

Optimization on Manifolds for Models using Second Order Differences

Ronny Bergmann*

Technische Universität Chemnitz

Kolloquium des Instituts für Mathematik,
Universität zu Lübeck,

June 6, 2019.

*joint with with M. Bačák, P.-Y. Gouzenbourger, F. Laus, J. Persch, G. Steidl, A. Weinmann

Contents

1. Introduction
2. Second Order Differences
3. Second Order Total Variation
4. Acceleration of Bézier Curves

1. Introduction

Manifold-Valued Images

New data acquisition modalities lead to non-Euclidean range

- Interferometric synthetic aperture radar (InSAR)
- Surface normals, GPS data, wind, flow,...
- Diffusion tensors in magnetic resonance imaging (DT-MRI), covariance matrices
- Electron backscattered diffraction (EBSD)



InSAR-Data of Mt. Vesuvius

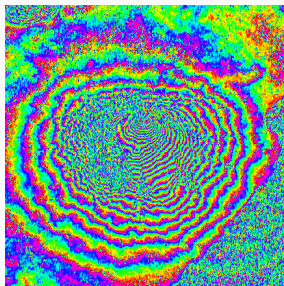
[Rocca, Prati, Guarneri, 1997]

phase-valued data, $\mathcal{M} = \mathbb{S}^1$

Manifold-Valued Images

New data acquisition modalities lead to non-Euclidean range

- Interferometric synthetic aperture radar (InSAR)
- Surface normals, GPS data, wind, flow,...
- Diffusion tensors in magnetic resonance imaging (DT-MRI), covariance matrices
- Electron backscattered diffraction (EBSD)



InSAR-Data of Mt. Vesuvius

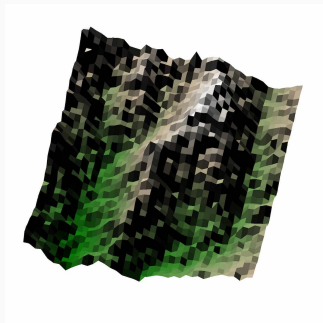
[Rocca, Prati, Guarneri, 1997]

phase-valued data, $\mathcal{M} = \mathbb{S}^1$

Manifold-Valued Images

New data acquisition modalities lead to non-Euclidean range

- Interferometric synthetic aperture radar (InSAR)
- Surface normals, GPS data, wind, flow,...
- Diffusion tensors in magnetic resonance imaging (DT-MRI), covariance matrices
- Electron backscattered diffraction (EBSD)



National elevation dataset

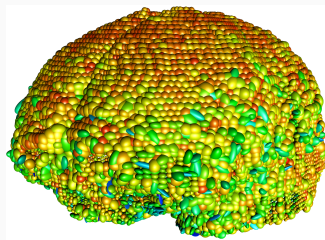
[Gesch et al., 2009]

directional data, $\mathcal{M} = \mathbb{S}^2$

Manifold-Valued Images

New data acquisition modalities lead to non-Euclidean range

- Interferometric synthetic aperture radar (InSAR)
- Surface normals, GPS data, wind, flow,...
- Diffusion tensors in magnetic resonance imaging (DT-MRI), covariance matrices
- Electron backscattered diffraction (EBSD)



diffusion tensors in human brain
from the Camino dataset

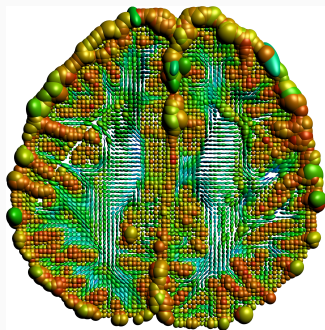
<http://cmic.cs.ucl.ac.uk/camino>

sym. pos. def. matrices, $\mathcal{M} = \text{SPD}(3)$

Manifold-Valued Images

New data acquisition modalities lead to non-Euclidean range

- Interferometric synthetic aperture radar (InSAR)
- Surface normals, GPS data, wind, flow,...
- Diffusion tensors in magnetic resonance imaging (DT-MRI), covariance matrices
- Electron backscattered diffraction (EBSD)



horizontal slice #28
from the Camino dataset

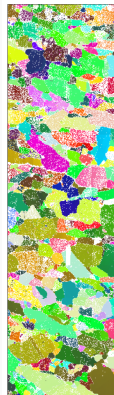
<http://cmic.cs.ucl.ac.uk/camino>

sym. pos. def. matrices, $\mathcal{M} = \text{SPD}(3)$

Manifold-Valued Images

New data acquisition modalities lead to non-Euclidean range

- Interferometric synthetic aperture radar (InSAR)
- Surface normals, GPS data, wind, flow,...
- Diffusion tensors in magnetic resonance imaging (DT-MRI), covariance matrices
- Electron backscattered diffraction (EBSD)



EBSD example from the MTEX toolbox
[Bachmann, Hielscher, since 2005]

Rotations (mod. symmetry),
 $\mathcal{M} = \text{SO}(3)(/\mathcal{S})$.

Manifold-Valued Images

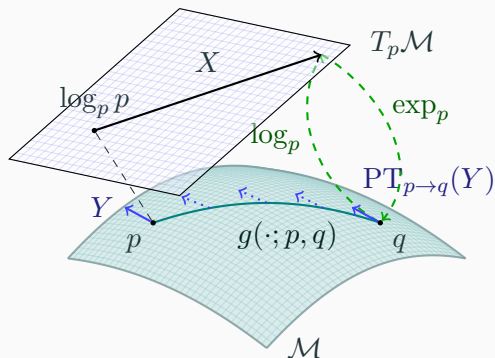
New data acquisition modalities lead to non-Euclidean range

- Interferometric synthetic aperture radar (InSAR)
- Surface normals, GPS data, wind, flow,...
- Diffusion tensors in magnetic resonance imaging (DT-MRI), covariance matrices
- Electron backscattered diffraction (EBSD)

Common properties

- Range of values is a Riemannian manifold
- Tasks from “classical” image processing, e.g.
 - denoising
 - inpainting
 - interpolation
 - labeling
 - deblurring

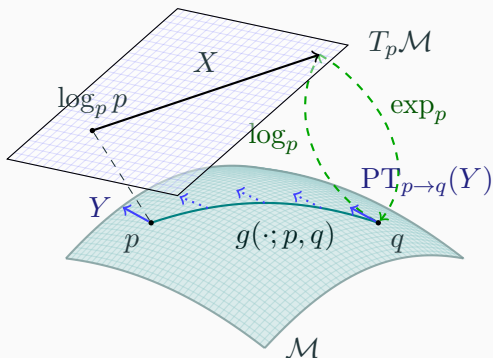
A d -dimensional Riemannian Manifold \mathcal{M}



A d -dimensional Riemannian manifold can be informally defined as a set \mathcal{M} covered with a 'suitable' collection of charts, that identify subsets of \mathcal{M} with open subsets of \mathbb{R}^d and a continuously varying inner product on the tangential spaces.

[Absil, Mahony, Sepulchre, 2008]

A d -dimensional Riemannian Manifold \mathcal{M}



Geodesic $g(\cdot; p, q)$ shortest path (on \mathcal{M}) between $p, q \in \mathcal{M}$

Tangent space $T_p \mathcal{M}$ at p , with inner product $(\cdot, \cdot)_p$

Logarithmic map $\log_p q = \dot{g}(0; p, q)$ “speed towards q ”

Exponential map $\exp_p X = g(1)$, where $g(0) = p$, $\dot{g}(0) = X$

Parallel transport $PT_{p \rightarrow q}(Y)$ of $Y \in T_p \mathcal{M}$ along $g(\cdot; p, q)$

Variational Methods on Manifolds

Variational methods model a trade-off between staying **close to the data** and **minimizing a certain property**

$$\mathcal{E}(p) = D(p; f) + \alpha R(p), \quad p \in \mathcal{M}$$

- $\alpha > 0$ is a weight
- \mathcal{M} is a Riemannian manifold
- given (input) data $f \in \mathcal{M}$
- data or similarity term $D(p; f)$
- regularizer / prior $R(p)$

Optimization on Manifolds

Let \mathcal{M} and \mathcal{N} be Riemannian Manifolds and $\mathcal{E}: \mathcal{N} \rightarrow \mathbb{R}$.

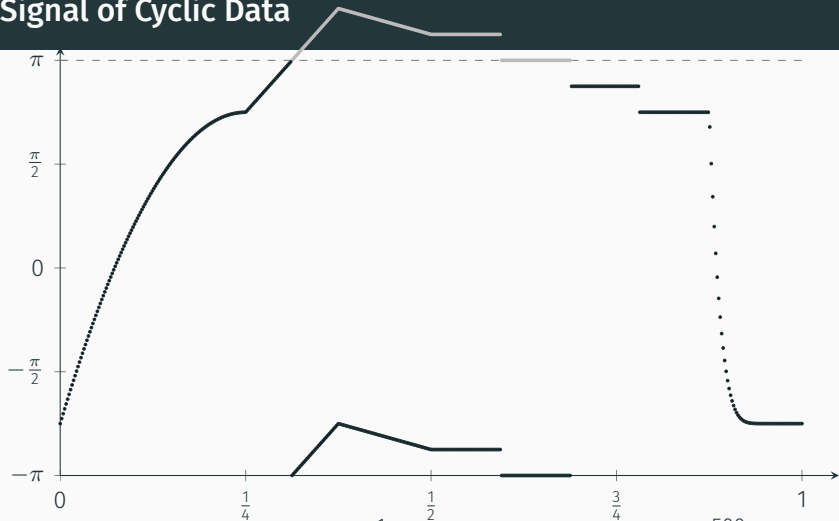
Consider the optimization problem

$$\arg \min_{p \in \mathcal{N}} \mathcal{E}(p)$$

where \mathcal{E} is

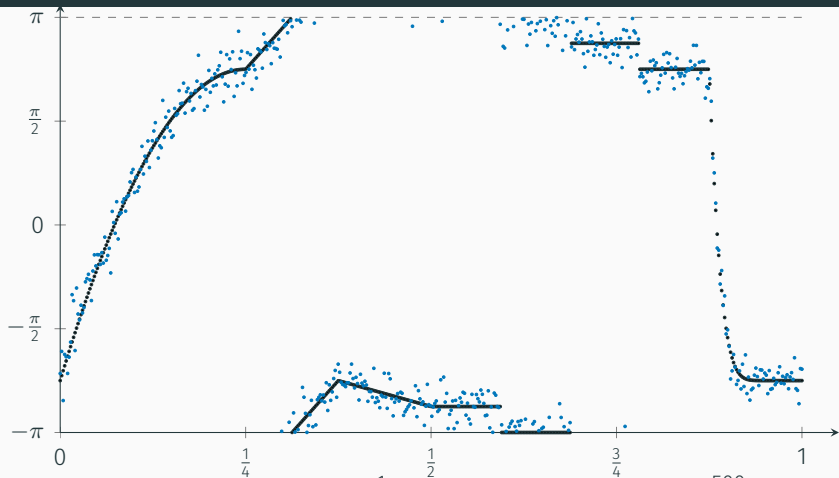
- (maybe) non-smooth
- (locally) convex
- high-dimensional,
 - a manifold valued signal, $\mathcal{N} = \mathcal{M}^d$
 - a manifold-valued image, $\mathcal{N} = \mathcal{M}^{d_1 \times d_2}$
- decomposable $\mathcal{E} = F + G$ in two (or even more) summands

A Signal of Cyclic Data



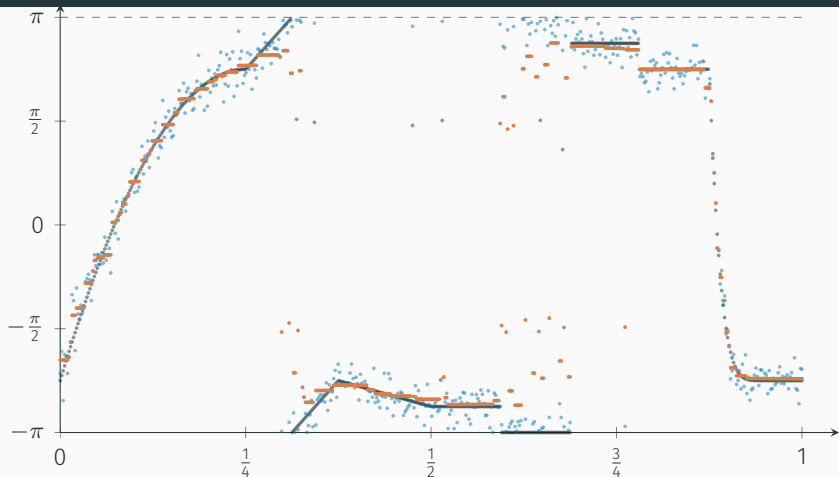
- A function $f: [0, 1] \rightarrow \mathbb{S}^1$ is sampled $\Rightarrow f_0 = (f_{0,i})_{i=1}^{500}$
- Data f stems from the gray plot via modulo
- Jumps $> \pi$ at $\frac{5}{16}$ and $\frac{11}{16}$ just from choice of representation

A Signal of Cyclic Data



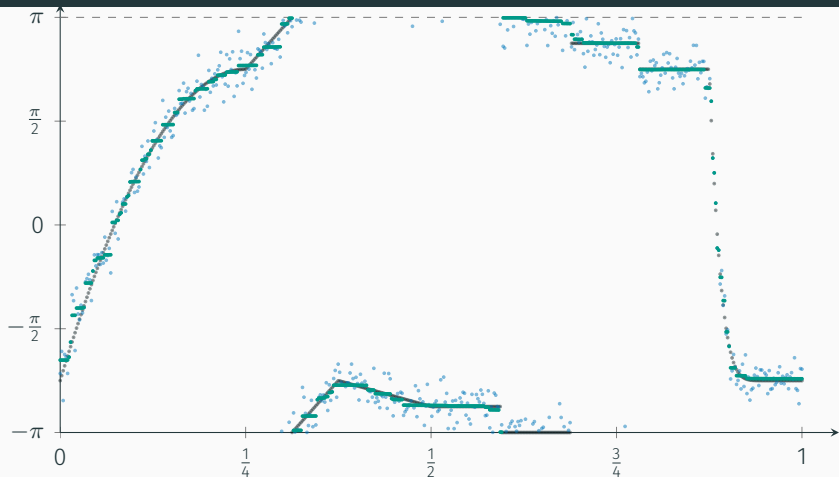
- A function $f: [0, 1] \rightarrow \mathbb{S}^1$ is sampled $\Rightarrow f_o = (f_{o,i})_{i=1}^{500}$
- Noise: wrapped Gaussian, $\sigma = 0.2$
- noisy $f_n = (f_o + \eta)_{2\pi}$

A Signal of Cyclic Data



- Comparison of f_o & f_n with f_R
- Denoised with CPPA and real-valued TV_1 , ($\alpha = \frac{3}{4}$, $\beta = 0$)
- Artefacts at the “jumps that are none” from representation

A Signal of Cyclic Data



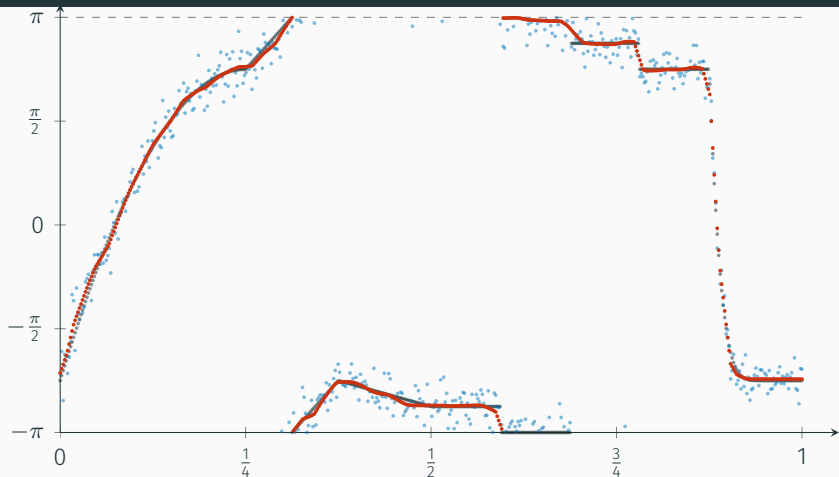
- Comparison of f_o & f_n with f_1
- Denoised with CPPA and TV_1 ($\alpha = \frac{3}{4}, \beta = 0$)
- but: stair casing

A Signal of Cyclic Data



- Comparison of f_o & f_n with f_2
- Denoised with CPPA and TV_2 ($\alpha = 0, \beta = \frac{3}{2}$)
- but: problems in constant areas

A Signal of Cyclic Data



- Comparison of f_o & f_n with f_3
- Denoised with CPPA and TV_1 & TV_2 ($\alpha = \frac{1}{4}$, $\beta = \frac{3}{4}$)
- combined: smallest mean squared error.

2. Second Order Differences

First and Second Order Differences

On \mathbb{R}^n

- line $\gamma(t) = x + t(y - x)$
- distance $\|x - y\|_2$
- first order model

[Rudin, Osher, Fatemi, 1992]

$$\sum_{i \in \mathcal{V}} \|f_i - u_i\|_2^2 + \alpha \sum_{i \in \mathcal{G} \setminus \{N\}} \|u_i - u_{i+1}\|_2$$

Riemannian manifold \mathcal{M}

- geodesic path $g(t; p, q)$
- geodesic distance $d: \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}$
- first order model

[Stekalovskiy, Cremers, 2011; Lellmann et al., 2013,
Weinmann, Demaret, Storath, 2014]

$$\sum_{i \in \mathcal{V}} d(f_i, u_i)^2 + \alpha \sum_{i \in \mathcal{G} \setminus \{N\}} d(u_i, u_{i+1})$$

First and Second Order Differences

On \mathbb{R}^n

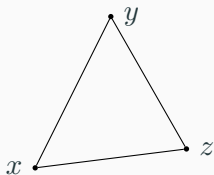
- line $\gamma(t) = x + t(y - x)$
- distance $\|x - y\|_2$
- first order model

[Rudin, Osher, Fatemi, 1992]

$$\sum_{i \in \mathcal{V}} \|f_i - u_i\|_2^2 + \alpha \sum_{i \in \mathcal{G} \setminus \{N\}} \|u_i - u_{i+1}\|_2$$

- second order difference

$$\|x - 2y + z\|_2$$



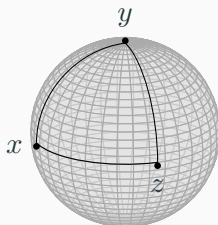
Riemannian manifold \mathcal{M}

- geodesic path $g(t; p, q)$
- geodesic distance $d: \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}$
- first order model

[Stekalovskiy, Cremers, 2011; Lellmann et al., 2013, Weinmann, Demaret, Storath, 2014]

$$\sum_{i \in \mathcal{V}} d(f_i, u_i)^2 + \alpha \sum_{i \in \mathcal{G} \setminus \{N\}} d(u_i, u_{i+1})$$

- How to model that on \mathcal{M} ?



$$\mathcal{M} = \mathbb{S}^2$$

First and Second Order Differences

On \mathbb{R}^n

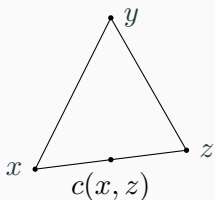
- line $\gamma(t) = x + t(y - x)$
- distance $\|x - y\|_2$
- first order model

[Rudin, Osher, Fatemi, 1992]

$$\sum_{i \in \mathcal{V}} \|f_i - u_i\|_2^2 + \alpha \sum_{i \in \mathcal{G} \setminus \{N\}} \|u_i - u_{i+1}\|_2$$

- second order difference

$$2\|\frac{1}{2}(x + z) - y\|_2$$



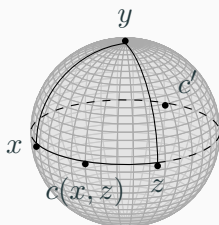
Riemannian manifold \mathcal{M}

- geodesic path $g(t; p, q)$
- geodesic distance $d: \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}$
- first order model

[Strekalovskiy, Cremers, 2011; Lellmann et al., 2013, Weinmann, Demaret, Storath, 2014]

$$\sum_{i \in \mathcal{V}} d(f_i, u_i)^2 + \alpha \sum_{i \in \mathcal{G} \setminus \{N\}} d(u_i, u_{i+1})$$

- **idea:** mid point formulation



$\mathcal{M} = \mathbb{S}^2$

First and Second Order Differences

On \mathbb{R}^n

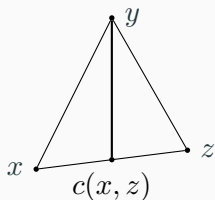
- line $\gamma(t) = x + t(y - x)$
- distance $\|x - y\|_2$
- first order model

[Rudin, Osher, Fatemi, 1992]

$$\sum_{i \in \mathcal{V}} \|f_i - u_i\|_2^2 + \alpha \sum_{i \in \mathcal{G} \setminus \{N\}} \|u_i - u_{i+1}\|_2$$

- second order difference

$$2\|c(x, z) - y\|_2$$



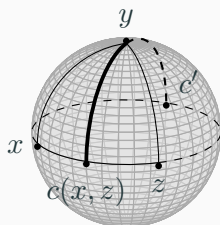
Riemannian manifold \mathcal{M}

- geodesic path $g(t; p, q)$
- geodesic distance $d: \mathcal{M} \times \mathcal{M} \rightarrow \mathbb{R}$
- first order model

[Strekalovskiy, Cremers, 2011; Lellmann et al., 2013, Weinmann, Demaret, Storath, 2014]

$$\sum_{i \in \mathcal{V}} d(f_i, u_i)^2 + \alpha \sum_{i \in \mathcal{G} \setminus \{N\}} d(u_i, u_{i+1})$$

- **idea:** mid point formulation



$$\mathcal{M} = \mathbb{S}^2$$

Absolute Second Order Difference

We denote the set of mid points between $x, z \in \mathcal{M}$ as

$$\mathcal{C}_{x,z} := \{c \in \mathcal{M} : c = g(\frac{1}{2}; x, z) \text{ for any geodesic } g(\cdot; x, z) : [0, 1] \rightarrow \mathcal{M}\}$$

and define the **Absolute Second Order Difference**

[Bergmann, Laus, et al., 2014; Bačák et al., 2016]

$$d_2(x, y, z) := \min_{c \in \mathcal{C}_{x,z}} d(c, y), \quad x, y, z \in \mathcal{M}.$$

For Optimization we need the differential and gradient of d_2 with respect to all its three arguments. For example for the first argument we have a **chain rule** of the distance and $g(\frac{1}{2}; \cdot, z)$

Differential and Gradient

The **differential** $D_p f = Df: T\mathcal{M} \rightarrow \mathbb{R}$ of a real-valued function $f: \mathcal{M} \rightarrow \mathbb{R}$ is the **push-forward** of f .

For a composition $F(p) = (g \circ h)(p) = g(h(p))$ of two functions $g, h: \mathcal{M} \rightarrow \mathcal{M}$ the **chain rule** reads

$$D_p F[X] = D_{h(p)} g [D_p h[X]],$$

where $D_p h[X] \in T_{h(p)}\mathcal{M}$ and $D_p F[X] \in T_{F(p)}\mathcal{M}$.

The **gradient** $\nabla f: \mathcal{M} \rightarrow T\mathcal{M}$ is the tangent vector fulfilling

$$(\nabla_{\mathcal{M}} f(p), Y)_p = Df(p)[Y] \text{ for all } Y \in T_p\mathcal{M},$$

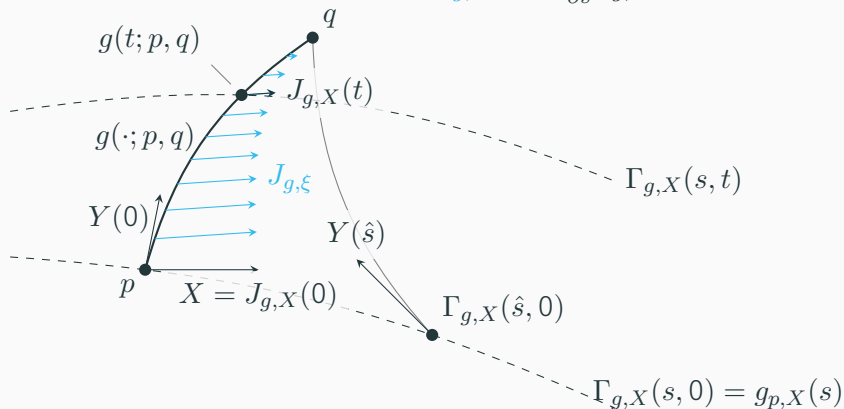
i.e. $\nabla f(p) \in T_p\mathcal{M}$ is a tangent vector at p .

The Differential of a Geodesic w.r.t. its Start Point

The **geodesic variation**

$$\Gamma_{g,X}(s,t) := \exp_{g_p,X(s)}(tY(s)), \quad s \in (-\varepsilon, \varepsilon), t \in [0,1], \varepsilon > 0.$$

is used to define the **Jacobi field** $J_{g,X}(t) = \frac{\partial}{\partial s} \Gamma_{g,X}(s,t)|_{s=0}$.



Then the differential reads $D_p g(t; \cdot, q)[X] = J_{g,X}(t)$.

Implementing Jacobi Fields on Symmetric Spaces

A manifold is **symmetric** if for every geodesic g and every $p \in \mathcal{M}$ the mapping $g(t) \mapsto g(-t)$ is an isometry at least locally around $p = g(0)$.

Then the system of ODEs characterizing the Jacobi field

$$\frac{D^2}{dt^2} J_{g,X} + R(J_{g,X}, \dot{g})\dot{g} = 0, \quad J_{g,X}(0) = X, \quad J_{g,X}(1) = 0$$

- has constant coefficients
- one can diagonalize the curvature tensor R ,
- let κ_ℓ denote its eigenvalues
- let $\{X_1, \dots, X_d\} \subseteq T_p\mathcal{M}$ be an ONB to these eigenvalues with $X_1 = \log_p q$.
- parallel transport $\Xi_j(t) = \text{PT}_{p \rightarrow g(t;p,q)} X_j, j = 1, \dots, d$

Implementing Jacobi Fields on Symmetric Spaces II

Decompose $X = \sum_{i=1}^d \eta_i X_i$. Then

$$D_p g(t; p, q)[X] = J_{g,X}(t) = \sum_{\ell=1}^d \eta_\ell J_{g,X_\ell}(t),$$

with

$$J_{g,X_\ell}(t) = \begin{cases} \frac{\sinh(d_g(1-t)\sqrt{-\kappa_\ell})}{\sinh(d_g\sqrt{-\kappa_\ell})} \Xi_\ell(t) & \text{if } \kappa_\ell < 0, \\ \frac{\sin(d_g(1-t)\sqrt{\kappa_\ell})}{\sin(\sqrt{\kappa_\ell}d_g)} \Xi_\ell(t) & \text{if } \kappa_\ell > 0, \\ (1-t)\Xi_\ell(t) & \text{if } \kappa_\ell = 0, \end{cases}$$

where $d_g = d(p, q)$ is the length of the geodesic.

Implementing the Gradient Using Adjoint Jacobi Fields.

The **adjoint Jacobi fields**

$$J_{g,\cdot}^*(t): T_{g(t;p,q)}\mathcal{M} \rightarrow T_p\mathcal{M}$$

are characterized by

$$(J_{g,X}(t), Y)_{g(t)} = (X, J_{g,Y}^*(t))_p, \text{ for all } X \in T_p\mathcal{M}, Y \in T_{g(t;p,q)}\mathcal{M}.$$

- computed using the same ODEs
- ⇒ calculate gradient of $f(x) = d(y, c), c = g(\frac{1}{2}; x, z)$, as

$$\nabla f(x) = J_{g,Y}^*(\frac{1}{2}), \quad g = g(\cdot; x, z), \quad Y = -\frac{\log_c y}{\|\log_c y\|_c}$$

- the gradient of iterated evaluations of geodesics
- ⇒ (sum of) composition of (adjoint) Jacobi fields

3. Second Order Total Variation

A Second Order TV-type Model on Manifolds

For \mathcal{M} -valued signals f we define

$$\mathcal{E}(u) := \sum_{i \in \mathcal{V}} d(f_i, u_i)^2 + \alpha \sum_{i \in \mathcal{G} \setminus \{N\}} d(u_i, u_{i+1}) + \beta \sum_{i \in \mathcal{G} \setminus \{1, N\}} d_2(u_{i-1}, u_i, u_{i+1})$$

A Second Order TV-type Model on Manifolds

For \mathcal{M} -valued signals f we define

$$\mathcal{E}(u) := \sum_{i \in \mathcal{V}} d(f_i, u_i)^2 + \alpha \sum_{i \in \mathcal{G} \setminus \{N\}} d(u_i, u_{i+1}) + \beta \sum_{i \in \mathcal{G} \setminus \{1, N\}} d_2(u_{i-1}, u_i, u_{i+1})$$

For images additionally: use

$$\|w - x + y - z\|_2 = 2 \left\| \frac{1}{2}(w + y) - \frac{1}{2}(x + z) \right\|_2 \text{ for}$$

Absolute Second Order Mixed Difference

$$d_{1,1}(w, x, y, z) := \min_{c \in \mathcal{C}_{w,y}, \tilde{c} \in \mathcal{C}_{x,z}} d(c, \tilde{c}), \quad w, x, y, z \in \mathcal{M}.$$

Proximal Map

For $\varphi: \mathcal{M}^n \rightarrow (-\infty, +\infty]$ and $\lambda > 0$ we define the **Proximal Map** as

[Moreau, 1965; Rockafellar, 1976; Ferreira, Oliveira, 2002]

$$\text{prox}_{\lambda\varphi}(p) := \arg \min_{u \in \mathcal{M}^n} \frac{1}{2} \sum_{i=1}^n d(u_i, p_i)^2 + \lambda\varphi(u).$$

- ! For a Minimizer u^* of φ we have $\text{prox}_{\lambda\varphi}(u^*) = u^*$.
- For $\varphi: \mathbb{R}^n \rightarrow \mathbb{R}$ proper, convex, lower semicontinuous:
 - the proximal map is unique.
 - PPA $x_k = \text{prox}_{\lambda\varphi}(x_{k-1})$ converges to $\arg \min \varphi$
- For $\varphi = \mathcal{E}$ not that useful

The Cyclic Proximal Point Algorithm

For $\varphi = \sum_{l=1}^c \varphi_l$ the

Cyclic Proximal Point-Algorithmus (CPPA) reads [Bertsekas, 2011; Bačák, 2014]

$$p^{(k+\frac{l+1}{c})} = \text{prox}_{\lambda_k \varphi_l}(p^{(k+\frac{l}{c})}), \quad l = 0, \dots, c-1, \quad k = 0, 1, \dots$$

On a Hadamard manifold \mathcal{M} :

convergence to a minimizer of φ if

- all φ_l proper, convex, lower semicontinuous
- $\{\lambda_k\}_{k \in \mathbb{N}} \in \ell_2(\mathbb{N}) \setminus \ell_1(\mathbb{N})$.

Ansatz.

- efficient proximal maps for every summand of $\mathcal{E}(u)$.
- speedup by parallelization

Proximal Maps for Distance and TV summands

Let $g(\cdot; p, q): [0, 1] \rightarrow \mathcal{M}$ be a geodesic between $p, q \in \mathcal{M}$.

Theorem (Distance term)

[Ferreira, Oliveira, 2002]

For $\varphi(p) = d^2(p, f)$ with fixed $f \in \mathcal{M}$ we have

$$\text{prox}_{\lambda\varphi}(p) = g\left(\frac{\lambda d(p, f)}{1 + \lambda d(p, f)}; p, f\right)$$

Theorem (First Order Difference Term)

[Weinmann, Demaret, Storath, 2014]

For $\varphi(p, q) = d(p, q)$ we have

$$\text{prox}_{\lambda\varphi}(p, q) = (g(t; p, q), g(1 - t; p, q))$$

with

$$t = \begin{cases} \frac{\lambda}{d(p, q)} & \text{if } \lambda < \frac{1}{2}d(p, q) \\ \frac{1}{2} & \text{else.} \end{cases}$$

Proximal Maps for the TV_2 Summands

To compute

$$\text{prox}_{\lambda d_2}(p) = \arg \min_{u \in \mathcal{M}^3} \left\{ \frac{1}{2} \sum_{i=1}^3 d(u_i, p_i)^2 + \lambda d_2(u_1, u_2, u_3) \right\}$$

We have

- a closed form solution for $\mathcal{M} = \mathbb{S}^1$ [Bergmann, Laus, et al., 2014]
- use a sub gradient descent (as inner problem) with

$$\nabla_{\mathcal{M}^3} d_2 = (\nabla_{\mathcal{M}} d_2(\cdot, p_2, p_3), \nabla_{\mathcal{M}} d_2(p_1, \cdot, p_3), \nabla_{\mathcal{M}} d_2(p_1, p_2, \cdot))^T.$$

where

- $\nabla_{\mathcal{M}} d_2(p_1, \cdot, p_3)(y) = -\frac{\log_y c(p_1, p_3)}{\|\log_{p_2} c(p_1, p_3)\|_{p_2}} \in T_y \mathcal{M}$
- $\nabla_{\mathcal{M}} d_2(\cdot, p_2, p_3)$ and analogously $\nabla_{\mathcal{M}} d_2(p_1, p_2, \cdot)$
using (adjoint) Jacobi fields and a chain rule

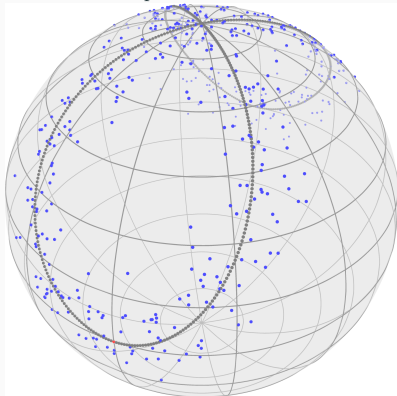
[Bačák et al., 2016]

Bernoulli's Lemniscate on the sphere \mathbb{S}^2

$$\gamma(t) := \frac{a\sqrt{2}}{\sin^2(t) + 1} (\cos(t), \cos(t) \sin(t), 1)^T, \quad t \in [0, 2\pi], a = \frac{\pi}{2\sqrt{2}}.$$

Generate a **sphere-valued signal** by

$$\gamma_S(t) = \exp_p(\gamma(t)), p = (0, 0, 1)^T$$



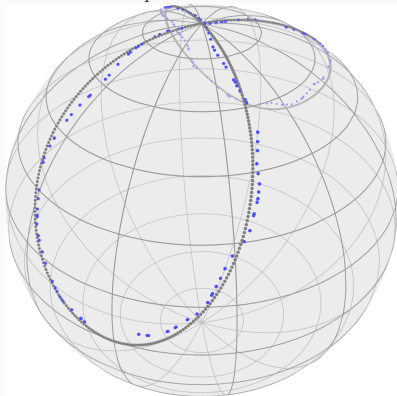
noisy lemniscate of Bernoulli on \mathbb{S}^2 , Gaussian noise, $\sigma = \frac{\pi}{30}$, on $T_p\mathbb{S}^2$.

Bernoulli's Lemniscate on the sphere \mathbb{S}^2

$$\gamma(t) := \frac{a\sqrt{2}}{\sin^2(t) + 1} (\cos(t), \cos(t) \sin(t), 1)^T, \quad t \in [0, 2\pi], a = \frac{\pi}{2\sqrt{2}}.$$

Generate a **sphere-valued signal** by

$$\gamma_S(t) = \exp_p(\gamma(t)), p = (0, 0, 1)^T$$



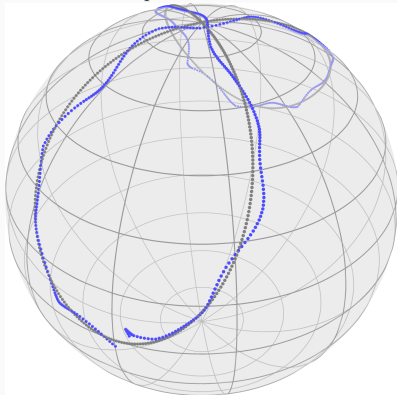
reconstruction with TV_1 , $\alpha = 0.21$, $MAE = 4.08 \times 10^{-2}$.

Bernoulli's Lemniscate on the sphere \mathbb{S}^2

$$\gamma(t) := \frac{a\sqrt{2}}{\sin^2(t) + 1} (\cos(t), \cos(t) \sin(t), 1)^T, \quad t \in [0, 2\pi], a = \frac{\pi}{2\sqrt{2}}.$$

Generate a **sphere-valued signal** by

$$\gamma_S(t) = \exp_p(\gamma(t)), p = (0, 0, 1)^T$$



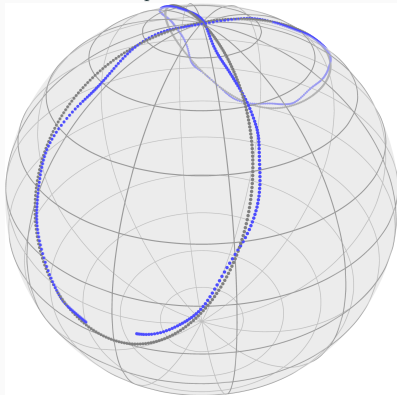
reconstruction with TV_2 , $\alpha = 0$, $\beta = 10$, $MAE = 3.66 \times 10^{-2}$.

Bernoulli's Lemniscate on the sphere \mathbb{S}^2

$$\gamma(t) := \frac{a\sqrt{2}}{\sin^2(t) + 1} (\cos(t), \cos(t) \sin(t), 1)^T, \quad t \in [0, 2\pi], a = \frac{\pi}{2\sqrt{2}}.$$

Generate a **sphere-valued signal** by

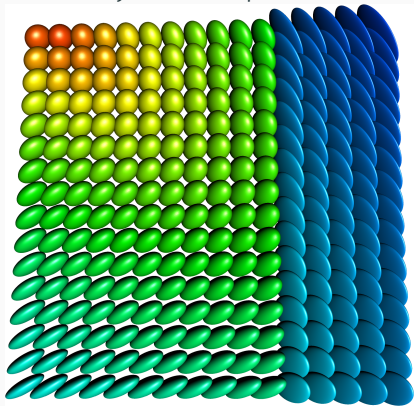
$$\gamma_S(t) = \exp_p(\gamma(t)), p = (0, 0, 1)^T$$



reconstruction with TV_1 & TV_2 , $\alpha = 0.16$, $\beta = 12.4$, $MAE = 3.27 \times 10^{-2}$.

Inpainting of $\mathcal{P}(3)$ -valued Images

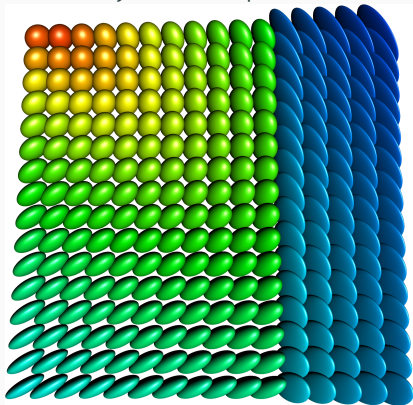
Draw symmetric positive definite 3×3 matrices as ellipsoids



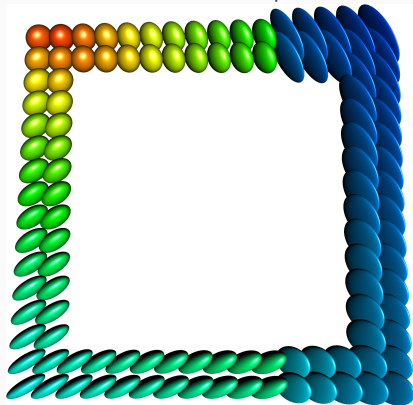
original data

Inpainting of $\mathcal{P}(3)$ -valued Images

Draw symmetric positive definite 3×3 matrices as ellipsoids



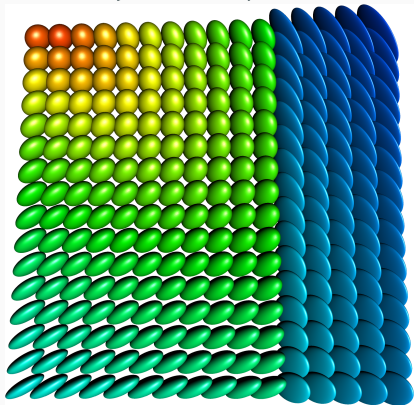
original data



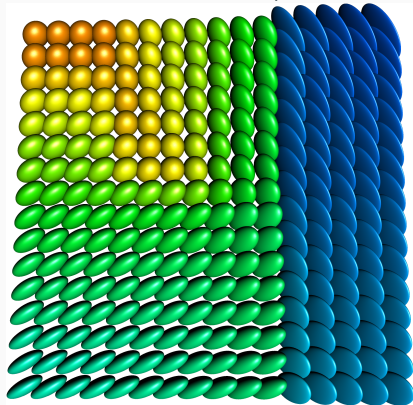
lost (a lot of) data

Inpainting of $\mathcal{P}(3)$ -valued Images

Draw symmetric positive definite 3×3 matrices as ellipsoids



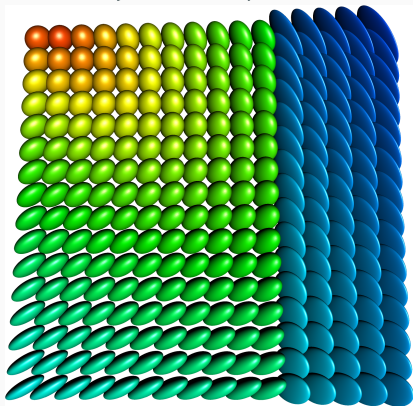
original data



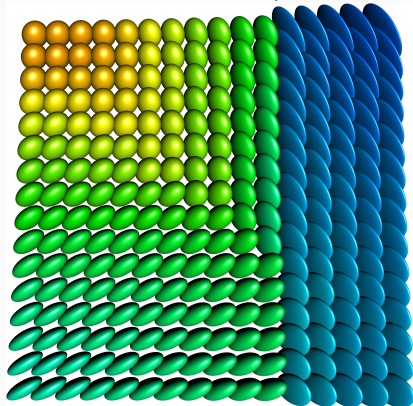
inpainted with $\alpha = \beta = 0.05$,
MAE = 0.0929

Inpainting of $\mathcal{P}(3)$ -valued Images

Draw symmetric positive definite 3×3 matrices as ellipsoids



original data



inpainted with $\alpha = 0.1$,
MAE = 0.0712

4. Acceleration of Bézier Curves

Data Fitting on Manifolds

Given data points d_0, \dots, d_n on a Riemannian manifold \mathcal{M} and time points $t_i \in I$, find a “nice” curve $\gamma: I \rightarrow \mathcal{M}$, $\gamma \in \Gamma$, such that $\gamma(t_i) = d_i$ (interpolation) or $\gamma(t_i) \approx d_i$ (approximation).

Data Fitting on Manifolds

Given data points d_0, \dots, d_n on a Riemannian manifold \mathcal{M} and time points $t_i \in I$, find a “nice” curve $\gamma: I \rightarrow \mathcal{M}$, $\gamma \in \Gamma$, such that $\gamma(t_i) = d_i$ (interpolation) or $\gamma(t_i) \approx d_i$ (approximation).

- Γ set of geodesics & approximation: geodesic regression
[Rentmeesters, 2011; Fletcher, 2013; Boumal, Absil, 2011]
- Γ Sobolev space of curves: Infinite-dimensional problem
[Samir et al., 2012]
- Γ composite Bézier curves; LSs in tangent spaces
[Arnould et al., 2015; Gousenbourger, Massart, Absil, 2018]
- Discretized curve, $\Gamma = \mathcal{M}^N$,
[Boumal, Absil, 2011]

Data Fitting on Manifolds

Given data points d_0, \dots, d_n on a Riemannian manifold \mathcal{M} and time points $t_i \in I$, find a “nice” curve $\gamma: I \rightarrow \mathcal{M}$, $\gamma \in \Gamma$, such that $\gamma(t_i) = d_i$ (interpolation) or $\gamma(t_i) \approx d_i$ (approximation).

- Γ set of geodesics & approximation: geodesic regression
[Rentmeesters, 2011; Fletcher, 2013; Boumal, Absil, 2011]
- Γ Sobolev space of curves: Infinite-dimensional problem
[Samir et al., 2012]
- Γ composite Bézier curves; LSs in tangent spaces
[Arnould et al., 2015; Gousenbourger, Massart, Absil, 2018]
- Discretized curve, $\Gamma = \mathcal{M}^N$,
[Boumal, Absil, 2011]

This talk.

“nice” means minimal (discretized) acceleration (“as straight as possible”) for Γ the set of composite Bézier curves.

Closed form solution for $\mathcal{M} = \mathbb{R}^d$: Natural (cubic) splines.

(Euclidean) Bézier Curves

Definition

[Bézier, 1962]

A **Bézier curve** β_K of degree $K \in \mathbb{N}_0$ is a function $\beta_K: [0, 1] \rightarrow \mathbb{R}^d$ parametrized by **control points** $b_0, \dots, b_K \in \mathbb{R}^d$ and defined by

$$\beta_K(t; b_0, \dots, b_K) := \sum_{j=0}^K b_j B_{j,K}(t),$$

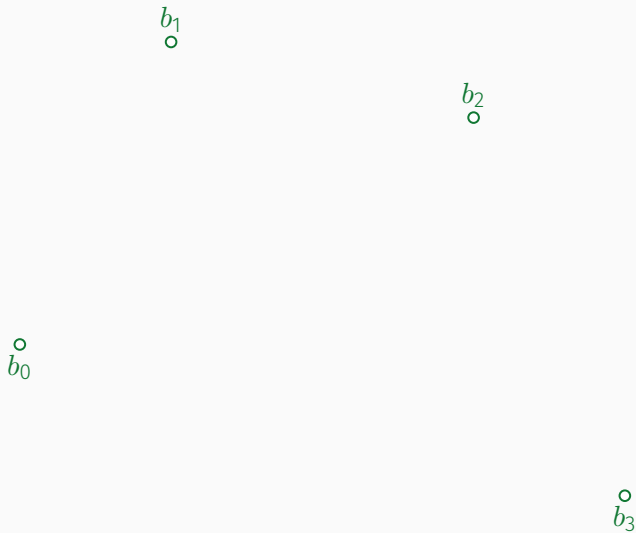
[Bernstein, 1912]

where $B_{j,K} = \binom{K}{j} t^j (1-t)^{K-j}$ are the **Bernstein polynomials** of degree K .

Evaluation via **Casteljau's algorithm**.

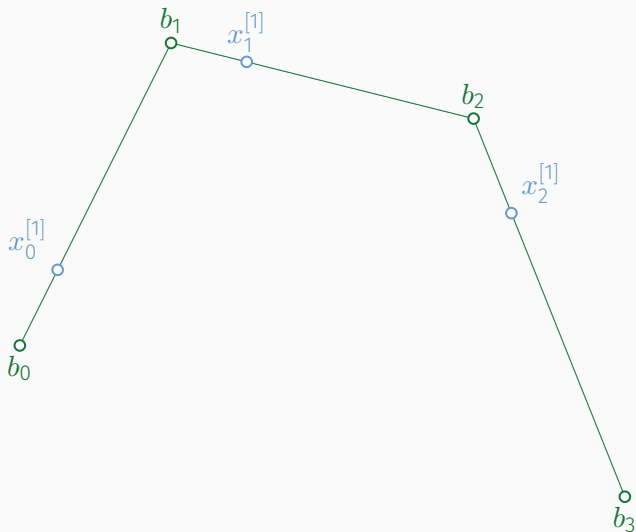
[de Casteljau, 1959]

Illustration of de Casteljau's Algorithm



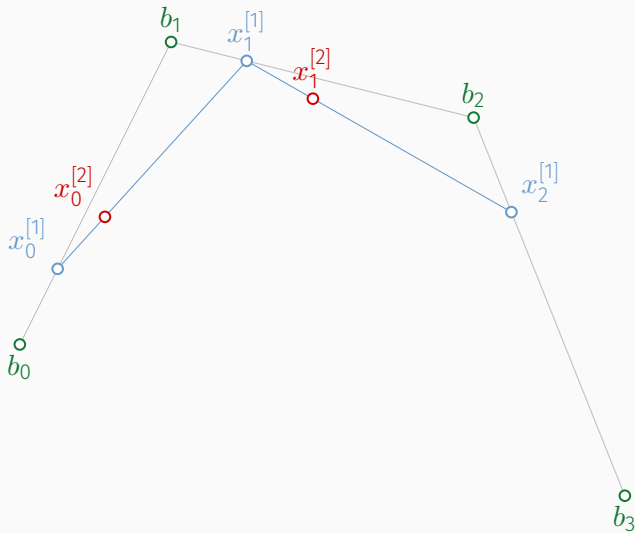
The set of control points b_0, b_1, b_2, b_3 .

Illustration of de Casteljau's Algorithm



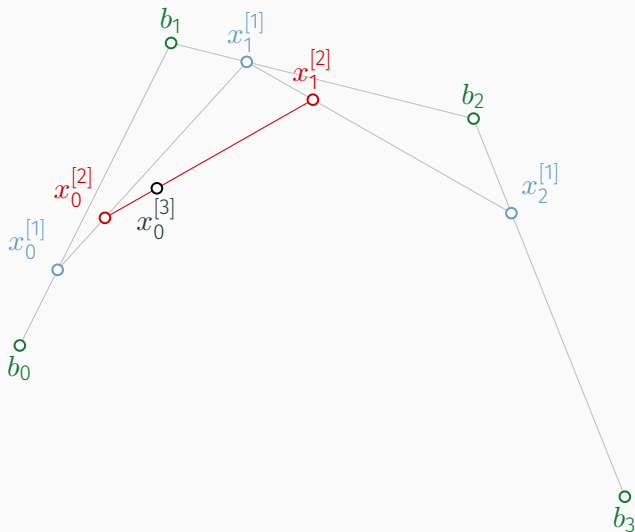
Evaluate line segments at $t = \frac{1}{4}$, obtain $x_0^{[1]}, x_1^{[1]}, x_2^{[1]}$.

Illustration of de Casteljau's Algorithm



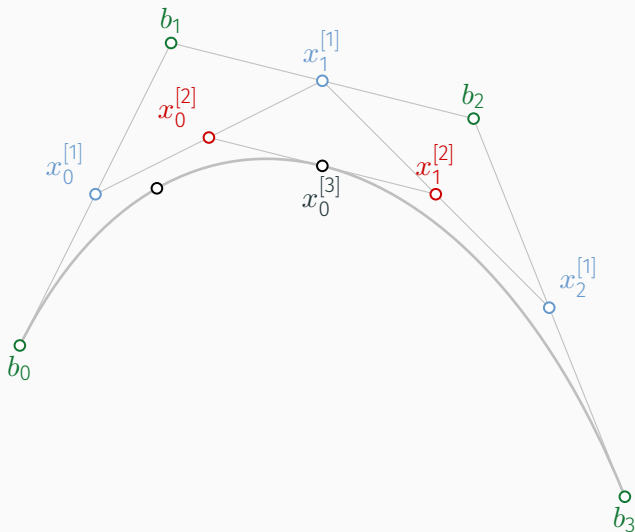
Repeat evaluation for new line segments to obtain $x_0^{[2]}, x_1^{[2]}$.

Illustration of de Casteljau's Algorithm



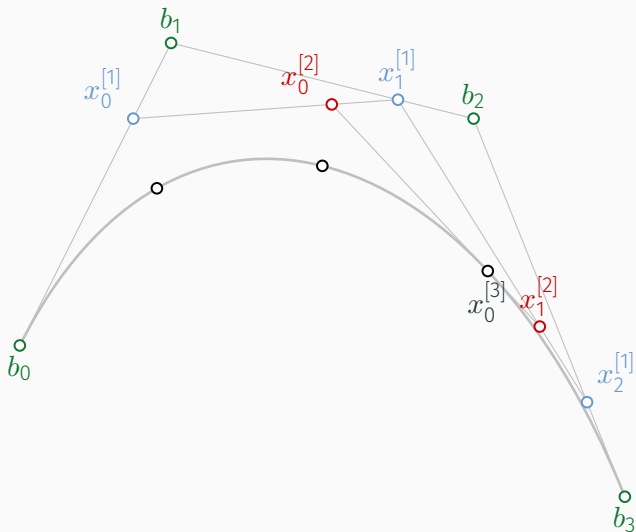
Repeat for the **last segment** to obtain $\beta_3(\frac{1}{4}; b_0, b_1, b_2, b_3) = x_0^{[3]}$.

Illustration of de Casteljau's Algorithm



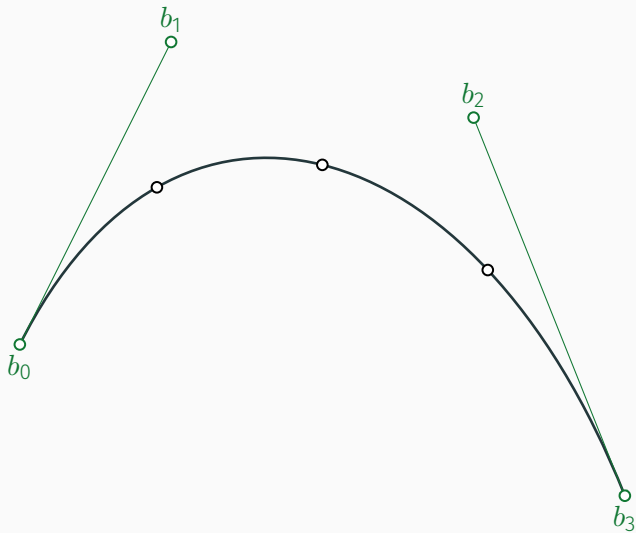
Same procedure for evaluation of $\beta_3(\frac{1}{2}; b_0, b_1, b_2, b_3)$.

Illustration of de Casteljau's Algorithm



Same procedure for evaluation of $\beta_3(\frac{3}{4}; b_0, b_1, b_2, b_3)$.

Illustration of de Casteljau's Algorithm



Complete curve $\beta_3(t; b_0, b_1, b_2, b_3)$.

Composite Bézier Curves

Definition

A **composite Bézier curve** $B: [0, n] \rightarrow \mathbb{R}^d$ is defined as

$$B(t) := \begin{cases} \beta_K(t; b_0^0, \dots, b_K^0) & \text{if } t \in [0, 1], \\ \beta_K(t - i; b_0^i, \dots, b_K^i), & \text{if } t \in (i, i + 1], \quad i = 1, \dots, n - 1. \end{cases}$$

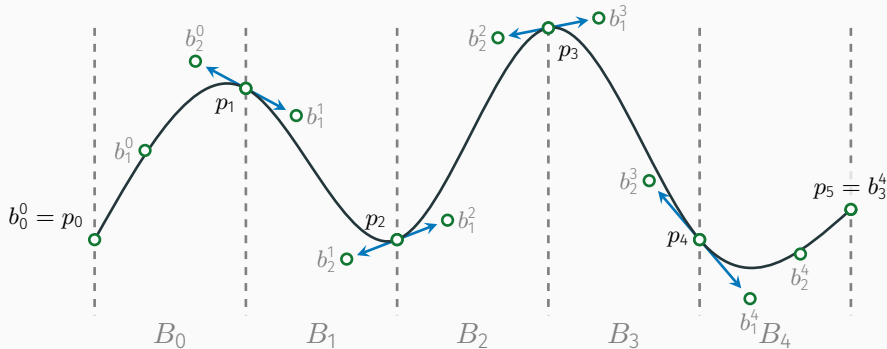
Composite Bézier Curves

Definition

A **composite Bézier curve** $B: [0, n] \rightarrow \mathbb{R}^d$ is defined as

$$B(t) := \begin{cases} \beta_K(t; b_0^0, \dots, b_K^0) & \text{if } t \in [0, 1], \\ \beta_K(t - i; b_0^i, \dots, b_K^i), & \text{if } t \in (i, i + 1], \quad i = 1, \dots, n - 1. \end{cases}$$

Denote i th segment by $B_i(t) = \beta_K(t; b_0^i, \dots, b_K^i)$ and $p_i = b_0^i$.



Composite Bézier Curves

Definition

A **composite Bezier curve** $B: [0, n] \rightarrow \mathbb{R}^d$ is defined as

$$B(t) := \begin{cases} \beta_K(t; b_0^0, \dots, b_K^0) & \text{if } t \in [0, 1], \\ \beta_K(t - i; b_0^i, \dots, b_K^i), & \text{if } t \in (i, i + 1], \quad i = 1, \dots, n - 1. \end{cases}$$

Denote i th segment by $B_i(t) = \beta_K(t; b_0^i, \dots, b_K^i)$ and $p_i = b_0^i$.

- **continuous** iff $B_{i-1}(1) = B_i(0)$, $i = 1, \dots, n - 1$
 $\Rightarrow b_K^{i-1} = b_0^i = p_i$, $i = 1, \dots, n - 1$
- **continuously differentiable** iff $p_i = \frac{1}{2}(b_{K-1}^{i-1} + b_1^i)$

Bézier Curves on a Manifold

Definition.

[Park, Ravani, 1995; Popiel, Noakes, 2007]

Let \mathcal{M} be a Riemannian manifold and $b_0, \dots, b_K \in \mathcal{M}$, $K \in \mathbb{N}$.

The (generalized) Bézier curve of degree k , $k \leq K$, is defined as

$$\beta_k(t; b_0, \dots, b_k) = g(t; \beta_{k-1}(t; b_0, \dots, b_{k-1}), \beta_{k-1}(t; b_1, \dots, b_k)),$$

if $k > 0$, and

$$\beta_0(t; b_0) = b_0.$$

Bézier Curves on a Manifold

Definition.

[Park, Ravani, 1995; Popiel, Noakes, 2007]

Let \mathcal{M} be a Riemannian manifold and $b_0, \dots, b_K \in \mathcal{M}$, $K \in \mathbb{N}$.

The (generalized) Bézier curve of degree k , $k \leq K$, is defined as

$$\beta_k(t; b_0, \dots, b_k) = g(t; \beta_{k-1}(t; b_0, \dots, b_{k-1}), \beta_{k-1}(t; b_1, \dots, b_k)),$$

if $k > 0$, and

$$\beta_0(t; b_0) = b_0.$$

- Bézier curves $\beta_1(t; b_0, b_1) = g(t; b_0, b_1)$ are geodesics.
- composite Bézier curves $B: [0, n] \rightarrow \mathcal{M}$ completely analogue (using geodesics for line segments)

Bézier Curves on a Manifold

Definition.

Let \mathcal{M} be a Riemannian manifold and $b_0, \dots, b_K \in \mathcal{M}$, $K \in \mathbb{N}$. [Park, Ravani, 1995; Popiel, Noakes, 2007]

The (generalized) Bézier curve of degree k , $k \leq K$, is defined as

$$\beta_k(t; b_0, \dots, b_k) = g(t; \beta_{k-1}(t; b_0, \dots, b_{k-1}), \beta_{k-1}(t; b_1, \dots, b_k)),$$

if $k > 0$, and

$$\beta_0(t; b_0) = b_0.$$

The Riemannian composite Bézier curve $B(t)$ is

- **continuous** iff $B_{i-1}(1) = B_i(0)$, $i = 1, \dots, n-1$
 $\Rightarrow b_{K-1}^{i-1} = b_0^i =: p_i$, $i = 1, \dots, n-1$
- **continuously differentiable** iff $p_i = g(\frac{1}{2}; b_{K-1}^{i-1}, b_0^i)$

Bézier Curves on a Manifold

Definition.

Let \mathcal{M} be a Riemannian manifold and $b_0, \dots, b_K \in \mathcal{M}$, $K \in \mathbb{N}$. [Park, Ravani, 1995; Popiel, Noakes, 2007]

The (generalized) Bézier curve of degree k , $k \leq K$, is defined as

$$\beta_k(t; b_0, \dots, b_k) = g(t; \beta_{k-1}(t; b_0, \dots, b_{k-1}), \beta_{k-1}(t; b_1, \dots, b_k)),$$

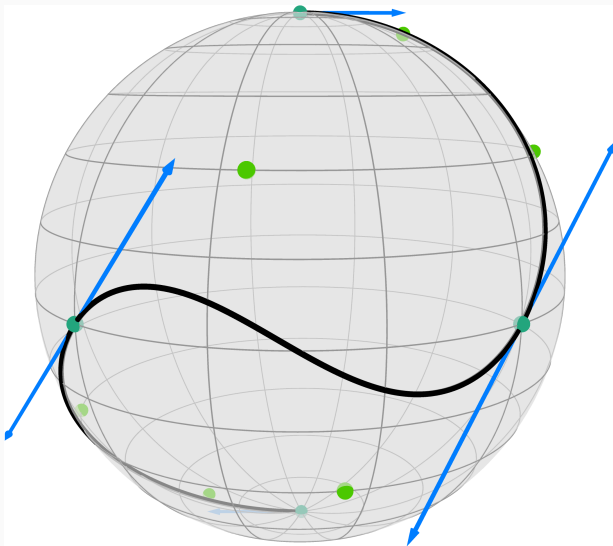
if $k > 0$, and

$$\beta_0(t; b_0) = b_0.$$

The Riemannian composite Bézier curve $B(t)$ is

- **continuous** iff $B_{i-1}(1) = B_i(0)$, $i = 1, \dots, n - 1$
 $\Rightarrow b_{K-1}^{i-1} = b_0^i =: p_i$, $i = 1, \dots, n - 1$
- **continuously differentiable** iff $b_{K-1}^{i-1} = g(2; b_1^i, p_i)$.

Illustration of a Composite Bézier Curve on the Sphere \mathbb{S}^2



The directions, e.g. $\log_{p_j} b_j^1$, are now tangent vectors.

A Variational Model for Data Fitting

Let $d_0, \dots, d_n \in \mathcal{M}$. A model for **data fitting** reads

$$\mathcal{E}(B) = \frac{\lambda}{2} \sum_{k=0}^n d_{\mathcal{M}}^2(B(k), d_k) + \int_0^n \left\| \frac{D^2 B(t)}{dt^2} \right\|_{B(t)}^2 dt, \quad \lambda > 0,$$

where $B \in \Gamma$ is from the set of continuously differentiable composite Bezier curve of degree K with n segments.

A Variational Model for Data Fitting

Let $d_0, \dots, d_n \in \mathcal{M}$. A model for data fitting reads

$$\mathcal{E}(B) = \frac{\lambda}{2} \sum_{k=0}^n d_{\mathcal{M}}^2(B(k), d_k) + \int_0^n \left\| \frac{D^2 B(t)}{dt^2} \right\|_{B(t)}^2 dt, \quad \lambda > 0,$$

where $B \in \Gamma$ is from the set of continuously differentiable composite Bezier curve of degree K with n segments.

- **Goal:** find minimizer $B^* \in \Gamma$
- finite dimensional optimization problem in the control points b_j^i , i.e. on \mathcal{M}^L with
 - $L = n(K - 1) + 2$
 - $\lambda \rightarrow \infty$ yields interpolation ($p_k = d_k$) $\Rightarrow L = n(K - 2) + 1$

A Variational Model for Data Fitting

Let $d_0, \dots, d_n \in \mathcal{M}$. A model for data fitting reads

$$\mathcal{E}(B) = \frac{\lambda}{2} \sum_{k=0}^n d_{\mathcal{M}}^2(B(k), d_k) + \int_0^n \left\| \frac{D^2 B(t)}{dt^2} \right\|_{B(t)}^2 dt, \quad \lambda > 0,$$

where $B \in \Gamma$ is from the set of continuously differentiable composite Bezier curve of degree K with n segments.

- **Goal:** find minimizer $B^* \in \Gamma$
- finite dimensional optimization problem in the control points b_j^i , i.e. on \mathcal{M}^L with
 - $L = n(K - 1) + 2$
 - $\lambda \rightarrow \infty$ yields interpolation ($p_k = d_k$) $\Rightarrow L = n(K - 2) + 1$
- **On $\mathcal{M} = \mathbb{R}^m$:** closed form solution, natural (cubic) splines

Discretizing the Data Fitting Model

We discretize the absolute second order covariant derivative

$$\int_0^n \left\| \frac{D^2 B(t)}{dt^2} \right\|_{\gamma(t)}^2 dt \approx \sum_{k=1}^{N-1} \frac{\Delta_s d_2^2[B(s_{i-1}), B(s_i), B(s_{i+1})]}{\Delta_s^4}.$$

on equidistant points s_0, \dots, s_N with step size $\Delta_s = s_1 - s_0$.

Evaluating $\mathcal{E}(B)$ consists of evaluation of geodesics and squared (Riemannian) distances

- $(N + 1)K$ geodesics to evaluate the Bézier segments
- N geodesics to evaluate the mid points c
- N squared distances to obtain the second order absolute finite differences squared

Gradient of the Discretized Data Fitting Model

For the gradient of the discretized data fitting model

$$\mathcal{E}(B) = \frac{\lambda}{2} \sum_{k=0}^n d_{\mathcal{M}}^2(B(k), d_k) + \sum_{k=1}^{N-1} \frac{\Delta_s d_2^2[B(s_{i-1}), B(s_i), B(s_{i+1})]}{\Delta_s^4}.$$

we

- identified first and last control points $p_i = b_{K-1}^{i-1} = b_0^i$
- plug in the constraint $b_{K-1}^{i-1} = g(2; b_1^i, p_i)$
 - ⇒ Introduces a further chain rule for the differential
 - ⇒ reduces the number of optimization variables.
- concatenation of adjoint Jacobi fields (evaluated at the points s_i) yields the gradient $\nabla_{\mathcal{N}} \mathcal{E}$.

Gradient Descent on a Manifold

Let $\mathcal{N} = \mathcal{M}^L$ be the product manifold of \mathcal{M} ,

Input.

- $\mathcal{E}: \mathcal{N} \rightarrow \mathbb{R}$,
- its gradient $\nabla_{\mathcal{N}}\mathcal{E}$,
- initial data $q^{(0)} = b \in \mathcal{N}$
- step sizes $s_k > 0, k \in \mathbb{N}$.

Output: $\hat{q} \in \mathcal{N}$

$k \leftarrow 0$

repeat

$$q^{(k+1)} \leftarrow \exp_{q^{(k)}}(-s_k \nabla_{\mathcal{N}}\mathcal{E}(q^{(k)}))$$

$$k \leftarrow k + 1$$

until a stopping criterion is reached

return $\hat{q} := q^{(k)}$

Armijo Step Size Rule

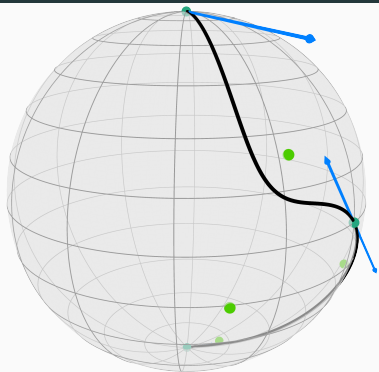
Let $q = q^{(k)}$ be an iterate from the gradient descent algorithm, $\beta, \sigma \in (0, 1), \alpha > 0$.

Let m be the smallest positive integer such that

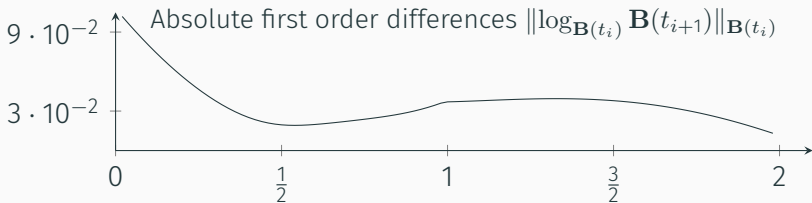
$$\mathcal{E}(q) - \mathcal{E}(\exp_q(-\beta^m \alpha \nabla_{\mathcal{N}} \mathcal{E}(q))) \geq \sigma \beta^m \alpha \|\nabla_{\mathcal{N}} \mathcal{E}(q)\|_q$$

holds. Set the step size $s_k := \beta^m \alpha$.

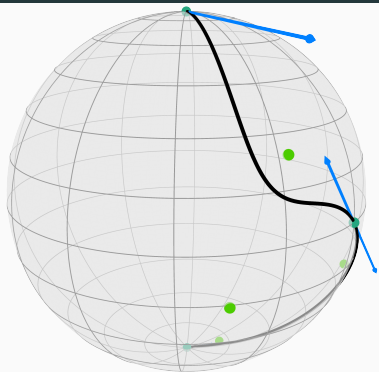
Minimizing with Known Minimizer



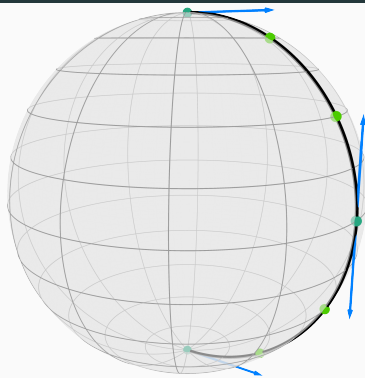
Original



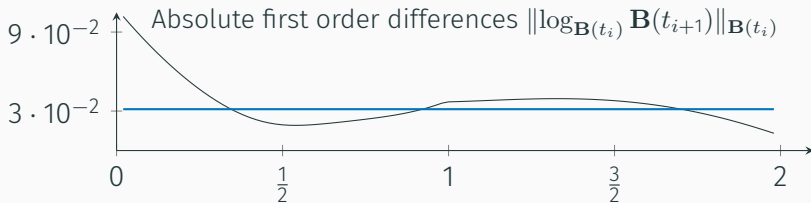
Minimizing with Known Minimizer



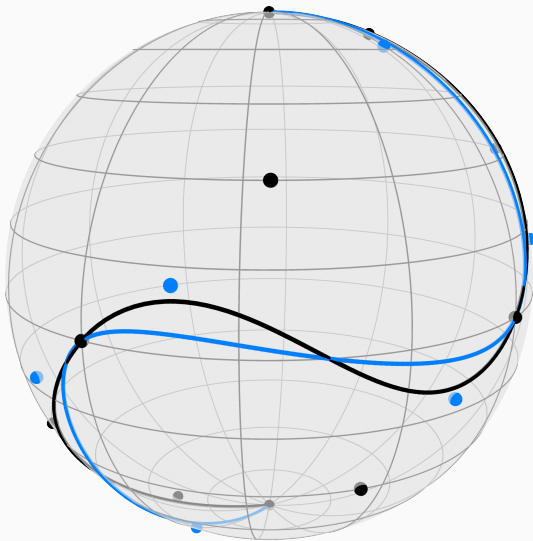
Original



Minimized



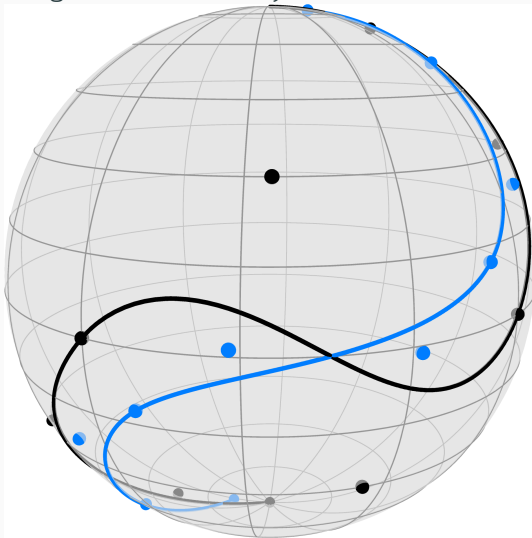
Interpolation by Bézier Curves with Minimal Acceleration.



A comp. Bézier curve (black) and its minimizer (blue).

Approximation by Bézier Curves with Minimal Acceleration.

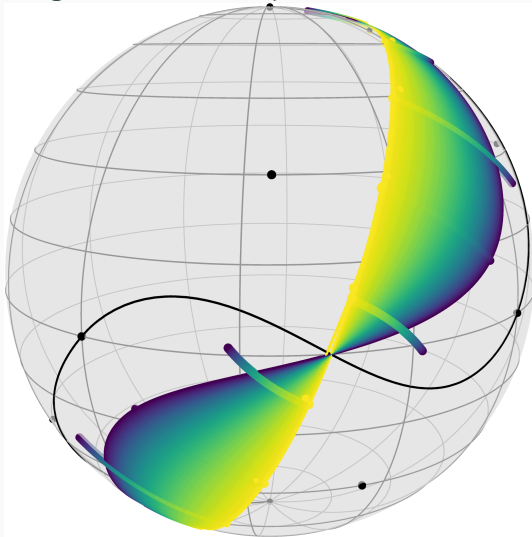
In the following video λ is slowly decreased from 10 to 0.



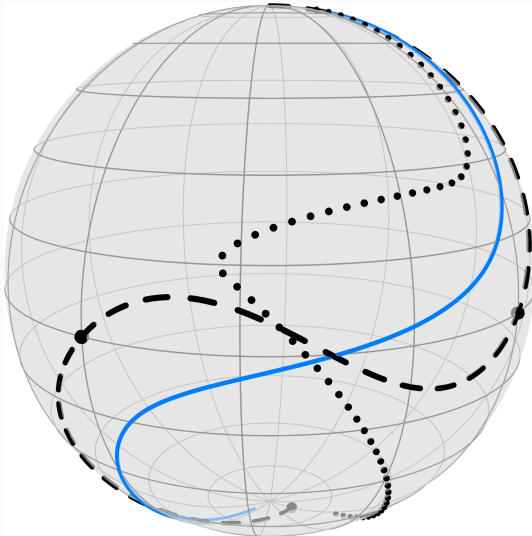
The initial setting, $\lambda = 10$.

Approximation by Bézier Curves with Minimal Acceleration.

In the following video λ is slowly decreased from 10 to 0.



Summary of reducing λ from 10 (violet) to zero (yellow).

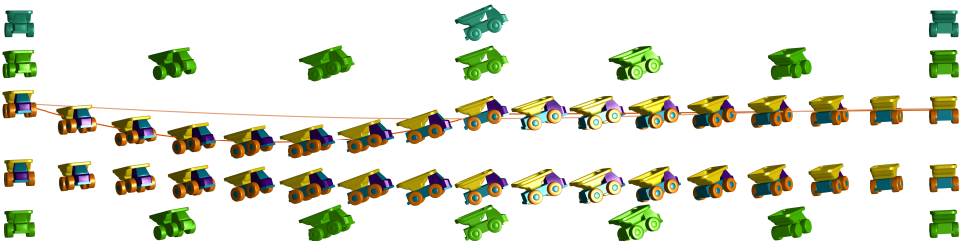


This curve (dashed) is “too global” to be solved in a tangent space (dotted) correctly, while our method (blue) still works.

An Example of Rotations $\mathcal{M} = \text{SO}(3)$

Initialization with approach from composite splines

[Gousenbourger, Massart, Absil, 2018]



Our method outperforms the approach of solving linear systems in tangent spaces, **but** their approach can serve as an initialization.

Models in manifold-valued imaging.

- Infimal Convolution [Bergmann, Fitschen, et al., 2017 ; Bergmann, Fitschen, et al., 2018]
- TGV [Bergmann, Fitschen, et al., 2018; Bredies et al., 2018]
- Nonlocal TV using the Graph Laplacian [Bergmann, Tenbrinck, 2018]
- denoising using second order statistics [Laus et al., 2017]

Algorithms In manifold-valued imaging.

- Douglas–Rachford splitting on Hadamard manifolds [Bergmann, Persch, Steidl, 2016]
- Half-quadratic Minimization (iteratively reweighted least squares) [Bergmann, Chan, et al., 2016; Grohs, Sprecher, 2016]

Summary

We defined second order differences on Riemannian manifolds.

We introduced two variational models:

- second order total variation
- minimizing the acceleration of a Bézier curve

We presented two algorithms to minimize the these models

- Cyclic Proximal Point Algorithm
- Gradient Descent

Future Work

- further models (Bézier surfaces, manifolds with no closed form for Jacobi fields,...)
- further algorithms, e.g. for constraint optimization
- further manifolds, e.g. infinite dimensional ones

Implement Algorithms in **Manopt.jl** an manifold optimization toolbox in Julia available at







<http://manoptjl.org>.

Paradigm:

Being able to use an(y) algorithm for a(ny) model directly on a(ny) manifold efficiently.

...in an open source programming language.

Selected References on Total Variation

-  Bačák, M.; Bergmann, R.; Steidl, G.; Weinmann, A. (2016). “A Second Order Non-Smooth Variational Model for Restoring Manifold-Valued Images”. *SIAM Journal on Scientific Computing* 38.1, A567–A597. DOI: 10.1137/15M101988X.
-  Bergmann, R.; Fitschen, J. H.; Persch, J.; Steidl, G. (2018). “Priors with Coupled First and Second Order Differences for Manifold-Valued Image Processing”. *Journal of Mathematical Imaging and Vision* 60.9, pp. 1459–1481. DOI: 10.1007/s10851-018-0840-y.
-  Bergmann, R.; Tenbrinck, D. (2018). “A graph framework for manifold-valued data”. *SIAM Journal on Imaging Sciences* 11.1, pp. 325–360. DOI: 10.1137/17M1118567.
-  Laus, F.; Nikolova, M.; Persch, J.; Steidl, G. (2017). “A Nonlocal Denoising Algorithm for Manifold-Valued Images Using Second Order Statistics”. *SIAM Journal on Imaging Sciences* 10.1, pp. 416–448. DOI: 10.1137/16M1087114.
-  Lellmann, J.; Strelakovsky, E.; Koetter, S.; Cremers, D. (2013). “Total variation regularization for functions with values in a manifold”. *IEEE ICCV 2013*, pp. 2944–2951. DOI: 10.1109/ICCV.2013.366.
-  Weinmann, A.; Demaret, L.; Storath, M. (2014). “Total variation regularization for manifold-valued data”. *SIAM Journal on Imaging Sciences* 7.4, pp. 2226–2257. DOI: 10.1137/130951075.

Selected References on Bézier Curves



Arnould, A.; Gousenbourger, P.-Y.; Samir, C.; Absil, P.-A.; Canis, M. (2015). "Fitting Smooth Paths on Riemannian Manifolds : Endometrial Surface Reconstruction and Preoperative MRI-Based Navigation". *GSI2015*. Ed. by F.Nielsen; F.Barbaresco. Springer International Publishing, pp. 491–498. DOI: 10.1007/978-3-319-25040-3_53.



Bergmann, R.; Gousenbourger, P.-Y. (2018). "A variational model for data fitting on manifolds by minimizing the acceleration of a Bézier curve". *Frontiers in Applied Mathematics and Statistics*. DOI: 10.3389/fams.2018.00059. arXiv: 1807.10090.



Boumal, N.; Absil, P. A. (2011). "A discrete regression method on manifolds and its application to data on $SO(n)$ ". *IFAC Proceedings Volumes (IFAC-PapersOnline)*. Vol. 18. PART 1, pp. 2284–2289. DOI: 10.3182/20110828-6-IT-1002.00542.



Gousenbourger, P.-Y.; Massart, E.; Absil, P.-A. (2018). "Data fitting on manifolds with composite Bézier-like curves and blended cubic splines". *Journal of Mathematical Imaging and Vision*. accepted. DOI: 10.1007/s10851-018-0865-2.



Samir, C.; Absil, P.-A.; Srivastava, A.; Klassen, E. (2012). "A Gradient-Descent Method for Curve Fitting on Riemannian Manifolds". *Foundations of Computational Mathematics* 12.1, pp. 49–73. DOI: 10.1007/s10208-011-9091-7.