

A TOTAL VARIATION MINIMIZATION TEMPLATE LIBRARY

Master Thesis

written by
Pascal Debus

supervised by
Markus Sprecher,
Prof. Dr. Philipp Grohs
Seminar for Applied Mathematics
ETH Zurich

October 1, 2015

Contents

1	Introduction	5
1.1	Grayscale images	5
1.1.1	Edge preservation	6
1.1.2	Discretization	6
1.2	Color Images	7
1.3	Manifold-valued Images	7
1.4	Objective and Outline of this work	7
2	Theory	8
2.1	Generalization of the functional	8
2.1.1	The Riemannian distance	8
2.2	Algorithms	9
2.2.1	Proximal Point	9
2.2.2	IRLS	10
2.3	Riemannian Newton method	11
2.3.1	Gradient	11
2.3.2	Hessian	11
2.3.3	Newton equation	11
2.3.4	Newton equation for the TV functional	12
2.3.5	Tangent space restriction	13
2.4	Manifolds	13
2.4.1	Euclidian	13
2.4.2	Sphere S^n	14
2.4.3	Special Orthogonal Group SO(n)	15
2.4.4	Symmetric Positive Definite Matrices SPD(n)	16
2.4.5	Grassmannian Gr(n,p)	18
2.5	Fréchet derivatives of matrix logarithm and square root	21
2.5.1	Derivative of the matrix square root	21
2.5.2	Derivative of the matrix logarithm	22
3	The TVMT Library	23
3.1	Capabilities	23
3.1.1	Supported Manifolds	23
3.1.2	Data	23
3.1.3	Functionals	23
3.1.4	Minimizer	23
3.1.5	Visualizations	24
3.2	Design concepts	24
3.2.1	Goals	24
3.2.2	Levels of parallelization	25
3.2.3	C++ techniques	27

3.3	Components	28
3.3.1	Manifold classes	28
3.3.2	Data class	30
3.3.3	Functional class	31
3.3.4	TV Minimizer class	31
3.3.5	Visualization class	32
3.3.6	Utility functions	34
3.3.7	External Components	34
3.4	Using TVTML	34
3.4.1	Prerequisites	34
3.4.2	Installation	35
3.4.3	Compilation of own projects using CMake	35
3.4.4	Tutorial and typical use cases	35
4	Applications and Numerical Experiments	40
4.1	Image denoising	40
4.1.1	Grayscale	40
4.1.2	Color	40
4.1.3	Inpainting	41
4.1.4	Recolorization	42
4.1.5	Volume images	42
4.2	SO(2) and SO(3) images data	43
4.2.1	Synthetic data	43
4.2.2	Fingerprint orientation data	44
4.2.3	Reconstruction of a dense optical flow field	44
4.3	SPD(3) image data	45
4.3.1	Synthetic data	45
4.3.2	Diffusion Tensor Magnetic Resonance Imaging	46
4.3.3	3D DT MRI data	47
4.4	Comparison IRLS and Proximal Point minimizers	48
4.5	Sensitivity to variations of the starting value	48
5	Outlook	49
5.1	Extensions	49
5.2	Recursive computation on subdomains	49

List of Figures

1.1	Comparison total variation	6
3.1	Calculation using pixel-wise kernels	26
3.2	SIMD parallelization	26
3.3	SO(3) cube visualization	32
3.4	SPD(3) ellipsoid visualization	33
3.5	3D SPD(3) Volume Visualization of a helix	33
3.6	3D Volume image renderer	34
4.1	Color image "Lena" linear vectorial denoising	41
4.2	Large image "mathematicians" linear-vectorial denoising	41
4.3	Denoising linear vectorial	42
4.4	Recolorization	42
4.5	Denoising 3D Grayscale Volume Data	43
4.6	Inpainting of synthetic SO(3) picture	43
4.7	Fingerprint orientation denoising	44
4.8	Dense optical flow reconstruction	45
4.9	Denoising of synthetic SPD(3) picture	46
4.10	Denoising DT-MRI data	47
4.11	Denoising 3D DTI-MRI data	48

List of Algorithms

2.1	Parallel proximal point algorithm	10
2.2	IRLS algorithm	11
2.3	Riemannian Newton method for real-valued functions	12

Chapter 1

Introduction

Various forms of noise occur in many forms of data acquisition, transmission and processing. This noise needs to be removed in order to obtain a meaningful interpretation of the data, to enable further processing or, as in many image processing applications, just for aesthetical reasons. A common everyday example for a noisy image is taking a picture with a digital camera (e.g. integrated in a smart phone) in a weakly illuminated room: Especially the dark areas of the picture are not uniform in color and brightness but have small variations from pixel to pixel.

A noise removal algorithm needs to remove these small variations but at the same time not alter important features of the data. In the case of images important features are for example the edges separating areas of different colors and providing the necessary sharpness of the picture. These edges on the other hand are characterized by large variations. This distinction between small and large variations is also helpful in the task of inpainting, which tries to restore the picture at unknown or damaged regions.

The method of total variation(TV) noise removal, which has the above described capabilities, was first introduced by Rudin, Osher and Fatemi [18] in 1992 for the case of real-valued, that means grayscale images. Their method is briefly summarized in the following section.

1.1 Grayscale images

Let $u_0 : \Omega \subset \mathbb{R} \rightarrow \mathbb{R}$ describe the original, noise-free image, where the image domain Ω is usually a rectangular or cuboid subset of \mathbb{R}^2 or \mathbb{R}^3 , respectively. Assuming the original pictures is corrupted by gaussian noise $n : \Omega \rightarrow \mathbb{R}$ with zero mean and variance σ^2 the noisy picture is given by $u : \Omega \rightarrow \mathbb{R}$, where $u = u_0 + n$. The edge preserving denoising of the picture is then equivalent to the solution $u^* : \Omega \rightarrow \mathbb{R}$ of the following constrained optimization problem:

$$u^* = \operatorname{argmin}_{f: \Omega \rightarrow \mathbb{R}} \int_{\Omega} |\nabla u| \quad \text{s.t.} \tag{1.1}$$

$$\int_{\Omega} (u - u_0) = 0, \quad \text{and} \quad \int_{\Omega} (u - u_0)^2 = \sigma^2 \tag{1.2}$$

The first term $TV(u) = \int_{\Omega} |\nabla u|$ is called the total variation of u . Rudin, Osher and Fatemi then use a partial differential equation (PDE) approach to solve the corresponding Euler-Lagrange equation for (1.1). Later Chambolle and Lions [10] showed that (1.1) is equivalent to the minimization of the functional

$$J(u) = \frac{1}{2} \|u - u_0\|_2^2 + \lambda \int_{\Omega} |\nabla u| \tag{1.3}$$

1.1.1 Edge preservation

A basic intuition why the L^1 norm in ([18]) is better suited for conserving sharp discontinuities such as edges can be seen from the following plot [25].

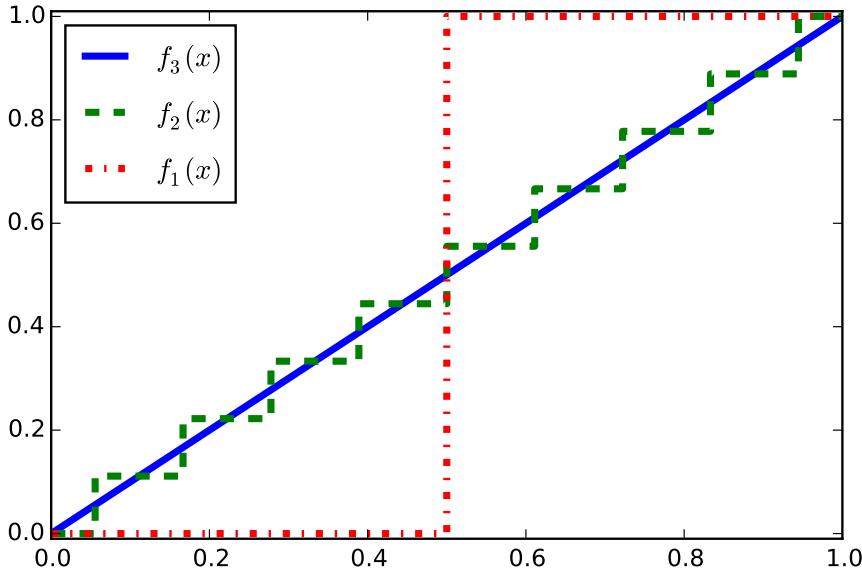


Figure 1.1: Plots of three functions with ($N = 1, 10, 100$) steps and a total variation equal to 1.0

Function	$\int_{[0,1]} \nabla f $	$\int_{[0,1]} \nabla f ^2$
f_1	1.0	1.0
f_2	1.0	0.1
f_3	1.0	0.01

One can see that the L^2 variation term favours continuous transitions such as f_3 rather than the sudden jump in f_1 whereas the total variation is the same for all cases.

1.1.2 Discretization

For a grayscale image $u : \Omega \subset \mathbb{R}^2 \rightarrow \mathbb{R}$ we choose Ω as a discrete grid of pixels $\{1, 2, \dots, m\} \times \{1, 2, \dots, n\} \subset \mathbb{R}^2$. The picture u then takes the values $u_{i,j} := u((i,j)) \in [0, 1]$ at the points $(i, j) \in \Omega$ and we can use a forward finite difference discretization of the gradient $\nabla u := (u_x, u_y)^T$ where

$$(u_x)_{i,j} = \begin{cases} u_{i,j+1} - u_{i,j}, & 1 \leq j < n \\ 0, & j = n \end{cases} \quad (1.4)$$

$$(u_y)_{i,j} = \begin{cases} u_{i+1,j} - u_{i,j}, & 1 \leq i < m \\ 0, & i = m \end{cases}.$$

We then arrive at the *isotropic* TV functional

$$TV_{iso}(u) = \sum_{i,j} \sqrt{(u_{i+1,j} - u_{i,j})^2 + (u_{i,j+1} - u_{i,j})^2}. \quad (1.5)$$

This TV term corresponds to the formulation originally proposed by Rudin et al [18], where $|\nabla u| = \sqrt{u_x^2 + u_y^2}$. Another possibility, which allows for more flexibility in the reweighting process described in 2.2.2 and is even necessary for the proximal point algorithm described in 2.2.1, is the *anisotropic* version of the functional

$$TV_{aniso}(u) = \sum_{i,j} |u_{i+1,j} - u_{i,j}| + |u_{i,j_1} - u_{i,j}|. \quad (1.6)$$

For the total functional we only have to add the pixel-wise differences to the original picture u_0

$$J(u) = \sum_{i,j} (u_{i,j} - (u_0)_{i,j})^2 + \lambda TV(u). \quad (1.7)$$

1.2 Color Images

The next step in the development of image denoising algorithms was their generalization to color images. From a mathematical perspective this just means considering pictures from $\Omega \rightarrow C \simeq \mathbb{R}^3$ where the form and additional properties of C depend on the chosen color model.

In the most simple case of linear models, like RGB for instance, one could choose C as $[0, 1]^3$ and consider denoising each component individually (channel-by-channel model) or consider \mathbb{R}^3 as a normed vector space of tuples (x_R, x_G, x_B) (linear-vectorial model).

For the nonlinear models, especially the so-called chromaticity-brightness model, Chang and Kang [16] showed the closest resemblance to human perception. In this case we can take $C = S^2 \times [0, 1]$ such that the chromaticity takes values on the sphere S^2 considered as a submanifold of the Euclidian space \mathbb{R}^3 , while the brightness is real-valued, as in the case of grayscale images.

1.3 Manifold-valued Images

In the last section we have already seen that, depending on the chosen color model, pixels can take their values on a manifold and are usually represented by their matrices. This data arises in a variety of application such as Diffusion Tensor Magnetic Resonance Imaging (DTI-MRI), computer vision and robotics to name just a few.

1.4 Objective and Outline of this work

In this work we will introduce an extendable multi-threaded C++ template library for the purpose of TV Minimization of manifold-valued images. So far the implemented minimization algorithms are based on the iteratively reweighted least squares (IRLS) adaption suggested by Sprecher and Grohs [14] as well as a proximal point algorithm by Weinmann et al [26]. We extend the implementation to 3D images cubes, the Grassmann manifold and also provides some quasi-analytic expressions for derivatives of the Riemannian distance function.

In the following chapter 2 a short summary of the necessary theory, a description of the algorithms and relevant properties for each of the implemented manifolds. After that the chapter 3 introduces the library itself in particular its capabilities, design concepts, structure, installation and usage in the form of some typical use cases. In Chapter 4 numerical experiments are conducted, showing various application of the library as well as convergence behavior and comparisons between IRLS and proximal point based minimizers.

Finally, chapter 5 concludes with possible extensions and adaptions of the library, in particular possibility of recursive splitting of the image domain into smaller subproblems and the transition to distributed architectures.

Chapter 2

Theory

2.1 Generalization of the functional

The functional defined in (1.7) needs a vector space structure in order for differences to make sense. Hence, the functional needs to be transformed to be still valid in the more general setting of manifold-valued pixels. Since on \mathbb{R} we have $|x - y| = d(x, y)$ for the Euclidian distance, the generalization to metric spaces (X, d) is the appropriate way.

To also include the case of 3D pictures and shorten the notation we will use a graph G to specify over which pairs of pixels the the sums in (1.7) are to be taken. let V be an index set of all pixels in our picture. Denote by $E \subset V \times V$ the set of directed edges in G and by $n(i) := \{j \in V : (i, j) \in E\}$ the set of i 's neighbors. Then

$$TV_{iso}(u) = \sum_{i \in V} \sqrt{\sum_{j \in n(i)} d^2(u_i, u_j)} \quad (2.1)$$

$$TV_{aniso}(u) = \sum_{(i, j) \in E} d^2(u_i, u_j). \quad (2.2)$$

Since we used forward discretization, $n(i)$ just contains the next grid neighbors in each dimension, i.e. for $u_i = u((j, k, l))$ the neighbors are $n(i) = \{u((j+1, k, l)), u((j, k+1, l)), u((i, k, l+1))\}$. To also cover inpainting problems, let $V_k \subset V$ be the index set of pixels where the pixels of the (noisy) original image u_0 are known. Our generalized functional is given by

$$J(u) = \frac{1}{2} \sum_{i \in V_k} d^2(u_i, (u_0)_i) + \lambda TV(u). \quad (2.3)$$

2.1.1 The Riemannian distance

The metric d that is going to be used will of course depend on the manifold. Distances on Riemannian manifolds can be measured using smooth curves $\gamma[a, b] \rightarrow M$ on the manifold. The length of this curve is then

$$L(\gamma) = \int_a^b \sqrt{\langle \dot{\gamma}(t), \dot{\gamma}(t) \rangle_{\gamma(t)}} dt \quad (2.4)$$

and the Riemanian distance can then consequently be defined as

$$d : M \times M \rightarrow \mathbb{R}, \quad dist(x, y) = \inf_{\Gamma} L(\gamma) \quad (2.5)$$

where Γ denotes the set of smooth curves connecting $x, y \in M$ and $\langle \cdot, \cdot \rangle_x$ denotes the inner product on $T_x M$.

This rather general definition is used by [2] whereas for our case Pennec et al's [7] approach using Riemannian exponential and logarithm map is more convenient. Their definitions are based on *geodesics* because they are precisely the curves that realize the above minimum.

Let M be a connected, geodesically complete Riemannian manifold. The exponential mapping $\exp_x : T_x M \rightarrow M$ is defined by $\exp_x(\nu) := \gamma(1)$ where $\gamma : \mathbb{R} \rightarrow M$ is the unit geodesic with $\gamma(0) = x$ and $\dot{\gamma}(0) = \nu$. Thus, the function maps the tangent vector ν to the point y on the manifold reached after moving along γ for $t = 1$. The logarithm map is its inverse and hence provides us with the tangent vector ν that belongs to the geodesic connecting x and y .

A nice overview of Pennec's reinterpretation of vector space operations was given in [21].

Operation	Vector space	Riemannian manifold
Subtraction	$\nu = y - x$	$\nu = \log_x(y)$
Addition	$y = x + v$	$y = \exp_x(\nu)$
Distance	$\text{dist}(x, y) = \ y - x\ $	$\text{dist}(x, y) = \ \log_x(y)\ _x$

2.2 Algorithms

The next topic that must be addressed is the minimization of the above defined functional which brings about another challenge in the form of its non-differentiability. In the implementation, this problem is solved using two different methods which are based on either working with a regularized version of the functional or by approximations of the gradient using proximal mappings. The latter is used by the so-called proximal point algorithm which will be shortly summarized in the next section before proceeding to the iteratively reweighted least squares (IRLS) algorithm.

2.2.1 Proximal Point

The proximal point algorithm for manifold valued data which is implemented in the TVMT library is based on the work of Weinmann et al[26] and belongs to the class of proximal splitting methods. A survey on these methods for Euclidian space data is provided in [11]. The general scope in the real case are convex optimization problems of the form

$$\text{minimize}_{x \in \mathbb{R}^n} f_1(x) + \dots + f_m(x) \quad (2.6)$$

where the $f_i : \mathbb{R}^n \rightarrow]-\infty, \infty]$ are convex function but not necessarily differentiable. As we can see this is also true for the functional (2.3) even for the simple Euclidian case were the summands are given by $d(x, y) = |x - y|$.

The splitting means considering every summand f_i individually and minimize it using its proximal mapping

$$\text{prox}_{f_i} x = \operatorname{argmin}_{y \in \mathbb{R}^n} \left(f_i(y) + \frac{1}{2} \|x - y\|_2^2 \right). \quad (2.7)$$

For the case of a differentiable function f the minimization problem (2.7) is trivial and we see that $\text{prox}_f x = x - \nabla f$, which can be interpreted as a gradient descent step.

In addition to that, one can show that the minimizers of f are exactly the fixed points of the proximal mapping (See [22], §2.3).

Application to manifolds

Let M be a Riemannian manifold and Ω as defined in ???. Due to the square root involved in the isotropic case, the method can only be applied to the anisotropic version of the functional (2.3) and

leads in the 2D case to the following splitting

$$J(u) = \sum_{i,j} F_{ij}(u) + \lambda \sum_{i,j} G_{ij}(u) + \lambda \sum_{i,j} H_{ij}(u) \quad (2.8)$$

$$F_{ij}(u) = d^2(u_{i,j}, (u_0)_{i,j}) \quad (2.9)$$

$$G_{ij}(u) = d(u_{i,j}, u_{i,j+1}) \quad (2.10)$$

$$H_{ij}(u) = d(u_{i,j}, u_{i+1,j}) \quad (2.11)$$

and proximal mappings of the form

$$\text{prox}_{\lambda G_{ij}} x = \underset{y \in M^{m \times n}}{\text{argmin}} \left(\lambda G_{ij}(y) + \frac{1}{2} d^2(x, y) \right). \quad (2.12)$$

We now only present the formulae relevant for the implementation and finally the algorithm itself. For details on derivations and convergence and existence proofs consider [26].

The proximal mappings itself can be computed using unit speed geodesics. Here $[x, y]_t$ denotes the point reached by following the unit speed geodesic starting and x in direction y for a time t .

$$(\text{prox}_{\lambda G_{ij}} u)_{ij} = [u_{i,j}, u_{i,j+1}]_{t_{TV}} \quad (2.13)$$

$$(\text{prox}_{\lambda H_{ij}} u)_{ij} = [u_{i,j}, u_{i+1,j}]_{t_{TV}} \quad (2.14)$$

$$(\text{prox}_{\lambda F_{ij}} u)_{ij} = [u_{i,j}, (u_0)_{i,j}]_{t_{l^2}} \quad (2.15)$$

Lastly, the times in the case of G_{ij} and F_{ij} are computed by

$$t_{TV} = \begin{cases} \lambda, & \text{if } \lambda < d(u_{i,j}, u_{i,j+1}) \\ \lambda / d(u_{i,j}, u_{i,j+1}), & \text{else} \end{cases} \quad (2.16)$$

$$t_{l^2} = \frac{\lambda}{1 + \lambda} d(u_{i,j}, (u_0)_{i,j}) \quad (2.17)$$

and for H_{ij} analogously.

In summary, the parallel version of the proximal algorithm now works by computing, using the rules above, for every pixel $u_{i,j}$ a proximal mapping to its next neighbors on the grid and one to the original picture u_0 , i.e. $[u_{i,j}, v]_t$ with $v \in \{u_{i-1,j}, u_{i+1,j}, u_{i,j-1}, u_{i,j+1}, (u_0)_{i,j}\}$. Next, the intrinsic (Karcher) mean [17] of these five mappings is computed and the pixel updated to a new value $u'_{i,j}$.

Finally, we can state the algorithm

Algorithm 2.1 Parallel proximal point algorithm

$u \leftarrow u_0$

2.2.2 IRLS

The IRLS approach to dealing with non-differentiable terms in the functional is by adding additional terms for regularization which means in the continuous (and isotropic) case

$$TV_\epsilon = \int_{\Omega} \omega_\epsilon |\nabla u|^2 = \int_{\Omega} \frac{|\nabla u|^2}{\sqrt{|\nabla u|^2 + \epsilon^2}} \quad (2.18)$$

for a small $\epsilon > 0$. In this form, the functional becomes differentiable and (??) can be interpreted as weighted L^2 norm.

The IRLS algorithm, described in more detail in [23], works by alternating between reweighting and minimization step. First, the weights are computed, then the the minimization is performed

using the regularized. The steps for the isotropic functional are as follows

$$w_i^{new} = W_{iso}^\epsilon(u)_i := \left(\sum_{j \in n(i)} d(u_i, u_j) + \epsilon^2 \right)^{-\frac{1}{2}} \quad \forall i \in V \quad (2.19)$$

$$u^{new} = U(w) := \operatorname{argmin}_{u \in \Omega} \sum_{i \in V_k} d^2(u_i, (u_0)_i) + \lambda \sum_{i \in V} w_i \sum_{j \in n(i)} d^2(u_i, u_j) \quad (2.20)$$

while the anisotropic steps

$$w_{i,j}^{new} = W_{aniso}^\epsilon(u)_{i,j} := (d(u_i, u_j) + \epsilon^2)^{-\frac{1}{2}}, \quad \forall (i, j) \in E \quad (2.21)$$

$$u^{new} = U(w) := \operatorname{argmin}_{u \in \Omega} \sum_{i \in V_k} d^2(u_i, (u_0)_i) + \lambda \sum_{(i,j) \in E} w_{i,j} d^2(u_i, u_j). \quad (2.22)$$

Further details on the derivation and proofs on convergence, existance and uniqueness of solution for different manifold classes, such as Hadamard spaces or the sphere, can be found in [?]. The minimization can in principle be performed using various methods from smooth optimization theory. Due to its quadratic convergence rate, here the Newton method was chosen. Finally, the algorithm is given as

Algorithm 2.2 IRLS algorithm

$u \leftarrow u_0$

2.3 Riemannian Newton method

2.3.1 Gradient

Let M be a Riemannian manifold and $T_x M$ the tangent space at $x \in M$. Furthermore, let $f : M \rightarrow \mathbb{R}$ a smooth function defined on M . We define the *Riemannian gradient* $\operatorname{grad} f$ to be the unique tangent vector $\xi \in T_x M$ satisfying

$$\langle \operatorname{grad} f(x), \xi \rangle_x = Df(x)[\xi] \quad (2.23)$$

where $Df(x)[\xi] = \xi[f]$, when $\xi : C^\infty(M) \rightarrow \mathbb{R}$ is interpreted as derivation acting on f . The Riemannian gradient shares many properties of its Euclidian counterpart, in particular defining the direction of steepest ascent, such that first order methods like gradient descent can already be implemented for manifolds.

2.3.2 Hessian

For second order methods, such as the Newton algorithm, we also need a corresponding *Riemannian Hessian*. Following again the work of Absil et al [2], the Hessian is realized as linear endomorphism of the tangent space $T_x M$

$$\operatorname{Hess} f(x)[\xi] = \nabla_\xi \operatorname{grad} f(x) \quad (2.24)$$

where ∇ denotes the Riemannian connection on M .

2.3.3 Newton equation

Now let $x \in M$ and $\gamma : [a, b] \rightarrow M$ be a unit geodesic with $\gamma(0) = x$ and $\dot{\gamma}(0) = \nu \in T_x M$ passing through $y \in M$. Thus, we can set $y = \exp_x(\nu)$ and $\nu = \log_x(y)$ as suggested in 2.1.1. Using

the definitions of gradient and Hessian above one can expand f to second order around x in the following way

$$\begin{aligned} f(\exp_x(\nu)) &= f_x(\nu) = f(x) + \langle \text{grad } f(x), \nu \rangle_x + \frac{1}{2} \langle \text{Hess } f(x)[\nu], \nu \rangle_x + \mathcal{O}(\|\nu\|_x^3) \quad \Leftrightarrow \quad (2.25) \\ f(y) &= f(x) + \langle \text{grad } f(x), \log_x(y) \rangle_x + \frac{1}{2} \langle \text{Hess } f(x)[\log_x(y)], \log_x(y) \rangle_x + \mathcal{O}(\|\log_x(y)\|_x^3). \end{aligned}$$

The first line of (2.25) suggest that, for given fixed x , we can also interpret f as a function of $\nu \in T_x M$ and perform the optimization f on the tangent space, on which $\text{grad } f$ and $\text{Hess } f$ have been defined. This finally leads to the following generalization of the *Newton equation* for the correction term $\eta_x \in T_x M$:

$$\text{Hess } f(x)[\eta_x] = -\text{grad } f(x) \quad (2.26)$$

The solution of the Newton equation η_x is a tangent vector of $T_x M$ and, as (2.25) already implies, can be mapped back to the manifold M using the exponential map such that $y' = \exp_x(\eta_x)$. The whole iterative procedure, based on [2], Algorithm 5 is summarized in the following listing.

Algorithm 2.3 Riemannian Newton method for real-valued functions

$u \leftarrow u_0$

2.3.4 Newton equation for the TV functional

We now show the general form of the linear system defined by (2.26) in the case of manifold-valued 3D images $u, u_0 : \Omega \rightarrow M^{Z \times Y \times X}$ where $\Omega = [1, Z] \times [1, Y] \times [1, X] \cap \mathbb{N}^3$ for $X, Y, Z \in \mathbb{N}$ and a corresponding functional $J : M^{Z \times Y \times X} \rightarrow \mathbb{R}$. Furthermore, the forward finite difference scheme described in ?? is used. To simplify notation, we also assume the regularization weights obtained from the last IRLS reweighting step to be all equal to 1 and thus consider only the bare $d^2(\cdot, \cdot)$ terms:

$$\begin{aligned} J(u_{111}, u_{112}, \dots, u_{ijk}, \dots, u_{ZYX}) &= \sum_{ijk} d^2(u_{ijk}, (u_0)_{ijk}) \quad (2.27) \\ &+ \lambda \sum_{ijk}^{Z-1, Y, X} d^2(u_{ijk}, u_{i+1, j, k}) + \lambda \sum_{ijk}^{Z, Y-1, X} d^2(u_{ijk}, u_{i, j+1, k}) + \lambda \sum_{ijk}^{Z, Y, X-1} d^2(u_{ijk}, u_{i, j, k+1}) \end{aligned}$$

We obtain the first derivatives

$$\begin{aligned} \frac{\partial J}{\partial u_{mnp}} &= d_x^2(u_{mnp}, (u_0)_{mnp}) + \lambda(d_x^2(u_{mnp}, u_{m+1, n, p}) + d_y^2(u_{m-1, n, p}, u_{mnp}) \\ &\quad + d_x^2(u_{mnp}, u_{m, n+1, p}) + d_y^2(u_{m, n-1, p}, u_{mnp}) \\ &\quad + d_x^2(u_{mnp}, u_{m, n, p+1}) + d_y^2(u_{m, n, p-1}, u_{mnp})), \end{aligned} \quad (2.28)$$

where $d_x^2((u, v)) := \frac{\partial}{\partial x} d^2(x, y) \Big|_{(x, y)=(u, v)}$.

The second derivatives, considering only the terms in the first line of (2.28) containing the fidelity term and discrete gradient components in z direction, are then

$$\frac{\partial^2 J}{\partial u_{ijk} \partial u_{mnp}} = \delta_{nj} \delta_{kp} \delta_{mi} d_{xx}^2(u_{ijk}, (u_0)_{ijk}) \quad (2.29)$$

$$+ \lambda \delta_{nj} \delta_{kp} \left[\delta_{mi} \left(d_{xx}^2(u_{ijk}, u_{i+1, j, k}) + d_{yy}^2(u_{ijk}, u_{i+1, j, k}) \right) \right] \quad (2.30)$$

$$+ \delta_{m-1, i} d_{xy}^2(u_{ijk}, u_{i+1, j, k}) + \delta_{m+1, i} d_{yy}^2(u_{i-1, j, k}, u_{ijk}) \Big]. \quad (2.31)$$

Considering the pattern of the Kronecker deltas and reshaping the tensor of second partial derivatives into a single matrix, we can see that we obtain a block band matrix, where sums of non-mixed

second derivatives of the squared distance function are located on the main diagonal, while all mixed derivatives belonging to same gradient component discretization directions (x , y or z) populate subdiagonals of equal distance to the main diagonal. In particular, discretization in z -direction are on the first, y -direction on Z th and x -direction on $Z \times Y$ th subdiagonals.

To illustrate this, consider the following illustration of the band structure

$$HJ = \begin{pmatrix} D & Z & Y & X & \\ Z & D & Z & Y & X \\ & Z & D & Z & Y \\ Y & & Z & D & Z & Y \\ & Y & & Z & D & Z \\ X & Y & & Z & D & Z \\ & X & Y & Z & D & Z \end{pmatrix}, \quad (2.32)$$

where D denotes a block, consisting of the sum of non-mixed second derivatives, according to the rules defined by the generalization of (2.29) to all discretization directions.

HJ is exactly the matrix representation of the Hessian operator $\text{Hess } J(u)$, while the representation GJ of $\text{grad } J(u)$ is just the vectorized version of the tensor of first derivatives (2.28). If the embedding space of the manifold is given by $\mathbb{R}^{n \times p}$ then the matrix HJ will be an element of $\mathbb{R}^{\tilde{D} \times \tilde{D}}$ with $\tilde{D} = np(XYZ)$. HJ is however, also sparse, with approximately $7\tilde{D}$ non-zero entries. Finally, solving the Newton equation (2.26) corresponds to solving the sparse linear system defined by HJ and GJ , which is also the most computationally demanding part of the algorithm.

2.3.5 Tangent space restriction

To decrease the dimension of the sparse linear system one can choose a basis of the tangent space $T_x M$ for each $x \in M$ and restrict the computed gradient and hessian to the tangent space by performing a basis transformation. Since $\dim T_x M = \dim M$ this means that we can represent tangent vectors, such as the gradients of the squared distance functions, or tangent space mappings like the Hessian using only $\dim M$ coefficients. We now have $D = \dim M(XYZ)$ such that the prefactor is now the intrinsic manifold dimension and not the dimension of the embedding space any more.

2.4 Manifolds

In this section we will present the relevant quantities needed to implement the IRLS and proximal point algorithm. These are the distance function and its derivatives, exponential and logarithm mapping and the tangent space basis transformation mapping. For the Grassmann manifold, due to its quotient manifold nature and because it was not covered in the original implementation based on [?] a more general introduction will be given.

2.4.1 Euclidian

The space in question is just \mathbb{R}^n , hence trivially a manifold and naturally a vector space such that all expression can be calculated using just basic multivariable calculus. The exponential and logarithm mappings are not be understand in the sense of e^x but according to 2.1.1 such that there are just addition and subtraction in the vector space.

Exponential map

$$\exp_x(r) = x + r \quad (2.33)$$

Logarithm map

$$\log_x(y) = y - x \quad (2.34)$$

Squared distance function

$$d(x, y) = \|x - y\|^2 = \sum_{i=1}^n (x_i - y_i)^2 \quad (2.35)$$

First derivative of the squared distance function

$$\frac{\partial d^2(x, y)}{\partial y} = 2x \quad (2.36)$$

Second derivatives of the squared distance function

$$\frac{\partial^2 d^2(x, y)}{\partial x \partial y} = \frac{\partial^2 d^2(x, y)}{\partial x^2} = 2\mathbb{1}_n; \quad (2.37)$$

Tangent space restriction map

$$T_x = \mathbb{1}_{n^2} \quad (2.38)$$

2.4.2 Sphere S^n

We consider the manifold $S^n := \{x \in \mathbb{R}^{n+1} : \|x\| = 1\} \subset \mathbb{R}^{n+1}$ and express the maps in terms of vectors $x, y \in \mathbb{R}^{n+1}$ of the Euclidian embedding space.

The tangent space $T_x S^n$ at $x \in \mathbb{R}$ is, as basic intuition suggests, given by the tangent hyperplane to sphere at x such that

$$T_x S^n := \{y \in \mathbb{R}^{n+1} : x^T y = 0\} \quad (2.39)$$

The standard euclidian inner product $\langle r, s \rangle = r^T s$, restricted to the tangent space, turns S^n into a Riemannian manifold.

Exponential map

$$\exp_x(r) = \cos(\|r\|_2)x + \frac{\sin(\|r\|_2)}{\|r\|_2}r \quad (2.40)$$

Logarithm map

$$\log_x(y) = \arccos(x^T y) \frac{y - x^T y x}{\|y - x^T y x\|_2} \quad (2.41)$$

Note that this is only well-defined for non-antipodal points $x, y \in S^n$.

Squared distance function

$$d(x, y) = \arccos(x^T y) \quad (2.42)$$

First derivative of the squared distance function

$$\frac{\partial d^2(x, y)}{\partial x} = \begin{cases} \frac{-2 \arccos(x^T y)}{\sqrt{1-(x^T y)^2}} y, & x^T y \in (-1, 1) \\ -2y, & x^T y = 1 \end{cases} \quad (2.43)$$

Second derivatives of the squared distance function

$$\frac{\partial^2 d^2(x, y)}{\partial x^2} = \begin{cases} \left[\frac{2}{1-(x^T y)^2} - \frac{2 \arccos(x^T y)}{(1-(x^T y)^2)^{\frac{3}{2}}} x^T y \right] yy^T + 2x^T y \mathbb{1}_{n+1}, & x^T y \in (-1, 1) \\ \frac{2}{3}yy^T + 2x^T y \mathbb{1}_{n+1}, & x^T y = 1 \end{cases} \quad (2.44)$$

And the mixed derivative:

$$\frac{\partial^2 d^2(x, y)}{\partial x \partial y} = \begin{cases} \left[\frac{2}{1-(x^T y)^2} - \frac{2 \arccos(x^T y)}{(1-(x^T y)^2)^{\frac{3}{2}}} x^T y \right] yx^T - \frac{-2 \arccos(x^T y)}{\sqrt{1-(x^T y)^2}} y \mathbb{1}_{n+1}, & x^T y \in (-1, 1) \\ \frac{2}{3}yx^T - 2 \mathbb{1}_{n+1}, & x^T y = 1 \end{cases} \quad (2.45)$$

Tangent space restriction map

As we can see from (2.39), the tangent space $T_x S^n$ at $x \in S^n$ is just the orthogonal complement of x in \mathbb{R}^{n+1} . Thus, constructing a basis of the tangent space just amounts to constructing an orthonormal basis $\{b_0, b_1, \dots, b_n\}$ of \mathbb{R}^{n+1} with $b_0 = x$, then $\mathcal{B} = \{b_i\}_{i=1}^n$ is the basis of the tangent space.

This can be done using the QR algorithm or, in the case of $S^2 \subset \mathbb{R}^3$ using the cross product. For our basis transformation mapping this means that

$$T_x : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^n \quad (2.46)$$

$$T_x = (b_1 | b_2 | \dots | b_n). \quad (2.47)$$

2.4.3 Special Orthogonal Group SO(n)

The special orthogonal group, considered as matrix group is defined as follows

$$SO(n) := \{X \mathbb{R}^{n \times n} : X^T X = \mathbb{1}_n, \det(X) = 1\}, \quad (2.48)$$

while its tangent space at $X \in SO(n)$ is

$$T_X SO(n) := \{XS : S^T = -S\} = X[\text{Skew}(n)]. \quad (2.49)$$

We equip $T_X SO(n)$ with the standard inner product $\langle R, S \rangle_X = \text{tr } R^T S$ of its embedding space such that $(SO(n), \langle \cdot, \cdot \rangle)$ becomes a Riemannian manifold.

Exponential map

$$\exp_X R = X \exp(X^T R) \quad (2.50)$$

Note that the exponential map on the right hand side denotes the matrix exponential. The same holds for the logarithm on the right hand side of the following expression.

Logarithm map

$$\log_X(Y) = X \log(X^T Y) \quad (2.51)$$

First derivatives of the distance function

For the computation of the derivatives of the squared distance function, there exists a general analytical result by Karcher[17] that simplifies further computations considerably:

Theorem 2.1 (Karcher). *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*

$$\frac{\partial d^2(x, y)}{\partial x}[\cdot] = -2 \langle \log_x(y), \cdot \rangle_x \quad (2.52)$$

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

Since with (2.51), we have a closed form expression for $\log_x(y)$ the first derivative can be expressed as

$$\frac{\partial d^2(X, Y)}{\partial X} = -2X \log(X^T Y). \quad (2.53)$$

Second derivatives of the distance function

For the computation of the second derivatives we can take the expression obtained using the above theorem as a starting point and follow the approach and notation of Magnus [20]. This allows us to express the derivatives as combinations of simple Kronecker product of the arguments, which also is very straightforward and compact to implement. The detailed derivations can be found in the appendix ?? while here we only represent the final results.

For the second derivative with respect to the first argument one readily arrives at

$$\frac{\partial^2 d^2(X, Y)}{\partial X^2} = -2 \left[\left((\log X^T Y)^T \otimes \mathbb{1}_n \right) + (\mathbb{1}_n \otimes X) D \log(X^T Y) (Y^T \otimes \mathbb{1}_n) K_{nn} \right], \quad (2.54)$$

where K_{nn} denotes the commutator matrix which transforms the column-wise vectorization of a matrix A to the vectorization of its transpose A^T .

The mixed derivative is given by

$$\frac{\partial^2 d^2(X, Y)}{\partial X \partial Y} = -2 (\mathbb{1}_n \otimes X) D \log(X^T Y) (\mathbb{1}_n \otimes X^T), \quad (2.55)$$

These expressions are quasi-analytical: Matrix logarithms, exponentials and the Frechet derivative of the logarithms need to be evaluated numerically. Details concerning the implementation of the latter are postponed to section 2.5.

Tangent space restriction map

The dimension the tangent space $T_X SO(n)$ at $X \in SO(n)$ is $d = \frac{n(n-1)}{2}$. We define a basis for the space of skew-symmetric matrices $\text{Skew}(n)$ in the following way. Let $K = \{(i, j) \in \mathbb{N}^2 : 1 \leq i < d, i < j \leq d\}$ and note that $|K| = d$. For all $k \in K$ define the basis vector $B^{(k)}$ by

$$B_{ij}^{(k)} = \begin{cases} \frac{1}{\sqrt{2}}, & i = k_1 \text{ and } j = k_2 \\ -\frac{1}{\sqrt{2}}, & i = k_2 \text{ and } j = k_1 \\ 0, & \text{else} \end{cases} \quad (2.56)$$

The basis of the tangent space is then $\mathcal{B}_{T_X SO(n)} = \{T^{(k)} = XB^{(k)}\}_{k \in K}$ and to obtain the basis we transformation map we vectorize each basis matrix and define a matrix $T_X \in \mathbb{R}^{n^2 \times d}$ whose columns are given by the vectorized basis matrices.

$$T_X : \mathbb{R}^{n^2} \rightarrow \mathbb{R}^d \quad (2.57)$$

$$T_X = (\text{vec } T^{(1)} | \text{vec } T^{(2)} | \cdots | \text{vec } T^{(d)}) \quad (2.58)$$

2.4.4 Symmetric Positive Definite Matrices SPD(n)

The cone of symmetric, positive definite matrices is defined as

$$SPD(n) := \{X \in \mathbb{R}^{n \times n} : X^T = X, y^T X y > 0 \forall y \in \mathbb{R}^n \setminus \{0\}\}. \quad (2.59)$$

For $X \in SPD(n)$ the tangent space at X is isomorphic to the set of symmetric matrices

$$T_X SPD(n) := X[\text{Sym}(n)], \quad (2.60)$$

while Riemannian metric we define on $T_X SPD(n)$ is given by

$$\langle R, S \rangle_X := \text{tr}(X^{-1} R X^{-1} S). \quad (2.61)$$

Exponential map

$$\exp_X(R) = X^{\frac{1}{2}} \exp \left(X^{-\frac{1}{2}} R X^{-\frac{1}{2}} \right) X^{\frac{1}{2}} \quad (2.62)$$

Logarithm map

$$\log_X(Y) = X^{\frac{1}{2}} \log\left(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}\right) X^{\frac{1}{2}} \quad (2.63)$$

Squared distance function

As in the case of the special orthogonal group the Riemannian metric (2.61) induces the distance function on $SPD(n)$ such that

$$d^2(X, Y) = \|\log\left(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}\right)\|_F^2 \quad (2.64)$$

where $\|\cdot\|_F$ denotes the Frobenius norm on $\mathbb{R}^{n \times n}$.

First derivatives of the squared distance function

For the first derivatives we can apply theorem 2.1 again but some care must be taken this time in the computation since the Riemannian metric on $SPD(n)$ depends also on the basepoint X of a tangent space $T_X SPD(n)$.

$$\frac{\partial d^2(X, Y)}{\partial X} = -2 \langle \log_X(Y), \cdot \rangle_X \quad (2.65)$$

$$= -2 \left\langle X^{\frac{1}{2}} \log\left(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}\right) X^{\frac{1}{2}}, \cdot \right\rangle_X \quad (2.66)$$

$$= -2 \left\langle X^{-1} X^{\frac{1}{2}} \log\left(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}\right) X^{\frac{1}{2}} X^{-1}, \cdot \right\rangle_{\mathbb{1}_n} \quad (2.67)$$

$$= -2 \left\langle X^{-\frac{1}{2}} \log\left(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}\right) X^{-\frac{1}{2}}, \cdot \right\rangle_{\mathbb{1}_n} \quad (2.68)$$

Second derivatives of the distance function

For the SPD matrices we proceed in the same way as for the orthogonal group and obtain

$$\begin{aligned} \frac{\partial^2 d^2(X, Y)}{\partial X^2} &= 2 \left[\left(X^{-\frac{1}{2}} \log\left(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}\right)^T \otimes \mathbb{1}_n \right) + \left(\mathbb{1}_n \otimes X^{-\frac{1}{2}} \log\left(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}\right) \right) \right. \\ &\quad \left. + \left(X^{-\frac{1}{2}} \otimes X^{-\frac{1}{2}} \right) D \log\left(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}\right) \left(\left(X^{-\frac{1}{2}} Y \otimes \mathbb{1}_n \right) + \left(\mathbb{1}_n \otimes X^{-\frac{1}{2}} Y \right) \right) \right] \times \dots \\ &\quad \dots \times \left(X^{-\frac{1}{2}} \otimes X^{-\frac{1}{2}} \right) D\left(X^{\frac{1}{2}}\right) \end{aligned} \quad (2.69)$$

and for the mixed derivatives

$$\frac{\partial^2 d^2(X, Y)}{\partial X \partial Y} = -2 \left(X^{-\frac{1}{2}} \otimes X^{-\frac{1}{2}} \right) D \log\left(X^{-\frac{1}{2}} Y X^{-\frac{1}{2}}\right) \left(X^{-\frac{1}{2}} \otimes X^{-\frac{1}{2}} \right) \quad (2.70)$$

Tangent space restriction map

For $SPD(n)$ we have that $d := \dim = \frac{n(n+1)}{2}$. In analogy to the previous manifold we define a manifold for the space of symmetric matrices $Sym(n)$. In this case we have $K = K_1 \cup K_2 = \{(i, i) \in \mathbb{N}^2 : 1 \leq i \leq \} \cup \{(i, j) \in \mathbb{N}^2 : 1 \leq i < d, i < j \leq d\}$ such that again $|K| = d$. For all $k \in K$ define the basis vector $B^{(k)}$ by

$$B_{ij}^{(k)} = \begin{cases} 1, & i = k_1 \text{ and } j = k_2 \\ 1, & i = k_2 \text{ and } j = k_1 \\ 0, & \text{else} \end{cases} \quad (2.71)$$

For $k \in K_1$, these are single entry diagonal matrices. The basis of the tangent space is then $\mathcal{B}_{T_X SPD(n)} = \{T^{(k)} = X^{\frac{1}{2}} B^{(k)} X^{\frac{1}{2}}\}_{k \in K}$ and we proceed completely analogous to the $SO(n)$ case.

2.4.5 Grassmannian $\text{Gr}(n,p)$

The Grassmann manifold is special among the manifolds so far considered due to the fact that it is a quotient manifold. As such, there are different possibilities for choosing equivalence classes and representatives thereof from some matrix space that need to be addressed before an implementation.

For positive integers n and $p \leq n$ the Grassmann manifold is defined as the set of p -dimensional linear subspaces of \mathbb{R}^n . Since a linear subspace $\mathcal{Y} \in \text{Gr}(n,p)$ can be specified using a basis, we can arrange its basis vectors as columns of a matrix $Y \in \mathbb{R}^{n \times p}$ such that its column space spans \mathcal{Y} . The rank of Y must necessarily be full and equal to p because of the linear independence of its columns. Hence, elements of $\text{Gr}(n,p)$ can be represented using elements of the *non-compact Stiefel manifold*

$$\tilde{St}(n,p) := \{Y \in \mathbb{R}^{n \times p} : \text{rank } Y = p\}. \quad (2.72)$$

Quotient representations

Observing now that post-multiplication by any invertible $G \in Gl(p)$ does not change the span of Y , we can form the equivalence classes

$$Y GL(p) := \{YG : G \in Gl(p)\} \quad (2.73)$$

consisting of all matrices having the same span as Y . These equivalence classes can be thought of as the distinct elements of the Grassmannian which leads to the following quotient manifold representation.

$$Gr(n,p) := \tilde{St}(n,p)/Gl(p) \quad (2.74)$$

This representation used by Absil et al[3] which is very general because only the rank is specified. In the next steps of presenting the relevant quantities for the algorithm we will follow Absil's derivation and notation but choose the quotient representation used by Edelman et al [12] which is based on the orthogonal group. This will simplify the most expressions and is also desirable from an algorithmic point of view as it removes more degrees of freedom in the choice of possibly unique representatives.

For the sake of completeness we also mention a completely different approach by Sato and Iwai [24] who choose $\mathbb{R}^{n \times n}$ as embedding space where elements of $Gr(n,p)$ are given by rank p orthogonal projection matrices. The presented application are, however, mostly eigenvalue problems while in the case of image denoising the increased memory requirements are disadvantageous.

The orthogonal group quotient representation of the Grassmann manifold is given by

$$Gr(n,p) = St(n,p)/O(p) \quad (2.75)$$

where we denote by $St(n,p) = \{Y \in \mathbb{R}^{n \times p} : Y^T Y = \mathbb{1}_p, \}$ the *compact Stiefel manifold* with the additional requirement that the basis spanning the subspace \mathcal{Y} be orthonormal now. The canonical quotient projection map is then given by

$$\pi : St(n,p) \ni Y \mapsto \text{span } Y = \mathcal{Y} \in Gr(n,p) \quad (2.76)$$

Locally unique representatives

Let $U \in St(n,p)$ span $\mathcal{U} \in Gr(n,p)$ and define the local affine cross section through U and orthogonal to the fiber $U[O(p)] = \pi^{-1}(\mathcal{U}) \subset St(n,p)$ by

$$S_U := \{V \in St(n,p) : U^T(V - U) = 0\} \subset St(n,p). \quad (2.77)$$

The equivalence class of $V \in St(n,p)$ is equal to $\pi^{-1}(\pi(V)) = V[O(p)]$ and to calculate its intersection with S_U we choose $R \in O(p)$ such that $VR \in V[O(p)]$ and obtain

$$VR \in S_U \Leftrightarrow U^T(VR - U) = 0 \Leftrightarrow R = (U^T V)^{-1} \quad (2.78)$$

which leads to the intersection

$$S_U \cap V[O(p)] = \{VR = V(U^T V)^{-1}\}. \quad (2.79)$$

which can be also empty if $U^T V$ is not invertible. Finally, we define a *cross-section mapping* σ_U restricted to the set

$$\mathcal{U}_U := \{\mathcal{V} = \text{span } V : U^T V \in GL(p)\} \quad (2.80)$$

by

$$\sigma_U : Gr(n, p) \supset \mathcal{U}_U \ni \mathcal{V} = \text{span } V \mapsto V(U^T V)^{-1} \in S_U \subset St(n, p) \subset \mathbb{R}^{n \times p}. \quad (2.81)$$

which is a diffeomorphism providing the differentiable structure.

From an algorithmic point of view the above considerations are important for two reasons. Firstly, this provides us with the means to give well-defined expressions for various quantities we want to compute using arbitrary representatives.

Example 2.1 (Average). *For the case of an average for, we can take representatives $Y_1, \dots, Y_n \in St(n, p)$ for $\mathcal{Y}_1, \dots, \mathcal{Y}_n \in Gr(n, p)$ and find a $U \in St(n, p)$ such that S_U has non-zero intersection with all the Y_i 's equivalence classes, which is equivalent to $U^T Y_i \in Gl(p)$. The average \mathcal{A} can then be written as*

$$\mathcal{A} := \pi \left(\sum_{i=1}^n \sigma_U(Y_i) \right) = \pi \left(\sum_{i=1}^n Y_i (U^T Y_i)^{-1} \right). \quad (2.82)$$

Secondly, it allows us to find a parametrization of $Gr(n, p)$ in terms of $\mathbb{R}^{n \times p}$ matrices. This is necessary to construct a local basis of the tangent base and make the dimension of the sparse linear system a function of the intrinsic manifold dimension $(n-p)p$ instead of the embedding dimension np .

Tangent space

Due to the quotient structure which forces us to work with representatives we cannot just use the usual method for finding the tangent space by differentiating curves on the manifold but have to start with "numerator" of the quotient $St(n, p)$ instead. For the Grassmann manifolds only tangent vectors of a special subspace of $T_Y St(n, p)$, the horizontal space, can modify the span of subspace and exactly those belong to the tangent space of $Gr(n, p)$.

Let $Y \in St(n, p) \subset \mathbb{R}^{n \times p}$. Then tangent space at Y ([4] for details of the derivation) to the compact Stiefel manifold is given by

$$\begin{aligned} T_Y St(n, p) &= \{Z \in \mathbb{R}^{n \times p} : Y^T Z + Z^T Y = 0\} \\ &= \{Y\Omega + Y_\perp K : \Omega \in \text{Skew}(p), K \in \mathbb{R}^{(n-p) \times p}\} \end{aligned} \quad (2.83)$$

where $Y_\perp \in \mathbb{R}^{n \times (n-p)}$ is defined such that $[Y, Y_\perp] \in O(n)$. The second representation of (2.83) already implies the decomposition into vertical and horizontal spaces we are going to perform next. The vertical space at Y is by definition the tangent space to the fiber $\pi^{-1}(\pi(Y))$

$$V_Y = T_Y \pi^{-1}(\pi(Y)) = T_Y Y[O(p)] = Y[\text{Skew}(p)], \quad (2.84)$$

while the horizontal space is defined as its orthogonal complement with respect to (2.83)

$$H_Y = V_Y^\perp = \{H \in T_Y St(n, p) : Y^T H = 0\} \simeq Y_\perp [\mathbb{R}^{(n-p) \times p}]. \quad (2.85)$$

Using this, the tangent space to $Gr(n, p)$ at $\pi(Y) = \mathcal{Y}$, along with its projector, is given by

$$T_{\mathcal{Y}} Gr(n, p) \simeq H_Y St(n, p) \simeq Y_\perp [\mathbb{R}^{(n-p) \times p}] \quad (2.86)$$

$$\pi_{Y_\perp} := \mathbb{1}_n - YY^T. \quad (2.87)$$

The problem that remains is to pick unique representative for a tangent vector $\xi \in T_Y Gr(n, p)$. This is resolved by demanding that the unique representative $\xi_{\diamond Y}$ should project to ξ via

$$d\pi(Y)\xi_{\diamond Y} = \xi \quad (2.88)$$

where $\pi : St(n, p) \rightarrow Gr(n, p)$ is the canonical quotient projection, such that $d\pi$ is a map between their tangent spaces. Using the cross section mapping (2.81), $\xi_{\diamond Y}$ can be computed by

$$\xi_{\diamond Y} = d\sigma_Y(\mathcal{Y})\xi. \quad (2.89)$$

$\xi_{\diamond Y}$ is called the *horizontal lift* of $\xi \in T_Y Gr(n, p)$ at $Y \in St(n, p)$.

[Hier bin ich mir nicht ganz sicher. Hab leider nichts gefunden zum Thema lokaler Basis Grassmann tangent space, daher einfach die Ueberlegung mit den $\{E_{ij}\}_{i=1, j=1}^{n,p}$. Auch nicht wie die tangent space basis dann mit dem horizontalen Lift zusammenarbeiten sollten. Vielleicht müsste man auch noch $d\sigma_Y(\mathcal{Y})B_{ij}$ berechnen.]

Finally, to obtain a basis for the tangent space we can choose $\{E_{ij}\}_{i=1, j=1}^{n,p}$, with the (i,j)th entry set to one and the rest zero, as a basis of $\mathbb{R}^{(n-p) \times p}$ and compute Y_{\perp} using a QR decomposition of Y . The orthogonal complement Y_{\perp} is then just given by $Q_2 \in \mathbb{R}^{n \times (n-p)}$ which is part of the decomposition of the orthogonal matrix $Q = [Q_1, Q_2] \in \mathbb{R}^{n \times n}$.

For the basis of the tangent space we get

$$\{B_{ij}\}_{i=1, j=1}^{n,p} = \{Y_{\perp} E_{ij}\}_{i=1, j=1}^{n,p}. \quad (2.90)$$

Exponential map

Let X, Y span \mathcal{X}, \mathcal{Y} , respectively and let $U\Sigma V^T$ denote the compact singular value decomposition of Y . Then

$$\text{Exp}_{\mathcal{X}}(\mathcal{Y}) = \text{span}(XV \cos \Sigma + U \sin \Sigma). \quad (2.91)$$

Logarithm map

Distance function

Using the previously defined exponential map, one can easily define a geodesic distance function on the Grassmann manifold which is induced by its Riemannian metric

$$g(X, Y) = \text{Tr} X^T Y \quad (2.92)$$

Then the distance function is given by the principal angles θ_i between the subspaces

$$d_g^2(X, Y) = \|\theta\|_2^2 = \sum_{i=1}^p \theta_i^2, \quad (2.93)$$

where the the principal angles can be obtained by computing the singular value decomposition of $X^T Y$.

$$X^T Y = U\Sigma V^T = U \cos \Theta V^T \quad (2.94)$$

$$\Sigma = \text{diag}(\sigma_1, \dots, \sigma_p) \quad (2.95)$$

$$\Theta = \text{diag}(\theta_1, \dots, \theta_p) = \text{diag}(\arccos \sigma_1, \dots, \arccos \sigma_p) \quad (2.96)$$

The distance function (2.93) has the disadvantage that due to the occurrence arccosine, analytical expression are much harder to obtain.

To avoid this problem, we follow Absil's [3] approach and choose an equivalent norm, the so-called projection Frobenius norm, given by

$$d_P^2(X, Y) = \frac{1}{2} \|XX^T - YY^T\|_F^2 = \sum_{i=1}^p \sin^2 \theta_i \quad (2.97)$$

First derivatives of the distance function

$$\frac{\partial d^2(X, Y)}{\partial X} = 2 (XX^T - YY^T) X \quad (2.98)$$

Second derivatives of the distance function

$$\frac{\partial^2 d^2(X, Y)}{\partial X \partial X} = 2 [(X^T X \otimes \mathbb{1}_n) + (\mathbb{1}_p \otimes (XX^T - YY^T)) + (X^T \otimes X) K_{np}] \quad (2.99)$$

The mixed derivative is given by

$$\frac{\partial^2 d^2(X, Y)}{\partial X \partial Y} = -2 [(X^T Y \otimes \mathbb{1}_n) + (X^T \otimes Y) K_{np}] \quad (2.100)$$

2.5 Fréchet derivatives of matrix logarithm and square root

To use the derivative expression computed above we need the so called Kronecker form of the Fréchet derivative. The Fréchet derivative of a matrix valued function $f : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n \times n}$ at a point $X \in \mathbb{R}^{n \times n}$ is a linear function mapping $E \in \mathbb{R}^{n \times n}$ to $L_f(X, E) \in \mathbb{R}^{n \times n}$ such that

$$f(X + E) - f(X) - L_f(X, E) = \mathcal{O}(\|E\|). \quad (2.101)$$

Chain rule and inverse function theorem also hold for the Fréchet derivative:

$$L_{f \circ g}(X, E) = L_f(g(X)), L_g(X, E)) \quad (2.102)$$

$$L_f(X, L_{f^{-1}}(f(X), E)) = E \quad (2.103)$$

As we did in our formulation of the derivatives of the distance function, it can also be represented in the Kronecker form in which is represented as map $K_f : \mathbb{R}^{n^2} \rightarrow \mathbb{R}^{n^2}$, such that $K_f(X) \in \mathbb{R}^{n^2 \times n^2}$ is defined by

$$\text{vec}(L_f(X, E)) = K_f(X) \text{vec}(E) \quad (2.104)$$

where $\text{vec} : \mathbb{R}^{n \times n} \rightarrow \mathbb{R}^{n^2}$ denotes the columnwise vectorization operator.

2.5.1 Derivative of the matrix square root

We start by considering the Fréchet derivative of $f(X) = X^2$, which is given by

$$L_{X^2}(X, E) = XE + EX. \quad (2.105)$$

Applying the inverse function theorem consequently leads to

$$L_{X^2}(X^{\frac{1}{2}}, L_{X^{\frac{1}{2}}}(X, E)) = X^{\frac{1}{2}} L_{X^{\frac{1}{2}}}(X, E) + L_{X^{\frac{1}{2}}}(X, E) X^{\frac{1}{2}} = E, \quad (2.106)$$

where the last equality shows that the Fréchet derivative of the matrix square root $L_{X^{\frac{1}{2}}}(X, E)$ satisfies the Sylvester equation

$$X^{\frac{1}{2}} L + LX^{\frac{1}{2}} = E, \quad L := L_{X^{\frac{1}{2}}}(X, E). \quad (2.107)$$

The Kronecker representation $K_{X^{\frac{1}{2}}}(X)$ can now be obtained be using the vectorization operator on both sides of the equation and rearrange the term to the form (2.104) which leads to

$$K_{X^{\frac{1}{2}}}(X) = \left[\left(\mathbb{1} \otimes X^{\frac{1}{2}} \right) + \left(X^{\frac{1}{2}T} \otimes \mathbb{1} \right) \right]^{-1}. \quad (2.108)$$

However, this straightforward approach has the disadvantage that the inverse of a $n^2 \times n^2$ matrix need to be computed which has complexity $\mathcal{O}((n^2)^3) = \mathcal{O}(n^6)$. In addition to that, the inverse needs to be found explicitly which is not numerically stable in general.

The Sylvester equation (2.107), on the other hand, can be solved with $\mathcal{O}(n^3)$ operations via Schur transformation. We choose E^{ij} , the single-entry matrices having 1 at (i, j) and zero everywhere else, as a basis for $\mathbb{R}^{n \times n}$ and solve the Sylvester equation for each of the n^2 basis matrix elements. By that, the total complexity can be reduced to $n^2\mathcal{O}(n^3) = \mathcal{O}(n^5)$ and we avoid the potentially problematic explicit computation of inverses altogether.

We then obtain the final Kronecker form of the derivative by constructing its rows from the vectorized, tranposed Fréchet derivatives:

$$\left(K_{X^{\frac{1}{2}}}\right)_{in+j,\cdot} = \text{vec}\left(L_{X^{\frac{1}{2}}}(X, E^{ij})^T\right) \quad (2.109)$$

2.5.2 Derivative of the matrix logarithm

For the logarithm we follow the approach described by Al-Mohy et al [5] which is based on the differentiation of the Padé approximant to $\log(1 + X)$. Since this is only applicable if the norm of X is sufficiently small, the use of an inverse scaling and squaring technique based on the relation

$$\log(X) = 2 \log(X^{\frac{1}{2}}) \quad (2.110)$$

is necessary.

Application of the chain rule leads to

$$L_{\log}(X, E_0) = 2 \log\left(X^{\frac{1}{2}}, L_{X^{\frac{1}{2}}}(X, E_0)\right). \quad (2.111)$$

The second argument on the right hand side can again be written as solution $E_1 := L_{X^{\frac{1}{2}}}(A, E_0)$ of an Sylvester-type equation

$$X^{\frac{1}{2}}E_1 + E_1X^{\frac{1}{2}} = E_0. \quad (2.112)$$

Repeating the procedure s times results in

$$L_{\log}(X, E_0) = 2^s L_{\log}\left(X^{\frac{1}{2^s}}, E_s\right) \quad (2.113)$$

$$X^{\frac{1}{2^s}}E_i + E_iX^{\frac{1}{2^s}} = E_{i-1}, \quad i = 1, \dots, s \quad (2.114)$$

where E_s is obtained by successively solving the set of Sylvester equations defined in the second line.

Finally, the Padé approximant of order m in its partial fraction form [15] is given by

$$r_m(X) = \sum_{j=1}^m \alpha_j^{(m)} (\mathbb{1} + \beta_j^{(m)} X)^{-1} X \quad (2.115)$$

where $\alpha_j^{(m)}, \beta_j^{(m)} \in (0, 1)$ are the m -point Gauss-Legendre quadrature weights and nodes.

The derivative of (2.115) is then easily computed as

$$L_{r_m}(X, E) = \sum_{j=1}^m \alpha_j^{(m)} (\mathbb{1} + \beta_j^{(m)} X)^{-1} E (\mathbb{1} + \beta_j^{(m)} X)^{-1} \quad (2.116)$$

which leads to the final approximation of the matrix logarithm derivative,

$$L_{\log}(X, E) \approx 2^s L_{r_m}\left(X^{\frac{1}{2^s}} - \mathbb{1}, E_s\right). \quad (2.117)$$

For the implementation of (2.117) we use algorithm 5.1 from [5] with fixed $m = 7$. The Kronecker representation is then constructed as in the square root case.

Chapter 3

The TVMT Library

The library which was developed in the course of this work is an easy-to-use, fast C++14 template library for TV minimization of manifold valued two- or three-dimensional images.

3.1 Capabilities

3.1.1 Supported Manifolds

- Real Euclidian space \mathbb{R}^n
- Sphere $S^n = \{x \in \mathbb{R}^{n+1} : \|x\| = 1\}$
- Special orthogonal group $SO(n) = \{R \in \mathbb{R}^{n \times n} : RR^T = \mathbb{1}, \det(R) = 1\}$
- Symmetric positive definite matrices $SPD(n) = \{S \in \mathbb{R}^{n \times n} : S = S^T, x^T S x > 0 \forall x \in \mathbb{R}^n\}$
- Grassmann manifold $Gr(n, p) =$

3.1.2 Data

- 2D and 3D images
- Input/Output via OpenCV integration supporting all common 2D image formats
- CSV input for matrix valued data
- Input methods for raw volume image data as well as the NIfTI [?] format for DT-MRI images
- Various methods to identify damaged areas for inpainting

3.1.3 Functionals

- isotropic (only possible for IRLS) or anisotropic TV functionals
- first order TV term
- weighting and inpainting possible

3.1.4 Minimizer

- IRLS
- PRPT

3.1.5 Visualizations

- OpenGL rotated cubes visualization for $\text{SO}(3)$ images
- OpenGL ellipsoid visualization for $\text{SPD}(3)$ images
- OpenGL volume renderer for 3D volume images

3.2 Design concepts

3.2.1 Goals

Performance

Since the core parts of the implementation are originally based on a matlab prototype by Sprecher [?], [9] and [21], one of the most important goals was of course a faster implementation with a smaller memory footprint. On the test platform with two hyperthreaded 2.8GHz cores Intel i5-2520 with AVX vector extensions the Matlab implementation froze for images larger than one Megapixel(MP). Hence, the C++ implementation should enable the algorithm to be tested in a much broader scope which is also closer to common picture sizes in image processing, especially since even smartphones today easily produce pictures in the Megapixel range. In addition to that, also other factors affecting the performance of the algorithm, such as cache locality and memory speed, can be investigated.

The main performance driver for this library is the multilevel-parallelization. Evidently, this does not include the formulation of the IRLS minimization algorithm itself, due to the fact that it is naturally a iterative method, but rather any possible subtask such as computation of the functional values, gradient and Hessian, for example. On top of that, we tried to maximize cache locality on the loop level and to free memory as soon as possible but keep data that is used very often and requires costly recomputation, like for example the IRLS weights.

In contrast to the original Matlab implementation, the computation of various quantities such as weights, first and second derivatives is not realized with tensor products any more. For Matlab, due the low speed of its internal loop constructs, the approach is justified but in a pure C++ implementation other factors are more important. One reason for the change is improved readability and maintainability of the code since tensor product usually tend to become very convoluted, especially for the manifolds with matrix representations. Also the modularization of the manifold class is not straightforward any more.

From a performance and parallelization perspective, contractions of tensor products are similar to Matrix products and usually require some sort of blocking scheme for parallelization. In addition to that, because certain reshapes of the image container prior to the computation are necessary, the dimensions to be summed over are not necessarily contiguous in memory such that a high a cache utilization is more difficult to achieve.

Finally, in order to formulate certain operations as tensor products, temporary tensors of the correct dimensionality need to be created, which are actually not necessary.

Another measure that significantly reduces the memory footprint for the IRLS minimizer, especially for manifolds with matrix representations, is to only save gradient and hessian in their local tangent space basis representation, such that the degrees of freedom correspond the intrinsic manifold dimension not the dimension of the embedding space. This also reduces the time to solve the sparse linear system.

Modularization and Extendability

In principle the programming paradigm in Matlab is still procedural resulting in a hierarchy of functions for the various tasks. Handling different types of manifolds then usually requires switch expressions in all functions that use manifold-specific functionality. Adding a support for a new

manifold to the algorithm or modifying existing manifold functionality makes a modification of all switch cases necessary. There is no single point of change but many source files need to be edited.

For the TVMT Library an object oriented and generic programming approach was chosen, which tries to model each variable component of the algorithm in a separate class, as independent of the other components as possible. Differences in each class are represented by specializations of their primary class template. The best example for that is the manifold class which has a specialization for every supported manifold type and due to the fact that the functions implemented in those class specializations are generally just functions of one or two elements of the manifold, they could also be used in other projects which require the same functionality.

Interfaces between classes are provided by giving classes higher in the hierarchy template parameters corresponding to lower classes: The class modelling the functional, for instance, has a manifold type template parameter, as described in more detail in the section 3.3. Like all other component classes, also the functional class can be extended by adding further specializations for other types of functionals, that include for example higher order terms or have different fidelity terms [?].

Those specializations also have the advantage that the code is just in one file, a single point of change to increase readability and maintainability.

3.2.2 Levels of parallelization

Parallelization takes place on two levels. The first one is shared memory multithreading implemented with the OpenMP language extensions. In most cases this is realized using the so-called *pixel-wise kernels* of the VPP library, which makes it possible to map an arbitrary function on all pixels of a set of image containers: The function is called for each tuple of pixels having the same coordinates in their respective image. For the parallel execution each processor is assigned a batch of image rows. If the pixel wise kernels are not applicable, for example if the needed subdomain of the image is too complicated, we use manual OpenMP loop parallelization. At the time of writing, this unfortunately also includes 3D images, such that we implemented an own version of 3D pixel-wise kernels to keep the code compact.

The alternative to the above described tensor product implementation is to use pixel-wise kernels to parallelize any operation to requires iteration over an image container. For most computations, in the case of computing derivatives, we only need the pixel and its next neighbor in a given dimension. For calculating the forward derivatives we just have to call the pixel-wise kernel with two subimages of our current working image: One with the last slice (of the given dimension) missing and one with the first slice missing. This is demonstrated by the following short listing 3.1 and further illustrated in 3.1:

Listing 3.1 Pixel-wise forward derivative computation

```

1 auto calc_first_arg_deriv = [&] (value_type& x, const weights_type& w, const <
2   value_type& i, const value_type& n) { MANIFOLD::deriv1x_dist_squared(i, n, x); x *= <
3   w; };
4
5 img_type YD1 = img_type(without_last_row);
6 vpp::pixel_wise(YD1, weightsY_ | without_last_row, data_.img_ | without_last_row, <
7   data_.img_ | without_first_row) | calc_first_arg_deriv;

```

The advantage is that even though we evaluate some function for a pair of neighboring pixels which are not adjacent in memory, the parallel processing is still always row-wise. Since rows in row major languages like C++ are contiguous in memory we can avoid frequent memory access on distant locations and consequently avoid cache thrashing to a certain degree.

The second level of parallelization is instruction level parallelism, also known as *Single Instruction Multiple Data* (SIMD), which uses the processor's vector extensions (e.g. SSE, AVX, NEON). The CPU provides some additional special SIMD registers with increased size of usually 128 bits to 512 bits such that multiple integer or floating point variables fit inside. Then an arithmetic operation

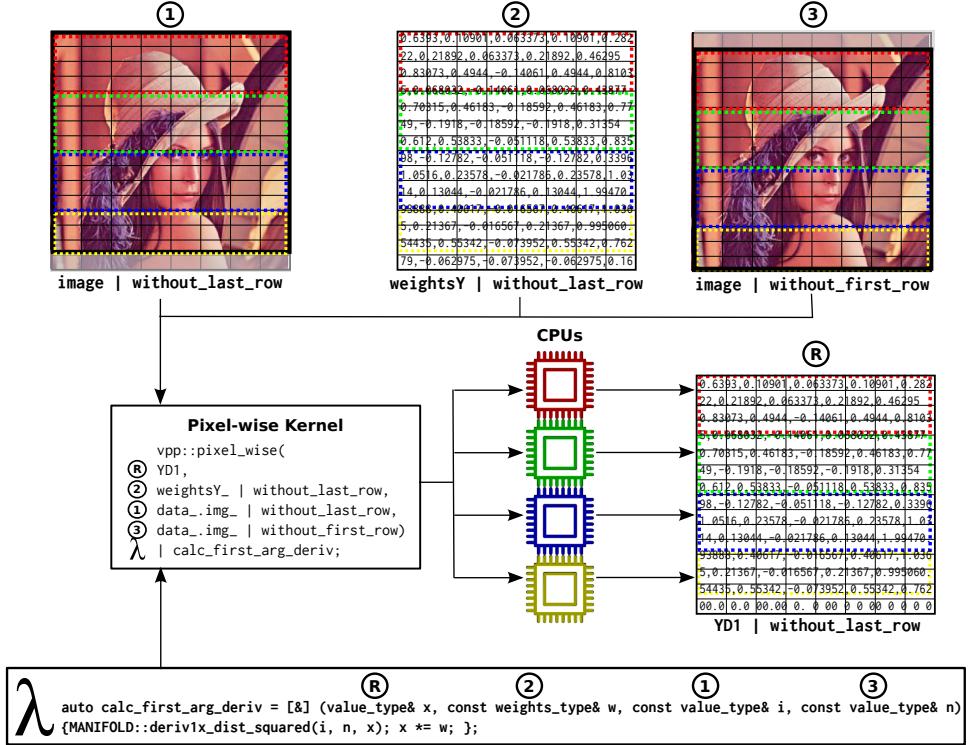


Figure 3.1: Parallel calculation of derivatives in y -direction and weighting using pixel-wise kernels. For each pixel position (i, j) in the three input pictures, ①, ② and ③, as well as the output picture ④ the pixel-wise kernel creates a tuple $(R_{ij}, 2_{ij}, 1_{ij}, 3_{ij}) = (YD1_{i,j}, \text{weightY}_{i,j}, \text{Image}_{i,j}, \text{Image}_{i,j+1})$ which is than used to call the specified lambda function. Depending on the row number of the pixel, the calls are executed by different CPU cores.

is simultaneously applied to all variables in the register (see figure 3.2) such that theoretically the amount of floating point operations is multiplied by the number of variables fitting in the registers.

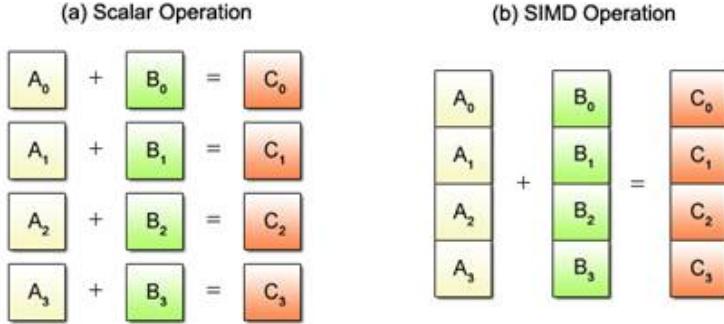


Figure 3.2: Instruction level parallelism using SIMD registers.

In order to really achieve this speedup the data must be aligned in memory which means the address of pixels in memory must always be multiple of the SIMD register size. Fortunately, that issue is handled by the VPP and Eigen libraries enabling the compiler to perform the necessary vectorization optimizations.

3.2.3 C++ techniques

The TVMT Library tries to take advantage of new C++11 and C++14 language features in order to speed up computations via compile-time optimizations and also make the code more compact and readable. The most important tools in that regard are lambda functions and variadic templates which are shortly described in the following section.

Lambda functions

A lambda function is basically a locally defined function object, which is able to capture variable from the surrounding scope that can but need not to be named. The corresponding Matlab language construct is an anonymous function or function handle, usually defined using the @ operator. The following listing shows the basic definitions and uses cases of lambda functions:

Listing 3.2 Lambda functions

```
1 int init = 5;
2 std::vector<int> v {1, 2, 3, 4};
3
4 // C++11 lambda function for adding integers
5 // init is captured by reference
6 auto f = [&] (int a, int b) {return a + b + init;};
7
8 // C++14 generic argument lambda function
9 // init is captured by value
10 auto g = [=] (auto a, auto b) {return a + b + init;};
11
12 // Call named lambda functions
13 int d = f(8, 3);
14 double e = g(1.0f, 5);
15
16 // or directly pass anonymous lambda function as argument
17 std::transform(v.begin(), v.end(), v.begin(), [] (auto x) { ++x; });
```

In the TVMT library lambda functions provide the connection between the static manifold methods and the pixel-wise kernels which apply them to the image containers. A typical case can be seen in the already introduce listing 3.1. Since lambda functions are only locally defined, in the scope where they are actually needed, one can avoid making the method list of the classes unnecessary long.

Variadic templates

With variadic templates it is possible to define functions which take a variable number of arguments. Obviously, this is also possible in other languages like Matlab or C with the most prominent example being the function printf. However this usually implemented using some list type (in C va.list), which adds additional overhead, whereas in C++ it is realized via some special kind of template metaprogramming which is recursive in nature. The recursion, in turn, is resolved at compile-time and leads to code that is actually equivalent to manually defining a function with the desired number of arguments and consequently there is no additional runtime effort.

Listing 3.3 Variadic template example

```
1 // Recursion base case
2 template<typename T>
3 T sum(T v) {
4     return v;
5 }
6
7 // Recursive template
8 template<typename T, typename... Args>
9 T sum(T first, Args... args) {
10     return first + sum(args...);
11 }
```

The main application for this constructs in TVML are the implementation of the Karcher mean, needed for the proximal point implementation, and TVML's own version of the 3D pixel-wise kernels.

3.3 Components

3.3.1 Manifold classes

The manifold template class encapsulates all information and methods related to the differential geometric structure of the data. This enables the generic implementation of the functionality higher in the class hierarchy such as functional evaluations or minimizers. The primary template has the following parameters

```
1 // Primary Template
2 template <enum MANIFOLD_TYPE MF, int N, int P=0>
3 struct Manifold {
4 };
```

where MF is an enumeration constant to specify the type of the manifold, N denotes the dimension of the representation space and P the dimension of subspaces, as in the case of Grassmann manifolds. In order to add a new manifold one just has to implement a specialization of this primary template.

So far, the manifold classes contains functionality necessary for TV minimization using either the IRLS or proximal point algorithm and furthermore some additional operations that are needed for supporting tasks like interpolation and smoothing. The classes are implemented using only static constants and methods: At no time is it necessary or desired to actually instantiate the class. The methods itself are usually unary or binary functions, with parameters and result all passed by reference to avoid copies. Since these methods are called very often, basically for every pair of neighboring pixels, they are all declared inline in order to support the compiler during the code optimization.

It is also possible to use these classes in other projects requiring similiar functionality like for instance when implementing a geodesic finite element solver.

In the following we will look at excerpts of the SPD implementation to illustrate which information and functionality a new manifolds class need to provide and to give an overview of the available functions.

Static constants

```
1 static const MANIFOLD_TYPE MyType;           // SPD
2 static const int manifold_dim ;              // N*(N+1)/2
3 static const int value_dim;                  // N*N
4
5 static const bool non_isometric_embedding;
```

The first constant just stores the manifold template template parameter introduced above, while `manifold_dim` and `value_dim` are the intrinsic dimension of the manifold and of its embedding space, respectively. Finally, the boolean constant is just a flag which tells the algorithm that special pre- and postprocessing for interpolation is necessary.

Type definitions

To enable the generic formulation of the algorithms the manifold classes provide a mapping between the types of their values, derivatives, tangent bases and underlying scalar type and their actual representation as matrix and vector data types of the Eigen linear algebra library. Examples can be seen in the following listing:

```
1 // Scalar and value typedefs
2 typedef double scalar_type;
3 typedef double dist_type;
4 typedef Eigen::Matrix<scalar_type, N, N> value_type;
5 // ...
6
7 // Tangent space typedefs
8 typedef Eigen::Matrix<scalar_type, N*N, N*(N+1)/2> tm_base_type;
9 // ...
```

```

10 // Derivative Typedefs
11 typedef value_type deriv1_type;
12 typedef Eigen::sMatrix<scalar_type, N*N, N*N> deriv2_type;
13 typedef Eigen::Matrix<scalar_type, N*(N+1)/2, N*(N+1)/2> restricted_deriv2_type;
14 // ...

```

Static methods

Finally, the following methods are implemented for the manifold classes

Riemannian distance function and its derivatives

```

1 inline static dist_type dist_squared(cref_type x, cref_type y);
2 // First derivatives
3 inline static void deriv1x_dist_squared(cref_type x, cref_type y, deriv1_ref_type<
... result);
4 inline static void deriv1y_dist_squared(cref_type x, cref_type y, deriv1_ref_type<
... result);
5 // Second derivatives
6 inline static void deriv2xx_dist_squared(cref_type x, cref_type y, <
... deriv2_ref_type result);
7 inline static void deriv2xy_dist_squared(cref_type x, cref_type y, <
... deriv2_ref_type result);
8 inline static void deriv2yy_dist_squared(cref_type x, cref_type y, <
... deriv2_ref_type result);

```

Exponential and Logarithm map

```

1 template <typename DerivedX, typename DerivedY>
2 inline static void exp(const Eigen::MatrixBase<DerivedX>& x, const Eigen::<
... MatrixBase<DerivedY>& y, Eigen::MatrixBase<DerivedX>& result);
3 inline static void log(cref_type x, cref_type y, ref_type result);
4
5 inline static void convex_combination(cref_type x, cref_type y, double t, <
... ref_type result);

```

The parameters of the exponential here are not the manifolds own typedefs but the base class of all Eigen matrix data types. The reason for using this construction is that the function can also be called with composite expressions (e.g. $XY + Z$) without a temporary copy. Most of the other functions are usually called with atomic expression only, hence there is no need to use this more complicated construction on a general basis.

The convex_combinations method denotes a point z on the manifold by following a unit time geodesic connecting the points x and y for a time t .

Karcher means

```

1 inline static void karcher_mean(ref_type x, const value_list& v, double tol=1e<
... -10, int maxit=15);
2 inline static void weighted_karcher_mean(ref_type x, const weight_list& w, const <
... value_list& v, double tol=1e-10, int maxit=15);
3
4 // Variadic templated version
5 template <typename V, class ... Args>
6 inline static void karcher_mean(V& x, const Args&... args);

```

Implementations for finding the Karcher mean of an arbitrary number of points. The first version requires the points to be stored in a std::vector container while the second version is based on variadic templates and expects the arguments just as a comma separated list after the first argument, where the final result will be stored. Creating the list for the first version eventually requires copying and is consequently slower but has an overloaded version which allows to compute a weighted Karcher mean

Tangent plane basis, projector and interpolation

```

1 // Basis transformation for restriction to tangent space
2 inline static void tangent_plane_base( cref-type x,  tm-base-ref-type result);
3 // Projection
4 inline static void projector( ref-type x);
5 // Interpolation pre- and postprocessing
6 inline static void interpolation_preprocessing( ref-type x);
7 inline static void interpolation_postprocessing( ref-type x);

```

The first function computes a basis of the tangent space at the point x and stores it in $result$ as the columns of a matrix.

The projector, if defined for the given manifold, will project a point of the ambient embedding spacing onto the manifold. In the case of Euclidian space, where the embedding space and the manifold are identical, this function does nothing and will be optimized out by the compiler. Nevertheless, it must exist or programs will not compile.

Interpolation pre- and postprocessing is necessary for instance for the SPD manifold. Other manifolds must just provide an empty implementation.

3.3.2 Data class

The data class handles anything related to storage, input and output of two- or threedimensional image data, as well as some support functions for detecting edges and damaged areas in a picture. In contrast to the manifold class, the data class needs to be instantiated such that a reference to the data object can be passed to any class which needs data access. In turn, in addition to the dimension of the picture, the data class takes a fully specialized manifold class type as a template parameter:

```

1 // Primary Template
2 template <typename MANIFOLD,  int DIM >
3 class Data {
4 };

```

There are basically four multi-dimensional arrays stored in the data class: The original noisy image, the current working image and, if applicable, arrays storing the inpainting and edge weight information. For storage, the n-dimensional VPP [13] image container is used.

This image container class works very well together with the Eigen vector and matrix data types, provides a variety of expressive loop- and iterator constructs and also takes care of the alignment of the image data in memory, which is a prerequisite for the *Single Instruction Multiple Data* (SIMD) optimization and vectorization by the compiler. Since the memory management of the container is based on std::shared_pointer it is also very easy to efficiently access subimages or slices of an image without any copies.

The most common input method for 2D and 3D are summarized in the following code snippet:

```

1 // 2D Input functions
2 void rgb_imread( const char* filename); // for R^3
3 void rgb_readBrightness( const char* filename); // for R
4 void rgb_readChromaticity( const char* filename); // for S^2
5 void readMatrixDataFromCSV( const char* filename,  const int nx,  const int ny);
6
7 // Synthetic SO/SPD picture
8 void create_nonsmooth_son( const int ny,  const int nx);
9 void create_nonsmooth_spd( const int ny,  const int nx);
10
11 // 3D Input functions
12 void rgb_slice_reader( const char* filename,  int num_slides);
13 void readMatrixDataFromCSV( const char* filename,  const int nz,  const int ny,  const int nx);
14 void readRawVolumeData( const char* filename,  const int nz,  const int ny,  const int nx);

```

The purpose and usage of most of these methods is self-explanatory. The CSV readers expect the data to be a linear list of pixels, where the components of each pixel are comma-separated and row-wise flattened, such that each line of the input file contains exactly one pixel. The order of the list is also row-wise for 2D or slice- than row-wise for 3D images, respectively.

The slice reader reads a series of images, following the filename scheme filenameX.ext, where X is the number of the slice to be read into an image cube at z-coordinate X.

3.3.3 Functional class

In addition to fully specialized Manifold and Data class types (third and fourth template parameters), there are three further template parameters that must be specified by the library user. The first one is the order of the functional which refers to order of the highest differential operator in the TV term of the functional. So far, only first order functionals are implemented which would correspond to setting `ord=FIRSTORDER` in the primary template shown below

```

1 //Primary Template
2 template <enum> FUNCTIONAL_ORDER ord, enum FUNCTIONAL_DISC disc, class MANIFOLD, class ~
3   DATA, int DIM=2>
4 class Functional{
5 };

```

The second template parameters `disc` determines whether the isotropic or the anisotropic version is to be used. Please note that for the proximal point algorithm only anisotropic is available. Finally, the last parameter specifies the dimensionality of the data.

The main purpose of the functional class is to provide methods for the computation of all functional-related quantities, such as evaluation of the functional, its gradient, hessian and construction of a local basis of the tangent spaces. That also means that in the IRLS case the functional class stores sparse linear system that needs to solved in each Newton step.

For users of the library, the most important methods are those for setting the λ and ϵ^2 parameters.

```

1 inline param_type getlambda() const { return lambda_; }
2 inline void setlambda(param_type lam) { lambda_=lam; }
3 inline param_type geteps2() const { return eps2_; }
4 inline void seteps2(param_type eps) { eps2_=eps; }

```

Should it be necessary, it is also possible to access some of the enumerated quantities directly using

```

1 // Evaluation functions
2 result_type evaluateJ();
3 void evaluateDJ();
4 void evaluateHJ();
5
6 void updateTMBase();
7
8 inline const gradient_type& getDJ() const { return DJ_; }
9 inline const sparse_hessian_type& getHJ() const { return HJ_; }
10 inline const tm_base_mat_type& getT() const { return T_; }

```

The functions in lines 2, 3 and 5 merely trigger a recomputation while the last three functions return references to these quantities.

3.3.4 TV Minimizer class

For the TV Minimizer class it makes sense to consider the IRLS and proximal point implementation separately. The primary template is shown in the following code snippet.

```

1 //Primary Template
2 template <enum> ALGORITHM AL, class FUNCTIONAL, class MANIFOLD, class DATA, enum ~
3   PARALLEL PAR=OMP, int DIM=2>
4 class TV_Minimizer{
5 };

```

As in the previous cases we have to provide fully specialized manifold, data and also functional types. Again, the last parameter specifies the dimension of the data. Of the remaining to `PAR` has the default value `OMP` which specifies the method of parallelization, in this case the OpenMP language extensions. Other methods including just serial execution could be added later. The remaining template parameter, `AL` specifies the minimizer to be used and can take the values `IRLS` or `PRPT` (Proximal point).

For IRLS, we have the following user interface

```

1 void first_guess(); // First guess for inpainting
2 void smoothening(int smooth_steps); // Simple averaging box filter
3 newton_error_type newton_step(); // perform one newton step
4 void minimize(); // full minimization

```

while for proximal point we have

```

1 use_approximate_mean(bool u) { use_approximate_mean_ = u; } // turn mean approximation on/off
2 void first_guess(); // First gues for \inpainting
3
4 void updateFidelity(double muk); // Update Fidelity part
5 void updateTV(double muk, int dim, const weights_mat& W); // Update TV part
6
7 void geod_mean(); // Calculate geodesic mean
8 void approx_mean(); // approximate mean using convex combinations
9
10 void prpt_step(double muk); // perform one proximal point step
11 void minimize(); // full minimization

```

3.3.5 Visualization class

This class provides visualizations of 3D volume data and so far SO(3) and SPD(3) visualizations by cubes and ellipsoids. If these are to be used in user code it is necessary to link against OpenGL, GLUT and GLEW libraries, which is explained in more detail in section 3.4.3. The visualization classes have the following primary template.

```

1 // Primary Template
2 template <enum MANIFOLD_TYPE MF, int N, class DATA, int dim=2>
3 class Visualization{
4 };

```

The class methods that are relevant to users of the libary are summarized here

```

1 void saveImage(std::string filename);
2 void GLInit(const char* windowname);
3
4 void paint_inpainted_pixel(bool setFlag);

```

The important function here is GLInit which initializes the rendering of the data. If one intends to also save the image, on has to specify a filename using saveImage *before* calling GLInit. Finally, paint_inpainted_pixel just sets a flag which decides whether inpainted pixels are not painted at all (setFlag = false, default value) or if they are visualized with the value they have at the time of rendering. Usually one wants to set this to true after the minimization to show the results.

A full example to this is presented in section 3.4.4

SO(3) Visualization

For the visualization of SO(3) data we just consider a unit volume cube centered at the origin of \mathbb{R}^3 with its front face normal vector parallel to the y-axis. Then the rotation matrix representing the SO(3) element is applied to the cube. To break the $O_h \simeq S_4 \times S_2$ symmetry of the cube, we give a different color to every face.



Figure 3.3: SO(3) Visualization as oriented, colored cubes

SPD(3) Visualization

For three-dimensional SPD matrices we have six degrees of freedom, which in the case of DT-MRI pictures correspond to the diffusion coefficients in different directions. Those can be visualized by ellipsoids using three degrees of freedom for their orientation in space and the remaining three for length of its semi-axis.

Starting with the unit sphere centered at the origin again, we compute eigenvector and eigenvalues for every SPD matrix. Due to the SPD property a full basis of eigenvectors with positive eigenvalues always exists. The diagonal matrix formed by the vector of eigenvalues is applied as a scaling transformation of the coordinate axis. The matrix whose columns are the computed eigenvectors can then be interpreted as a rotation (or principal axis transformation of the ellipsoid).

To avoid large size difference and overlaps between the ellipsoids we also normalize the eigenvalues using the mean diffusivity μ defined by

$$\mu = \frac{1}{3} \sum_{i=1}^3 \lambda_i. \quad (3.1)$$

Finally, the color is defined by normalizing the largest eigenvector, called the principal direction, and mapping its coordinates to the RGB colorspace, such that clusters of similar orientations can be more easily visually distinguished.

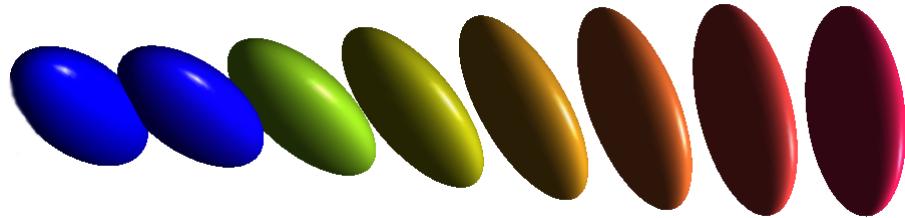


Figure 3.4: SPD(3) Visualization as oriented, colored ellipsoids

In the case of 3D SPD images the rendering window also provides some controls over the view. The up and down arrow keys can be used to zoom in and out of the picture, left and right keys pivot the camera and with the s key the image is saved using the filename specified before.

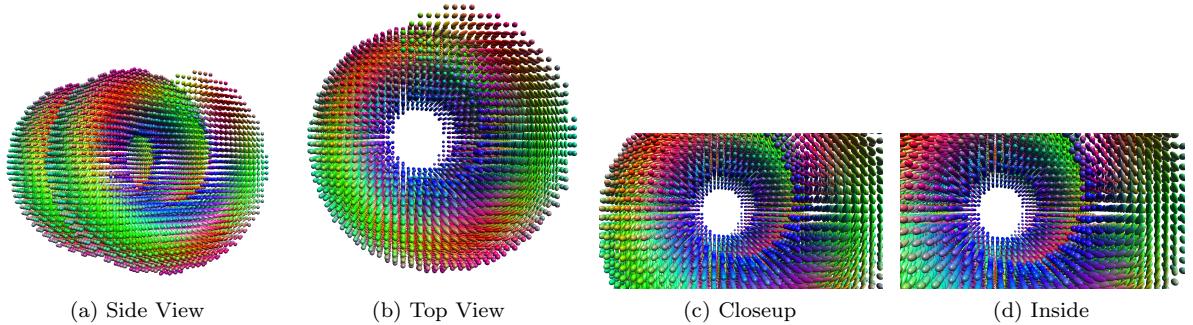


Figure 3.5: Example of the 3D data Visualization of SPD(3) images from different viewpoints.

3D Volume image rendering

The volume image renderer just transforms the data to a 3D texture which is then mapped onto cube rotating about the z-axis. Plasticity is created by setting the alpha channel of each displayed voxel to the intensity value of the corresponding data voxel, such that dark areas are more transparent. The following 3.6 shows the rendered volume from different angles.



Figure 3.6: 3D Volume image using texture based rendering

3.3.6 Utility functions

Matrix functions

In the file `matrix_utils.hpp` additional matrix functions not included in the Eigen library are implemented. So far, these are only the methods for the computation of the Fréchet derivatives of matrix square root and logarithm, as well as their Kronecker representations.

Function pointers utilities

Located in the file `func_ptr_utils.hpp` are some auxiliary functions needed to transform pointers to class member functions to plain C function pointers required by the OpenGL and GLUT library API.

3D pixel-wise kernels

The 3D version of the pixel-wise kernels along with useful tools based on them, for copying or filling 3D images.

3.3.7 External Components

Maybe put somewhere else...

Linear algebra handler: Eigen

Data storage and processing: Video++

3.4 Using TVTML

3.4.1 Prerequisites

The main dependencies of TVTML are the *Eigen* C++ template library for linear algebra and the *Video++* video and image processing library. Those libraries, as well as TVTML's core functionality are provided as header-only libraries. There are, however, some more recommended static libraries that are recommended to speed up the computation, enable easy I/O or are needed for visualization of the results. To administer all these different parts and because the header-only libraries require additional compiler flags for the code optimization, TVTML also relies on the *CMake* installation tool for installation and compilation of user code using TVTML.

The following lists shows the needed packages for the usage of TVTML:

- CMake ($\geq 2.8.0$)
- g++ ($\geq 4.9.1$), any C++14 compatible compiler should also be possible but is untested.
- Eigen ($\geq 3.2.5$)
- Video++

Recommended are also the following packages. They are needed if any of the described extended functionality needs to be used.

- OpenCV ($\geq 2.4.9$), for image input and output, edge detection for inpainting

- CGAL (≥ 4.3), for first guess interpolation during inpainting
- OpenGL (≥ 7.0), for visualizations of SPD, SO and any 3D data
- SuiteSparse ($\geq 4.2.1$), faster parallel sparse solver for the linear system in the IRLS algorithm

3.4.2 Installation

3.4.3 Compilation of own projects using CMake

3.4.4 Tutorial and typical use cases

The basic process of using the library is to explicitly specify the necessary template parameters for all needed components. For the sake of compactness and readability this should be done using `typedefs`. In the next step one can then instantiate the classes and start implementing.

Image denoising, vectorial color model

As a first example we show the denoising of a simple color picture using the IRLS minimizer. In a first step the necessary classes need to be included. For the sake of shortening the code we also switch to the `tvm` namespace of the library.

Listing 3.4 Inclusion of library headers

```

1 #include "../core/algo_traits.hpp"
2 #include "../core/tvmin.hpp"
3
4 using namespace tvm;

```

Next, we specify the manifold type and data type we want to use, in this case Euclidian \mathbb{R}^3 and a corresponding 2D image container

Listing 3.5 Specification of manifold and data type

```

1 typedef Manifold< EUCLIDIAN, 3 > mf_t;
2 typedef Data< mf_t, 2> data_t;

```

Note that the data type must be specified using the *fully* specialized manifold class type we defined in the line before.

Our data type is now ready for work such that we can read the input data in the next few lines.

Listing 3.6 Initialization and input of image data

```

1 data_t myData=data_t();           // Creating the data object
2 myData.rgb_imread(filename);     // Reading an image file , filename is a const char*

```

After the data object is ready we must specify the functional we want to use, which we choose to be first order TV, isotropic and 2D. Again, also the fully specified manifold and data class types need to be given as template parameter. The last template parameter, the dimension of the data, has default value 2 and can also be omitted in this case.

Listing 3.7 Defining the functional and setting parameters

```

1 typedef Functional<FIRSTORDER, ISO, mf_t, data_t, 2> func_t;
2
3 func_t myFunc(lambda, myData);    // Creation of the functional object
4 myFunc.seteps2(1e-10);           // Specify the epsilon parameter

```

For the instantiation of the functional we need to pass the λ for our functional as well as our newly created data object. The `seteps2` method sets the value of ϵ^2 for the reweighting computation. In case of the proximal point algorithm it should be set to zero.

Listing 3.8 Choosing the minimizer, smoothing and minimization

```
1 typedef TV_Minimizer< IRLS, func_t, mf_t, data_t, OMP, 2> tvmin_t;
2
3 tvmin_t myTVMIn(myFunc, myData); // Creation of minimizer object
4
5 myTVMIn.smoothening(5);           // smoothing to obtain better starting value
6 myTVMIn.minimize();              // Starts the minimization
```

Finally we choose the minimizer we want to use, in this case IRLS, and pass functional, manifold and data types as template parameters. The OMP parameter is not fully implemented yet and is supposed to provide choice between different parallelization schemes or also completely serial computation. The last parameter again has default value 2 and describes the dimension of the data. The complete listing of this example can be found in ??.

Colorization using color inpainting

In the following we show a more complicated example: Recolorization of an image where most ($\approx 99\%$) *color* information has been removed. This means that this problem is defined on the product manifold $S^2 \times \mathbb{R}$, optimization however will only take place on S^2 while the \mathbb{R} data part is only needed to obtain edge information. We also three auxiliary functions (removeColor, DisplayImage, recombineAndShow) here that are not shown in the code snippets but will be included in the full listing in the appendix ???. This time, we start by also obtaining some of the minimization parameters from the command line:

Listing 3.9 Include library files and read parameters from standard input

```
1 #include <iostream>
2 #include <string>
3 #include <cmath>
4
5 #include <opencv2/highgui/highgui.hpp>
6 #include " ../core/algo_traits.hpp"
7 #include " ../core/data.hpp"
8 #include " ../core/functional.hpp"
9 #include " ../core/tvmin.hpp"
10
11 #include <vpp/vpp.hh>
12 #include <vpp/utils/opencv_bridge.hh>
13
14 using namespace tvmtl;
15
16 int main(int argc, const char *argv[])
17 {
18     if (argc < 3){
19         std::cerr << "Usage : " << argv[0] << " image [lambda] [threshold]" << std::endl;
20         return 1;
21     }
22
23     double lam=0.01;
24     double threshold=0.01;
25
26     if (argc == 4){
27         lam=atof(argv[2]);
28         threshold=atof(argv[3]);
29     }
30
31     std::string fname(argv[1]);
32
33     // ...
34 }
```

Here threshold defines the percentage of color information that remains in the picture.

In the next step, we again define the necessary type definitions for manifold and data classes and create our data objects.

Listing 3.10 Manifold and Data class type definitions and instantiation

```
1 // typedefs
2 typedef Manifold< SPHERE, 3 > spheremf_t; // S^2
3 typedef Manifold< EUCLIDIAN, 1 > eucmf_t; // R
4
```

```

5 typedef Data< spheremf_t, 2> chroma_t; // Chromaticity part
6 typedef Data< eucmf_t, 2> bright_t; // Brightness part
7
8 // Instantiation
9 chroma_t myChroma=chroma_t();
10 bright_t myBright=bright_t();

```

When the data containers are ready we need to read the input picture, extract color and brightness information and store it in the respective objects. This problem is basically a color inpainting problem but we do not want the reconstructed color to blur across edges in the picture. This can be solved by making use of the edge weights array that is stored together with the image. We will detect edged in the brightness part of the picture and and use those edges in the Chromaticity denoising procedure.

Finally, we will remove the color in the following way: Create a random inpainting matrix where the probability a certain pixel is set to false is given by the threshold variable and then replace every RGB pixel by the mean of its three color components (those pixels are basically gray scale then).

The necessary steps are shown in the next listing

Listing 3.11 Color and brightness input, edge detection and color removal

```

1 myBright.rgb_readBrightness(argv[1]); // Extract brightness from filename argv[1]
2 myBright.findEdgeWeights(); // Detect edges and store in matrix
3
4 myChroma.rgb_readChromaticity(argv[1]); // Extract chromaticity from filename argv[1]
5 myChroma.inpaint_=true; // Turn inpainting on
6 myChroma.setEdgeWeights(myBright.edge_weights_); // Initiliazie chromaticity part edges with brightness part edges
7 myChroma.createRandInpWeights(threshold); // Create random inpainting matrix
8 removeColor(myChroma, myBright); // Remove color
9
10 // Recombine chromaticity and brightness and show the colorless image
11 recombineAndShow(myChroma, myBright, "colorless_" + fname, "Colors removed Picture");

```

The next part works almost exactly as in the last example. We define functional, set its parameters, then define the minimizer. The only difference is that we have to run first_guess before the minimization.

Listing 3.12 Functional and minimizer definition, first guess and minimization

```

1 typedef Functional<FIRSTORDER, ISO, spheremf_t, chroma_t> cfunc_t;
2 typedef TV_Minimizer< IRLS, cfunc_t, spheremf_t, chroma_t, OMP > ctvmin_t;
3
4 cfunc_t cFunc(lam, myChroma); // create functional object
5 cFunc.seteps2(1e-10); // set eps^2 parameter
6
7 ctvmin_t cTVMin(cFunc, myChroma); // create minimizer object
8 cTVMin.first_guess(); // first guess
9
10 std::cout << "Start TV minimization..." << std::endl;
11 cTVMin.minimize();
12
13 // Recombine Brightness and Chromaticity parts of recolored Picture
14 recombineAndShow(myChroma, myBright, "recolored_" + fname, "Recolored Picture");

```

Some visual results of the above code are also shown in section ??.

3D DT-MRI data denoising and visualization

As a final example we choose a more complicated manifold, SPD(3) in this case, as well as 3D data to also demonstrate the use of the visualization classes. Moreover, we use the proximal point algorithm in this example. The CSV reader just reads a list of pixels where the numerical values comprising the pixel are stored comma-separated one pixel per line. The CSV files has no header such that we provide the dimensions as command line parameters.

Listing 3.13 Initialization

```

1 #include <iostream>
2 #include <cstdlib>
3 #include <string>
4 #include <sstream>

```

```

5 #include "../core/algo_traits.hpp"
6 #include "../core/data.hpp"
7 #include "../core/functional.hpp"
8 #include "../core/tvmin.hpp"
9 #include "../core/visualization.hpp"
10
11 int main(int argc, const char *argv[])
12 {
13     int nz, ny, nx;
14     nz = std::atoi(argv[2]);
15     ny = std::atoi(argv[3]);
16     nx = std::atoi(argv[4]);
17
18     std::stringstream fname;
19     std::string nfname;
20     fname << "dti3d" << nz << "x" << ny << "x" << ny << ".png";
21     nfname = "noisy_" + fname.str();
22
23     // ...
24
25     return 0;
26 }

```

Since the meaning of the individual components should be clear by now do all the necessary type definitions at once in the next listing

Listing 3.14 Type definitions, Visualization type

```

1 using namespace tvmtl;
2
3 typedef Manifold< SPD, 3 > mf_t;
4 typedef Data< mf_t, 3> data_t;
5 typedef Functional<FIRSTORDER, ANISO, mf_t, data_t, 3> func_t;
6 typedef TV_Minimizer< PRPT, func_t, mf_t, data_t, OMP, 3 > tvmin_t;
7 typedef Visualization<SPD, 3, data_t, 3> visual_t;

```

The only innovation is the aforementioned Visualization class. The first 3 in its template parameter list is the embedding dimension of the manifold and the last 3 denotes the dimension of the data. Note that we needed to specify it for the functional and minimizer classes as well because the default value is 2. The remaining parameters specify the manifold type via an enumeration constant (in the same way one specifies it for the Manifold class) and the data type via a fully specialized data class type.

Before we start the minimization we want to display the original noisy data and eventually save it to a file once we have found a nice viewing angle in the rendering window. The necessary steps are as follows:

Listing 3.15 Data input and displaying the noisy data

```

1 data_t myData = data_t(); // Create data object
2 myData.readMatrixDataFromCSV(argv[1], nz, ny, nx); // Read from CSV file
3
4 visual_t myVisual(myData); // Create visualization object
5 myVisual.saveImage(nfname); // Specify file name to save a screenshot
6
7 std::cout << "Starting OpenGL-Renderer..." << std::endl;
8 myVisual.GLInit("SPD(3) Ellipsoid Visualization"); // Start the Rendering

```

In the last step we create functional and minimizer class, perform the minimization and display the denoised data again.

Listing 3.16 Minimization and final rendering

```

1 double lam=0.7;
2 func_t myFunc(lam, myData); // Functional object
3 myFunc.seteps2(0); // eps^2 should be 0 for PRPT
4
5 tvmin_t myTVMIn(myFunc, myData); // Minimizer object
6
7 std::cout << "Start TV minimization.." << std::endl;
8 myTVMIn.minimize();
9
10 std::string dfname = "denoised(prpt)_"; + fname.str();
11 myVisual.saveImage(dfname); // Specify name for denoised image

```

```
12  
13 std::cout << "Starting OpenGL-Renderer..." << std::endl;  
14 myVisual.GLInit("SPD(3) Ellipsoid Visualization"); // Render
```

The resulting picture for this example are also shown in 4.3.3.

Chapter 4

Applications and Numerical Experiments

In the first part of the chapter we apply the algorithms to a variety of problems in image processing, computer vision, medical imaging and related fields. The second part will be a comparison of the performance of the two implemented minimizer and close with with an numerical experiment investigating the dependence of the solution on the changes in the original picture. This is a first step in extending the IRLS algorithm towards recursive splitting into subdomain.

As briefly mentioned in the previous section, the test platform is a linux machine with two hyperthreaded 2.8GHz cores Intel i5-2520 (thus a total of four hardware threads) with AVX vector extensions and 8 GB RAM.

4.1 Image denoising

The very basic application of the algorithm is of course denoising of common 2D grayscale or color pictures. For grayscale pictures the TV minimization is performed over the Euclidian manifold $M = \mathbb{R}$, while for color picture, as already explained in the introduction, we have either $M = \mathbb{R}^3$ for the linear-vectorial model or $M = S^2 \times \mathbb{R}$ for the non-linear chromaticity-brightness model.

4.1.1 Grayscale

4.1.2 Color

In this example we perform the TV minimization of color images using the two different color models. In Figure 4.1 the among the image processing community well-known *Lena* picture, which is rather small in size. Minimization in the linear-vectorial color model using 5 IRLS iterations with one Newton step per reweighting is completed within *XXX* seconds.

Next, using the same model and parameters we denoise a different image with a size already in the megapixel range. The needed time, however, is with *XXX* seconds quite high. The result can be seen in Figure 4.2.

Finally, in Figure ?? we denoise a third picture using the chromaticity-brightness model. Here minimization over the product manifold $S^2 \times \mathbb{R}$ is performed by denoising the chromaticity(S^2) and the brightness(\mathbb{R}) separately, which has the added advantage of more fine-grained control over the process because two λ parameters can be chosen separately for each part, too.

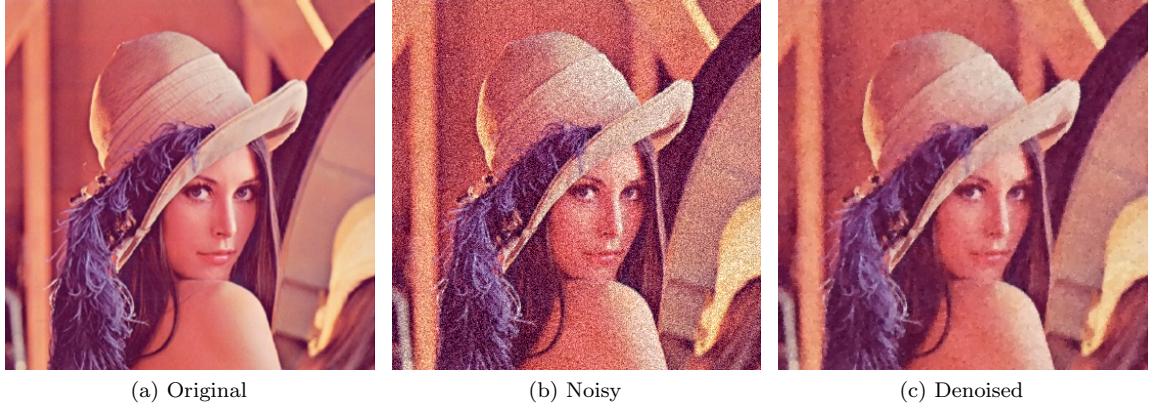


Figure 4.1: Denoising of a color images using the linear vectorial color model which corresponds to the manifold \mathbb{R}^3 (a) Original image "Lena.jpg", 361×361 px, 8 bit color depth (b) Componentwise gaussian noise $\mu = 0$, $\sigma = ?$ added (c) Denoised, IRLS with $\lambda = ?$, 5 IRLS steps, 1 newton steps per IRLS step

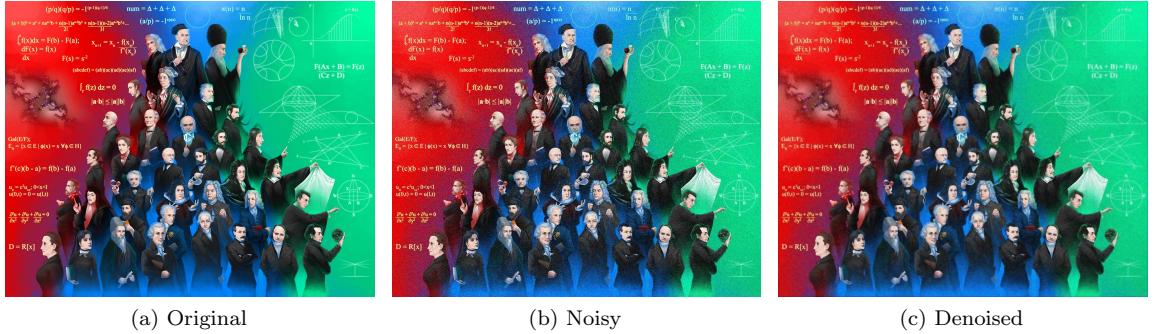


Figure 4.2: Denoising of a color images using the linear vectorial color model which corresponds to the manifold \mathbb{R}^3 (a) Original image "mathematicians.jpg", 1280×1024 px, 8 bit color depth (b) Componentwise gaussian noise $\mu = 0$, $\sigma = ?$ added (c) Denoised, IRLS with $\lambda = ?, ?$ IRLS steps, ? newton steps per IRLS step

4.1.3 Inpainting

We consider a damaged picture were a considerable part of the picture has been overpainted with blue color. In the first step we detect the damaged region which in this case is done via a simple color selector (e.g. all pixels with a blue value larger than 0.95). In principle many other selection methods known from common raster graphic editors could be implemented here as well.

Next, the a first guess is calculated using scattered linear interpolation and lastly the TV minimization itself is performed. The process is summarized in Figure 4.3.

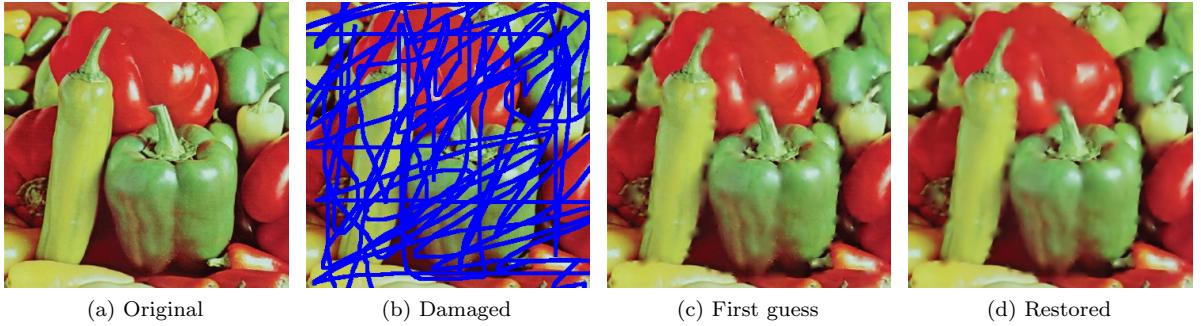


Figure 4.3: Denoising of a color images using the linear vectorial color model which corresponds to the manifold \mathbb{R}^3 (a) Original image "Pepper.png", 359×361 px, 8 bit color depth (b) Damaged by overpainting with blue color (c) First guess via componentwise scattered interpolation (d) Restored, IRLS with $\lambda = ?$, ? IRLS steps per IRLS step

4.1.4 Recolorization

Colorization, also known as color inpainting, because it is basically just a special of inpainting is performed in the next example. Here the picture is not necessarily noisy but we assume only the brightness of each pixel is known, while the chromaticity is known only for a low ratio $r = 0.01$ of all pixels. Note that this splitting implies that inpainting and TV minimization takes place only S^2 .

As in the previous example, we first have to detect all damaged, i.e. non-colored, pixels to inpaint. Again we use scattered interpolation to obtain the first guess which is depicted in Figure 4.3 (c). One can observe that the color runs over the edges of leaves. To avoid this we need to detect the edges in the brightness part using the Canny edge detector [8], for example, and set the edge weights for the chromaticity part accordingly. As a result, we indeed obtain sharp and clear edges in the final result 4.3 (d).



Figure 4.4: Recolorization using color inpainting in the Chromaticity-Brightness color model, corresponding to $S^2 \times \mathbb{R}$ (a) Original image "Basil.jpg", 300×179 px, 8 bit color depth (b) Image with a ratio of approximately 0.01 remaining colored pixels (c) First guess via componentwise scattered interpolation (d) Recolored, IRLS with $\lambda = ?$, ? IRLS steps per IRLS step

4.1.5 Volume images

We conclude the picture section with an example of a 3D volume image as they might occur in medical imaging from magnetic resonance imaging (MRI) or computed tomography. In this demonstration, however, we chose the example of the so-called *Boston teapot*, taken a volume image library [?] and added componentwise gaussian noise. The image represents only intensity values, hence minimization is performed over \mathbb{R} . The result are shown in Figure 4.5a.

For this picture we used the proximal point algorithm, because the memory and computational requirements of the IRLS for a picture of this size are very high: In section it was shown that the dimension of the sparse linear system is $\dim(M)XYZ$ which in this case amounts to 1.1×10^7 , which

is the length of the gradient while the Hessian will contain 7.7×10^7 non-zero-entries. The solution of a sparse linear system of that size is computationally very demanding while in comparison the geodesic averaging and karcher mean calculations simplify to mostly vectorized addition and subtraction operations on a simple manifold like \mathbb{R} .

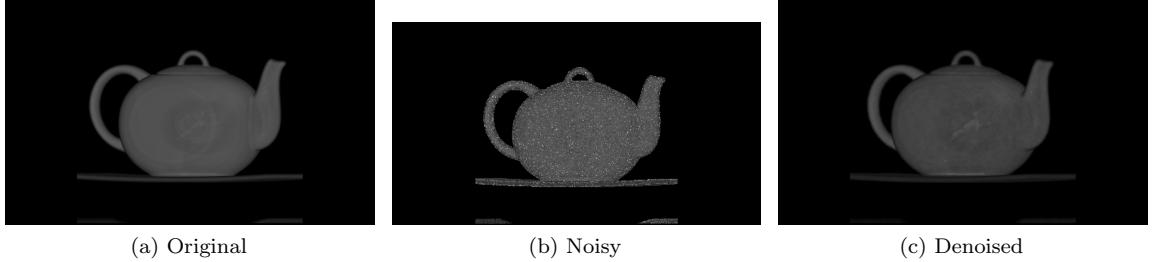


Figure 4.5: Denoising a 3D grayscale volume image (a) Original image "BostonTeapot.raw", $256 \times 256 \times 178$ px, 8 bit color depth (b) Componentwise gaussian noise $\mu = 0$, $\sigma = ?$ added (c) Denoised, Proximal point with $\lambda = ?$, 50 PRPT steps

4.2 SO(2) and SO(3) images data

4.2.1 Synthetic data

The following synthetic $SO(3)$ image is constructed in the following way. Let $\Omega = \{1, \dots, 30\}^2$ and define for every $(i, j) \in \Omega$ a rotation axis

$$v = \begin{cases} (2x, y, 0)^T, & x > 0.5 \\ (0, 2x, 0.5)^T, & \text{else} \end{cases}, \quad (4.1)$$

where $x = \frac{j}{30}$, $y = \frac{i}{30}$ and a rotation angle

$$\alpha = \begin{cases} x + y, & x > y \\ \frac{\pi}{2} + x - y, & \text{else} \end{cases}. \quad (4.2)$$

Then assign the corresponding $SO(3)$ element representing a rotation by α and about v . Noise is added componentwise and the noisy matrix is then projected back to $SO(3)$ using a the projector $P_{SO(n)}(A) = UV^T$ where $A = U\Sigma V^T$ is the singular value decomposition of A .

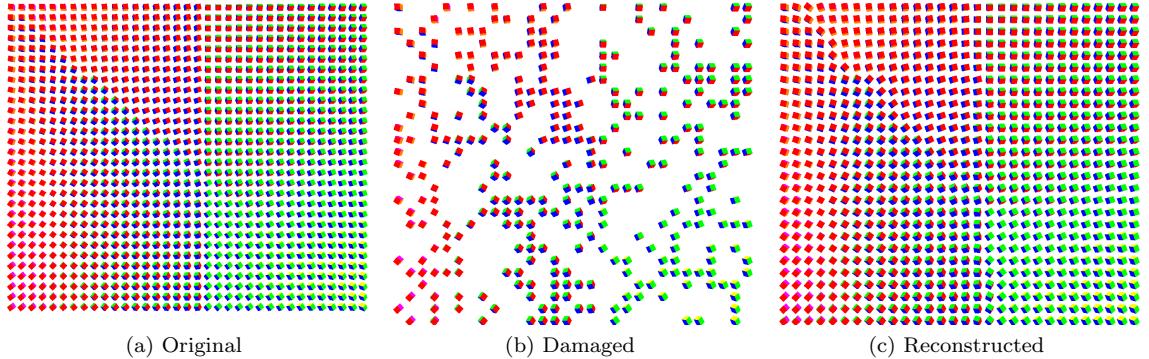


Figure 4.6: Inpainting of synthetic $SO(3)$ picture (a) Original image: Synthetic, nonsmooth $SO(3)$, 30×30 px (b) Threshold $p = ?$ (c) Denoised, IRLS with $\lambda = ?$, 5 IRLS steps, 1 Newton step per IRLS

4.2.2 Fingerprint orientation data

Fingerprint matching is based on extracting a set of particular features, called *minutiae*, which uniquely define the fingerprint. These features are usually ridge endpoint or ridge bifurcation points that are saved along with their position and orientation. This means that prior to minutia detection and extraction the calculation of an orientation field is necessary.

For pictures of fingerprint this is just a special form of edge detection which can be done by calculating Sobel derivatives for every pixel. Depending on the quality and noise level of the picture the computed orientation field can be very noisy itself which is another application for our TV algorithms.

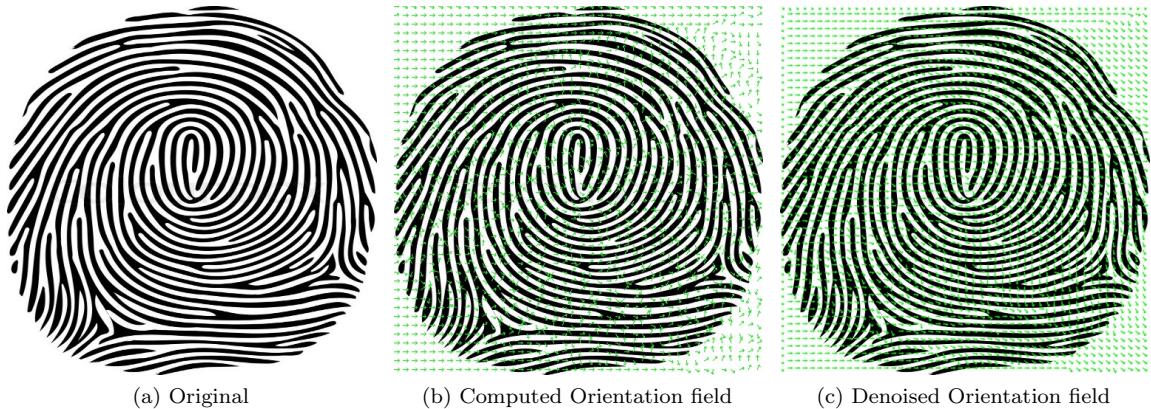


Figure 4.7: Denoising a orientation field from a fingerprint, orientations represented by $\text{SO}(2)$ elements (a) Original fingerprint (b) Orientation field computed using Sobel derivatives PLACEHOLDER (c) Denoised, IRLS with $\lambda = ?$, ? IRLS steps, ? Newton steps per IRLS step PLACEHOLDER

4.2.3 Reconstruction of a dense optical flow field

An optical flow is the pattern of apparent motion between two consecutive frame of a video sequence. This may be the result of either an actual movement of the depicted object or the result of a moving camera. Important applications are for example (anormal) motion detection, crowd behavior analysis, surveillance, video compression or image segmentation.

A *dense* optical flow field can be interpreted as a vector field where each vector describes the displacement of a point between from one frame to the next. If the set of points is restricted to only a few points of interest, a sparse feature set, we have *sparse* optical flow.

In the following example, we only use a sparse feature set for tracking and flow computation in a short video sequence. The traffic scene was taken from a crowds/high density moving object data set provided by [6]. At first, we compute the sparse optical flow using the Lucas-Kanade algorithm [19] implemented in the OpenCV library.

For the set of tracked features $\mathcal{F}_1 := \{F_i^{(1)}\}_{i=1}^{400} \subset \Omega \subset \mathbb{R}^2$ in the first frame the algorithm tries to identify each feature in the second frame resulting in a set of identified features $\mathcal{F}_2 := \{F_i^{(2)}\}_{i=1}^{N < 400} \subset \Omega \subset \mathbb{R}^2$ and corresponding displacement vectors $\mathcal{V}_{12} := \{V_i \mid V_i = F_i^{(2)} - F_i^{(1)}\}_{i=1}^N$.

We now assign to each pixel in our data an $SO(2)$ element in the following way

$$\alpha_i = \arctan\left(\frac{V_i^y}{V_i^x}\right) \quad (4.3)$$

$$I(i, j) = \begin{cases} \begin{pmatrix} \cos \alpha_i & -\sin \alpha_i \\ \sin \alpha_i & \cos \alpha_i \end{pmatrix} & (i, j) \in \mathcal{F}_2 \\ 0 & \text{otherwise} \end{cases} \quad (4.4)$$

Since we want to reconstruct the dense flow, this is an inpainting problem and we have to perform scattered interpolation before running the algorithm. The result can be seen in Figure 4.8.

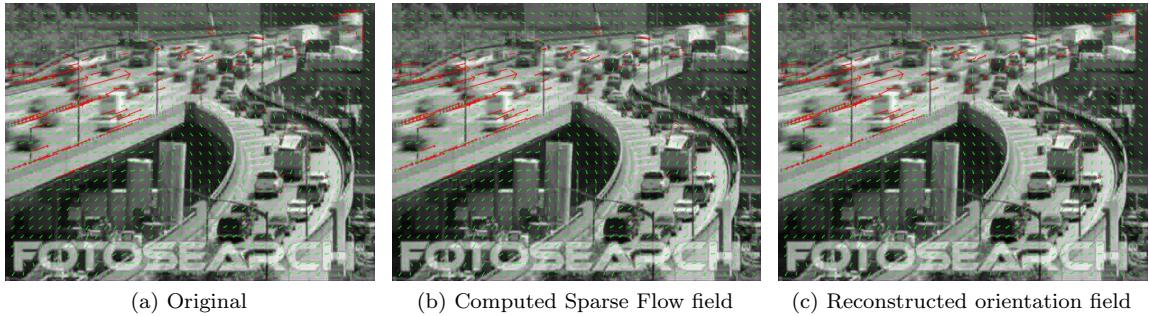


Figure 4.8: Reconstructing a dense flow from sparse feature tracking, orientations represented by $SO(2)$ elements (a) Frame of original video scene PLACEHOLDER (b) Sparse features tracked using PLACEHOLDER (c) Reconstructed, IRLS with $\lambda = ?$, ? IRLS steps, ? Newton steps per IRLS step

Of course the optimization could have also been performed on S^1 . Furthermore, there is also a more direct, variational approach for the calculation of the flow field which is also based on TV minimization but has a different fidelity term. This is one possibility for further extension of the library and is discussed in more detail in section 5.1

4.3 SPD(3) image data

4.3.1 Synthetic data

For the construction of the synthetic $SPD(3)$ image in Figure 4.9, let $\Omega = \{1, \dots, n\}^2$ and define for every $(i, j) \in \Omega$ a rotation axis

$$v = \begin{cases} (x, y, 2)^T, & x + y < 1 \\ (y, -x, 1)^T, & \text{else} \end{cases}, \quad (4.5)$$

where $x = \frac{j}{n}$, $y = \frac{i}{n}$ and a rotation angle

$$\alpha = \begin{cases} x + 2y, & x + y < 1 \\ y + 2x, & \text{else} \end{cases}. \quad (4.6)$$

Let R be the corresponding $SO(3)$ element representing a rotation by α and about v . Then define a diagonal matrix $D = \text{diag}(x + 0.2, y + 0.2, 0.5)$ and finally assign the matrix $A = R^T D R$ to the pixel.

Noise is then added by taking the matrix logarithm of every pixel, adding gaussian componentwise noise and apply the matrix exponential again.

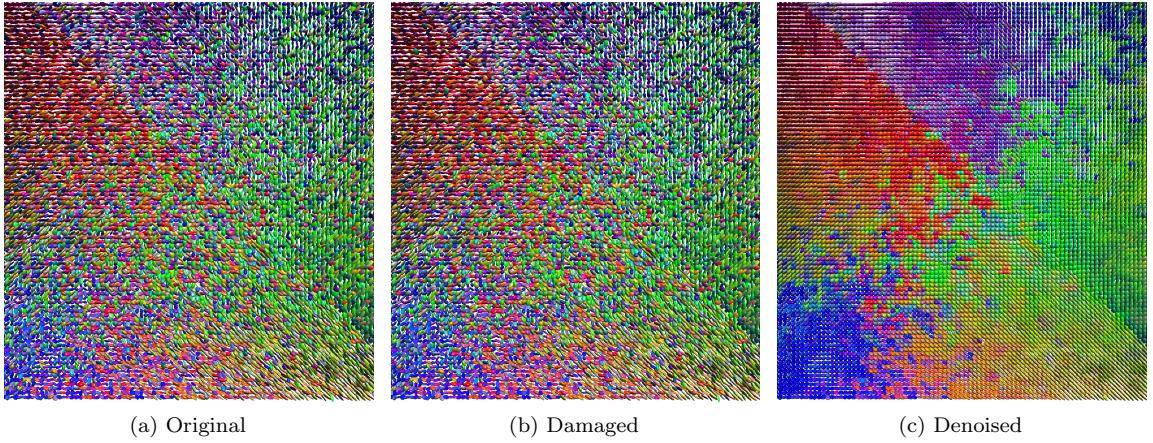


Figure 4.9: Denoising of synthetic $SPD(3)$ picture (a) Original image: Synthetic, nonsmooth $SPD(3)$, 100×100 px PLACEHOLDER (b) Threshold $p = ?$ (c) Denoised, IRLS with $\lambda = ?, 5$ IRLS steps, 1 Newton step per IRLS

4.3.2 Diffusion Tensor Magnetic Resonance Imaging

Diffusion Tensor Magnetic Resonance Imaging (DT-MRI) is a medical imaging method which is able to non-invasively measure diffusion coefficients of water molecules in living biological tissues. DT-MRI goes beyond CT or normal MR imaging methods which are only able to provide a single intensity value per voxel. Since water molecules can move easier along, for example axons, connecting the neurons in the brain, than they can move across it, the resulting anisotropic diffusion pattern can provide a lot of information about the structure of the brain.

DTI data sets are usually calculated from a set of diffusion weighted magnetic resonance imaging (DW-MRI) pictures. The basic magnetic resonance imaging works by applying an external magnetic along the z -axis such that the proton spins in the tissue align either parallel or antiparallel to it, however, while still precessing around the z -axis with the so-called Lamor frequency. An electromagnetic wave packet(HF-pulse) with that frequency leads to a collective state transition to a resonance state such that spin moments will be phase-synchronous before relaxing back to their original orientation with respect to the external field. The magnetic field created by having synchronized moments can be measured by a coil where an electric potential will be created. From the different relaxation times of different materials one can make conclusions about the structure of the tissue.

By applying an additional magnetic gradient field, the Lamor frequency of different layers of the probe can be modified such that only one layer of the material will resonate to the pulse. This provides an additional positional resolution of the imaging process.

The DTI image is finally computed using the *Stejskal-Tanner-equation* given by

$$A(\mathbf{g}) = A(0) \exp(-b\mathbf{g}^T \mathbf{D}\mathbf{g}) \quad (4.7)$$

where $A(\mathbf{g})$ denotes the signal strength, \mathbf{g} the magnetic gradient field and b some measurement related parameters. Solving this equation for \mathbf{D} finally leads to desired $SPD(3)$ matrix describing the diffusion coefficients and directions.

In Figure 4.10 the IRLS minimizer is applied to DTI data set provided by Barmpoutis [1]. On clearly identify regions of high anistropy in the picture, where the molecules are forced to diffuse in one preferred direction.

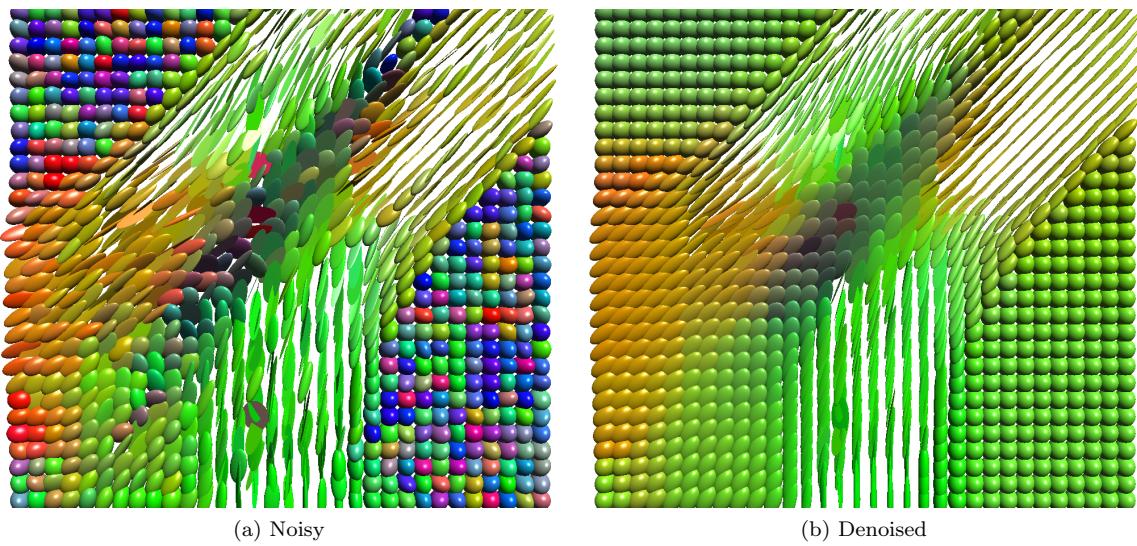


Figure 4.10: Denoising a DT-MRI image with pixel in $SPD(3)$ (a) Original DTI data, 32x32 pixel
 (b) Denoised, IRLS with $\lambda = ?$, ? IRLS steps, ? newton steps per IRLS step

4.3.3 3D DT MRI data

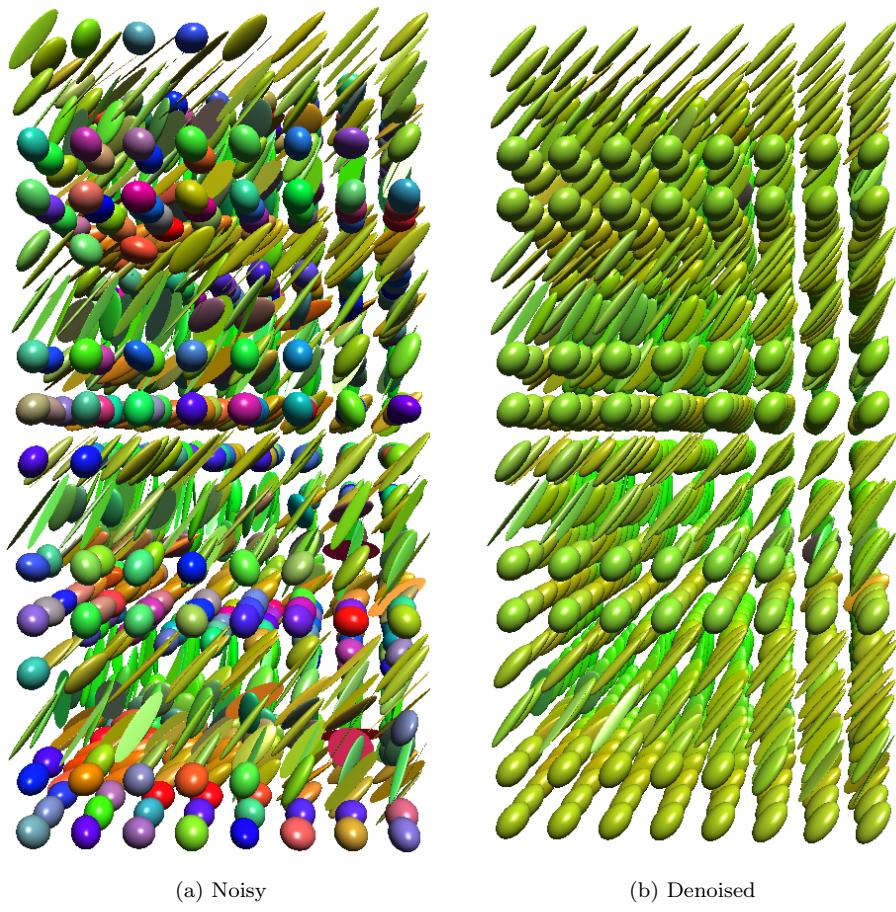


Figure 4.11: Denoising a 3D DT-MRI image with pixel in $\text{SPD}(3)$ (a) Componentwise gaussian noise $\mu = 0$, $\sigma = ?$ added (b) Denoised, Proximal point with $\lambda = ?$, 50 PRPT steps

4.4 Comparison IRLS and Proximal Point minimizers

4.5 Sensitivity to variations of the starting value

Chapter 5

Outlook

5.1 Extensions

- new manifolds and functionals
- automatic differentiation for matrix valued functions
- distributed memory architecture

5.2 Recursive computation on subdomains

Bibliography

- [1] A. Barmpoutis A., B. C. Vemuri, T. M. Shepherd, and J. R. Forder. Tensor splines for interpolation and approximation of dt-mri with applications to segmentation of isolated rat hippocampi. *IEEE Trans Med Imaging*, 26(11):1537–1546, 2007. (Cited on page 46.)
- [2] Absil. *Optimization on Matrix Manifolds*. Butterworth-Heinemann, 3 edition, January 1981. (Cited on pages 9, 11 and 12.)
- [3] P. A. Absil, R. Mahony, and R. Sepulchre. *Riemannian Geometry of Grassmann Manifolds with a View on Algorithmic Computation*, volume 80. Springer Netherlands, January 2004. (Cited on pages 18 and 20.)
- [4] P. A. Absil, R. Mahony, and R. Sepulchre. *Optimization algorithms on matrix manifolds*. Princeton University Press, 2009. (Cited on page 19.)
- [5] Awad H. Al-Mohy, Nicholas J. Higham, and Samuel D. Relton. Computing the fréchet derivative of the matrix logarithm and estimating the condition number. *SIAM Journal on Scientific Computing*, 35(4):C394–C410, 2013. (Cited on page 22.)
- [6] Saad Ali and Mubarak Shah. A lagrangian particle dynamics approach for crowd flow segmentation and stability analysis. In *CVPR*. IEEE Computer Society, 2007. (Cited on page 44.)
- [7] V. Arsigny, P. Fillard, X. Pennec, and N. Ayache. Geometric means in a novel vector space structure on symmetric positive definite matrices. *SIAM Journal on Matrix Analysis and Applications*, 29(1):328–347, 2007. (Cited on page 9.)
- [8] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(6):679–698, 1986. (Cited on page 42.)
- [9] M. Cavegn. Total variation regularization for geometric data. *Master thesis, ETH Zürich*, 2013. (Cited on page 24.)
- [10] A. Chambolle and P.-L. Lions. Image recovery via total variation minimization and related problems. *Numerische Mathematik*, 76(2):167–188, 1997. (Cited on page 5.)
- [11] PatrickL. Combettes and Jean-Christophe Pesquet. Proximal splitting methods in signal processing. In Heinz H. Bauschke, Regina S. Burachik, Patrick L. Combettes, Veit Elser, D. Russell Luke, and Henry Wolkowicz, editors, *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, volume 49 of *Springer Optimization and Its Applications*, pages 185–212. Springer New York, 2011. (Cited on page 9.)
- [12] Alan Edelman, Tomás A. Arias, and Steven T. Smith. The geometry of algorithms with orthogonality constraints. *SIAM J. Matrix Anal. Appl.*, 20(2):303–353, April 1999. (Cited on page 18.)

- [13] Matthieu Garrigues and Antoine Manzanera. Video++, a modern image and video processing c++ framework. Technical report, ENSTA-ParisTech, France, 2014. (Cited on page 30.)
- [14] P. Grohs, M. Sprecher, and T. Yu. Scattered manifold-valued data approximation. Technical Report 23, Seminar for Applied Mathematics, ETH Zürich, Switzerland, 2014. (Cited on page 7.)
- [15] Nicholas J. Higham. Evaluating padé approximants of the matrix logarithm. *SIAM Journal on Matrix Analysis and Applications*, 22(4):1126–1135, 2001. (Cited on page 22.)
- [16] S. H. Kang and R. March. Variational models for image colorization via chromaticity and brightness decomposition. *IEEE Transactions on Image Processing*, 16(9):2251–2261, 2007. (Cited on page 7.)
- [17] H. Karcher. Riemannian center of mass and mollifier smoothing. *Communications on Pure and Applied Mathematics*, 30(5):509–541, 1977. (Cited on pages 10 and 15.)
- [18] Stanley Osher Leonid I. Rudin and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–168, 1992. (Cited on pages 5, 6 and 7.)
- [19] Bruce D. Lucas and Takeo Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th International Joint Conference on Artificial Intelligence - Volume 2*, IJCAI’81, pages 674–679, San Francisco, CA, USA, 1981. Morgan Kaufmann Publishers Inc. (Cited on page 44.)
- [20] Jan R. Magnus and Heinz Neudecker. *Matrix differential calculus with applications in statistics and econometrics*. J. Wiley & Sons, Chichester, New York, Weinheim, 1999. (Cited on page 16.)
- [21] M. Nägelin. Total variation regularization for tensor valued images. *Bachelor thesis, ETH Zürich*, 2014. (Cited on pages 9 and 24.)
- [22] Neal Parikh and Stephen Boyd. Proximal algorithms. *Found. Trends Optim.*, 1(3):127–239, January 2014. (Cited on page 9.)
- [23] P. Rodriguez and B. Wohlberg. An iteratively reweighted norm algorithm for minimization of total variation functionals. *IEEE Signal Processing Letters*, 14(12):948 – 951, 2007. (Cited on page 10.)
- [24] Hiroyuki Sato and Toshihiro Iwai. Optimization algorithms on the grassmann manifold with application to matrix eigenvalue problems. *Japan Journal of Industrial and Applied Mathematics*, 31(2):355–400, 2014. (Cited on page 18.)
- [25] Andreas Wedel and Daniel Cremers. *Stereo Scene Flow for 3D Motion Analysis*. Springer, 2011. (Cited on page 6.)
- [26] A. Weinmann, L. Demaret, and M. Storath. Total variation regularization for manifold-valued data. *SIAM Journal on Imaging Sciences*, 7(4):2226–2257, 2014. (Cited on pages 7, 9 and 10.)