconds, however there would be possibilities to accelerate the computation (e.g. parallelization). It is hard to see any difference between the two restored images, the infinity norm of the difference is 0.05.

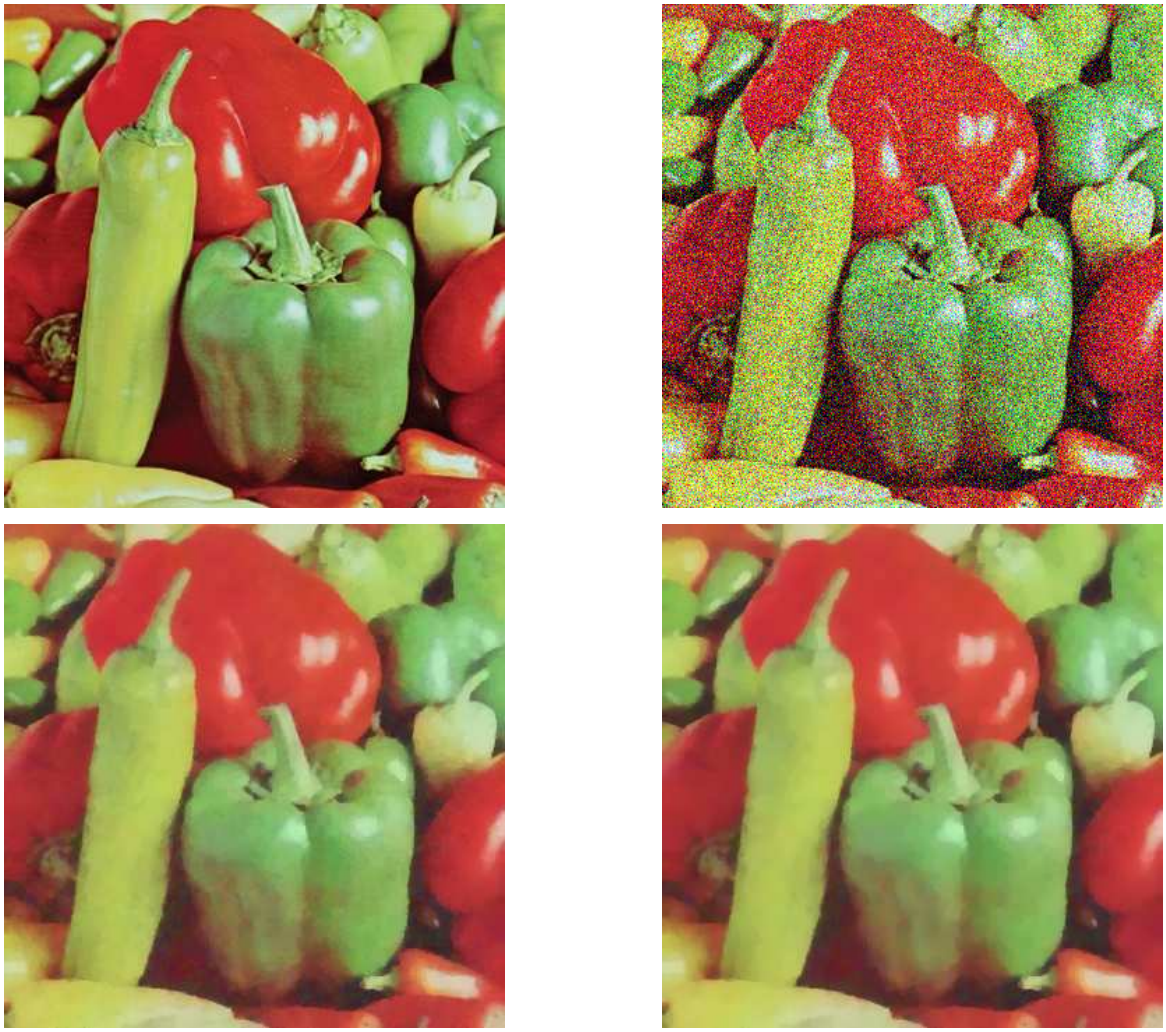

Figure 1: Denoising of a Color-image. The original image and the noisy image are on top. In the bottom the denoised images with the IRLS-algorithm (left) and the proximal point algorithm (right) are shown.

In Figure 2 we see an example of colorization applied to an image known among the image processing community as Lena. We assume that we know the brightness but the color of every pixel is only known with probability 0.01. We first detect the edges from the grayscale-image using a canny-edge detector. We then compute the color part by minimizing a weighted (according to the edges) TV-functional with small $\lambda$. We were not able to get a comparable result using the proximal point algorithm.

Figure 2: Colorization. From left to right: original image, image when 99% of the color was removed and restored image.

## 6.2 SPD-Valued Functions

The set of $n \times n$ positive definite matrices $SPD(n)$, as a Riemannian manifold, is the most studied example of manifolds of non-positive curvature and also the space used for DT-MRI. A detailed analysis of the geometry of this space can be found in [21]. For each positive definite matrix $X$ the tangent space $T_X SPD(n)$ at $X$ can be identified by the space of symmetric matrices. The inner product on $T_X SPD(n)$ given by

$$\langle R, S \rangle = \operatorname{trace}(X^{-1} R X^{-1} S)$$

defines a Riemannian metric on $SPD(n)$. This Riemannian metric induces the metric $d\colon SPD(n) \times SPD(n) \to \mathbb{R}_{\geqslant 0}$ given by

$$d(X,Y) = \|Log(X^{-1/2} Y X^{-1/2})\|_F, \tag{41}$$

where $Log$ denotes the matrix logarithm and $\|\cdot\|_F$ the Frobenius norm. For the computation of the derivatives of the squared distance we need the following result.

**Theorem 23** (Karcher [18])**.** *Let $M$ be a complete Riemannian manifold and $x, y \in M$ such that the geodesic connecting $x$ and $y$ is unique. Then the squared distance function to $y$ is differentiable at $x$ and we have*

$$\frac{\partial d^2(x,y)}{\partial x}[\cdot] = -2\langle log_x(y), \cdot \rangle_x$$

*where $\langle \cdot, \cdot \rangle_x$ is the Riemannian metric at $x \in M$.*

By [21] the exponential map on the space of SPD-matrices is given by

$$\exp_X(R) = X^{\frac{1}{2}} Exp(X^{-\frac{1}{2}} R X^{-\frac{1}{2}}) X^{\frac{1}{2}}, \tag{42}$$

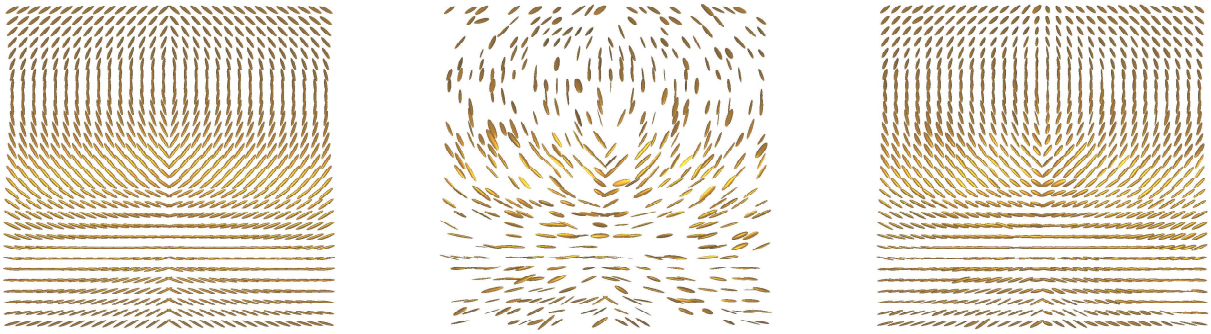

Figure 3: Denoising and Inpainting of a DT-MRI image. From left to right: original, noisy and restored image.

where $Exp$ denotes the matrix exponential. Hence the logarithm map is given by

$$\log_X(Y) = x^{\frac{1}{2}} Log(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}) X^{\frac{1}{2}}. \tag{43}$$

Therefore

$$\begin{aligned}
\frac{\partial d^2(X,Y)}{\partial X}[\cdot] &= -2\langle \log_X(Y), \cdot \rangle_X && (44)\\
&= -2\langle X^{\frac{1}{2}} Log(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}) X^{\frac{1}{2}}, \cdot \rangle_X && (45)\\
&= -2\langle X^{-1} X^{\frac{1}{2}} Log(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}) X^{\frac{1}{2}} X^{-1}, \cdot \rangle_{I_n} && (46)\\
&= -2\langle X^{-\frac{1}{2}} Log(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}) X^{-\frac{1}{2}}, \cdot \rangle_{I_n} && (47)
\end{aligned}$$

Hence we have explicit expressions for the first derivative. The second derivative can be computed from the expression of the first derivative using chain and product rules, as well as the matrix identities $Log(X^2) = 2Log(X)$ and

$$Log(Y) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!} (Y - I_n)^k$$

for all $Y \in SPD(n)$ such that the spectral radius of $Y - I_n$ is smaller than 1. See [16] for further details. An algorithm for TV-regularization of SPD-matrices following the explanations above was implemented by Mara Nägelin as part of her Bachelor thesis [22]. In Figure 3 on Page 17 we can see denoising and inpainting of an artificial $30 \times 30$ DT-MRI image. The noise is added by adding Gaussian noise to the Cholesky decomposition of the SPD-matrices. In this way it is guaranteed that the resulting matrices are SPD. We computed 6 IRLS-steps with 1 newtonstep each. The computation took 35 seconds.

# References


[1] P. A. Absil, R. Mahony, and R. Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009.

[2] T. Aykin and A. J. B. Babu. Multifacility location problems on a sphere. *International Journal of Mathematics and Mathematical Science*, 10(3), 1987.

[3] W. Ballmann. *Lectures on spaces of nonpositive curvature*, volume 25 of *DMV Seminar*. Birkhäuser Verlag, Basel, 1995.

[4] P. Blomgren and T. F. Chan. Color tv: Total variation methods for restoration of vector valued images. *Institute of Electrical and Electronics Engineers Transaction on Image Processing*, 7:304–309, 1996.

[5] M. R. Bridson and A. Haefliger. *Metric spaces of non-positive curvature*. Springer-Verlag, 1999.

[6] M. Cavegn. Total variation regularization for geometric data. *Master thesis, ETH Zürich*, 2013.

[7] A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1-2):89–97, 2004.

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

[9] T. F. Chan, G. H. Golub, and P. Mulet. A nonlinear primal-dual method for total variation-based image restoration, 1995.

[10] T. F. Chan, S. H. Kang, and J. Shen. Total Variation Denoising and Enhancement of Color Images Based on the {CB} and {HSV} Color Models . *Journal of Visual Communication and Image Representation*, 12(4):422–435, 2001.

[11] O. Christiansen, T.-M. Lee, J. Lie, U. Sinha, and T. F. Chan. Total variation regularization of matrix-valued images. *Internation Journal of Biomedical Imaging*, 2007.

[12] O. Coulon, D. C. Alexander, and S. R. Arridge. Diffusion tensor magnetic resonance image regularization. *Medical Image Analysis*, 8(1):47–67, 2004.

[13] I. Daubechies, R. Devore, M. Fornasier, and C. S. Güntürk. Iteratively reweighted least squares minimization for sparse recovery. *Communications on Pure and Applied Mathematics*, 2008.

[14] U. R. Dhar and J. R. Rao. Domain approximation method for solving multifacility location problems on a sphere. *The Journal of the Operational Research Society*, 33(7):pp. 639–645, 1982.

[15] M. Fornasier. *Theoretical Foundations and Numerical Methods for Sparse Recovery*. De Gruyter, 2010.

[16] P. Grohs, M. Sprecher, and T. Yu. Scattered manifold-valued data approximation. Technical Report 23, Seminar for Applied Mathematics, ETH Zürich, Switzerland, 2014.

[17] S. H. Kang and R. March. Variational models for image colorization via chromaticity and brightness decomposition. *Image Processing, Institute of Electrical and Electronics Engineers Transactions on*, 16(9):2251–2261, 2007.

[18] H. Karcher. Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, 30(5):509–541, 1977.

[19] J. Lellmann, E. Strekalovskiy, S. Koetter, and D. Cremers. Total Variation Regularization for Functions with Values in a Manifold. *Institute of Electrical and Electronics Engineers International Conference on Computer Vision*, 0:2944–2951, 2013.

[20] J. Milnor. *Topology From the Differentiable Viewpoint*. Princeton University Press, 1976.

[21] M. Moakher and M. Zéraï. The riemannian geometry of the space of positive-definite matrices and its application to the regularization of positive-definite matrix-valued data. *Journal of Mathematical Imaging and Vision*, 40(2):171–187, 2011.

[22] M. Nägelin. Total variation regularization for tensor valued images. *Bachelor thesis, ETH Zürich*, 2014.

[23] P. Perona. Orientation diffusions. *Image Processing, Biomedical Imaging Transactions on*, 7(3):457–467, 1998.

[24] M.J.D. Powell. On search directions for minimization algorithms. *Mathematical Programming*, 4(1):193–201, 1973.

[25] P. Rodríguez and B. Wohlberg. An iteratively reweighted norm algorithm for total variation regularization. *Signals, Systems and Computers*, 2006.

[26] L. Rudin and V. Caselles. Image recovery via multiscale total variation. In *in Proceedings of the Second European Conference on Image Processing*, pages 15–7, 1998.

[27] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60(1-4):259–268, November 1992.

[28] N. Sochen, R. Kimmel, and R. Malladi. A general framework for low level vision. *Institute of Electrical and Electronics Engineers Transaction on Image Processing*, 7:310–318, 1997.

[29] Karl-Theodor Sturm. Probability measures on metric spaces of nonpositive curvature. In *Heat kernels and analysis on manifolds, graphs, and metric spaces (Paris, 2002)*, volume 338 of *Contemporary Mathematics*, pages 357–390. American Mathematical Society, Providence, RI, 2003.

[30] B. Tang, G. Sapiro, and V. Caselles. Color image enhancement via chromaticity diffusion. *Institute of Electrical and Electronics Engineers Transactions on Image Processing*, 10:701–707, 2002.

[31] A. Weinmann, L. Demaret, and M. Storath. Total variation regularization for manifold-valued data. *Computing Research Repository*, 2013.

[32] Li Y.-f. and Feng X.-c. The split bregman method for l1 projection problems. *Chinese Journal of Electronics*, 38(11):2471, 2010.

[33] J. Yang, Y. Zhang, and W. Yin. A Fast Alternating Direction Method for TVL1-L2 Signal Reconstruction From Partial Fourier Data. *Institute of Electrical and Electronics Engineers Journal of Selected Topics in Signal Processing*, 4(2):288–297, 2010.

# Recent Research Reports

| Nr. | Authors/Title |
|---|---|

2014-29 H. Rauhut and Ch. Schwab
Compressive sensing Petrov-Galerkin approximation of high-dimensional parametric operator equations

2014-30 M. Hansen
A new embedding result for Kondratiev spaces and application to adaptive approximation of elliptic PDEs

2014-31 F. Mueller and Ch. Schwab
Finite elements with mesh refinement for elastic wave propagation in polygons

2014-32 R. Casagrande and C. Winkelmann and R. Hiptmair and J. Ostrowski
DG Treatment of Non-Conforming Interfaces in 3D Curl-Curl Problems

2014-33 U. Fjordholm and R. Kappeli and S. Mishra and E. Tadmor
Construction of approximate entropy measure valued solutions for systems of conservation laws.

2014-34 S. Lanthaler and S. Mishra
Computation of measure valued solutions for the incompressible Euler equations.

2014-35 P. Grohs and A. Obermeier
Ridgelet Methods for Linear Transport Equations

2014-36 P. Chen and Ch. Schwab
Sparse-Grid, Reduced-Basis Bayesian Inversion

2014-37 R. Kaeppeli and S. Mishra
Well-balanced schemes for gravitationally stratified media

2014-38 D. Schoetzau and Ch. Schwab
Exponential Convergence for hp-Version and Spectral Finite Element Methods for Elliptic Problems in Polyhedra