A Variational Model for Data Fitting on Manifolds by Minimizing the Acceleration of a Bézier Curve

Ronny Bergmann[a]

Technische Universität Chemnitz

Institutsseminar, Institut für Numerische Mathematik, TU Dresden,

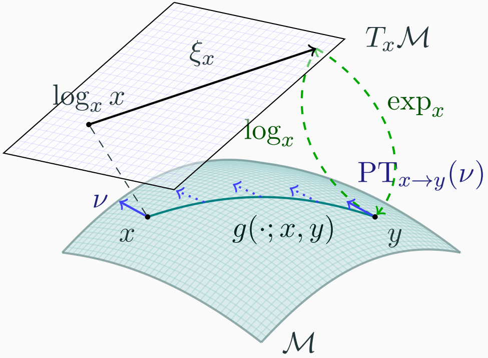
Dresden, November 27, 2018

[a] joint work with P.-Y. Gousenbourger, UCLouvain, Louvain-la-Neuve, Belgium.

Moin.

1. Riemannian Manifolds & Optimization
2. Data Fitting on Riemannian Manifolds
3. Bézier Curves and Generalized Bézier Curves
4. Discretized Acceleration of a Bézier Curve
5. Gradient Descent on a Manifold
6. Numerical Examples

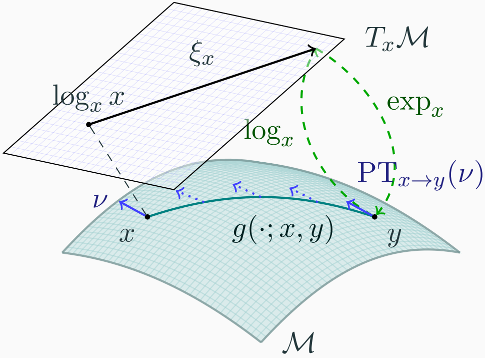
An m-dimensional Riemannian Manifold $\mathcal{M}$



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 continously varying inner product on the tangential spaces.

[Absil, Mahony, Sepulchre, 2008]

An m-dimensional Riemannian Manifold $\mathcal{M}$



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

Tangent space $\mathrm{T}_x\mathcal{M}$ at x, with inner product $\langle\cdot,\cdot\rangle_x$

Logarithmic map $\log_x y = \dot{g}(0;x,y)$ "speed towards y"

Exponential map $\exp_x \xi_x = g(1)$, where $g(0) = x$, $\dot{g}(0) = \xi_x$

Parallel transport $\mathrm{PT}_{x\to y}(\nu)$ of $\nu \in \mathrm{T}_x\mathcal{M}$ along $g(\cdot;x,y)$

Optimization on Manifolds

Let $\mathcal{N}$ be a Riemannian manifold and $E\colon \mathcal{N} \to \mathbb{R}$.

We aim to solve

$$\operatorname*{argmin}_{x\in\mathcal{N}} E(x)$$

- often: product manifold $\mathcal{N} = \mathcal{M}^n$
- for $n \in \mathbb{N}^2$: manifold-valued image processing

$\Rightarrow$ highdimensional problem

- locally: convexity defined via geodesics

Variational Methods on Manifolds

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

$$E(x) = D(x; f) + \alpha R(x)$$

- $\alpha > 0$ is a weight
- $f \in \mathcal{N}$ is given Data
- data or similarity term $D(x; f)$
- regularizer / prior $R(x)$

Differential and Gradient

The differential $D_x f = Df\colon T\mathcal{M} \to \mathbb{R}$ of a real-valued function $f\colon \mathcal{M} \to \mathbb{R}$ is the push-forward of f.

Intuition: Given $x \in \mathcal{M}$ and $\xi \in T_x\mathcal{M}$, then $Df(x)[\xi]$ is the directional derivative of f.

The gradient $\nabla f\colon \mathcal{M} \to T\mathcal{M}$ is the tangent vector fulfilling

$$\langle \nabla_{\mathcal{M}} f(x), \eta \rangle_x = Df(x)[\eta] \text{ for all } \eta \in T_x\mathcal{M}$$

$\Rightarrow$ gradient descent (with e.g. Armijo's rule)

Data Fitting on Manifolds

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

- Γ set of geodesics & approximation: geodesic regression [Rentmeesters, '11; Fletcher, '13; Boumal '13]
- Γ Sobolev space of curves: Inifinite-dimensional problem [Samir et. al.,'12]
- Γ composite Bézier curves; LSs in tangent spaces [Arnould et. al. '15; Gousenbourger, Massart, Absil, '18]
- Discretized curve, $\Gamma = \mathcal{M}^N$, [Boumal, Absil, '11]

This talk

"nice" means minimal (discretized) acceleration ("as straight as possible") for Γ the set of composite Bézier curves.
In Euclidean space: Natural cubic splines as closed form solution.

(Euclidean) Bézier Curves

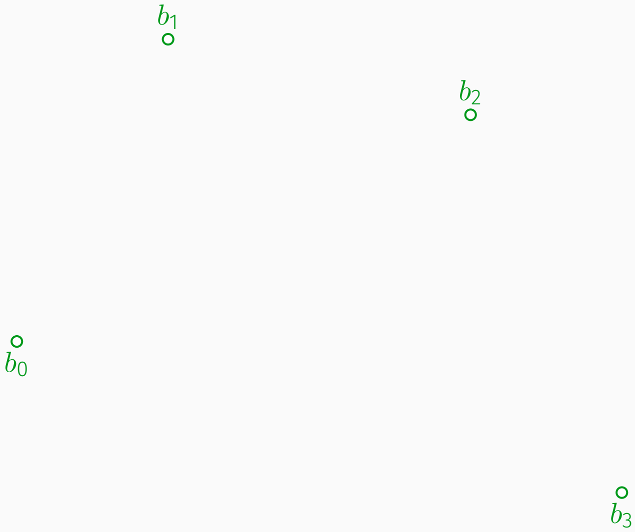
Definition [Bézier, '62]
A Bézier curve β_K of degree $K \in \mathbb{N}_0$ is a function $\beta_K : [0,1] \to \mathbb{R}^d$ parametrized by control points $b_0, \ldots, b_K \in \mathbb{R}^n$ and defined by

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

[Bernstein, 1912]
where $B_{j,K} = \binom{K}{j} t^j (1-t)^{K-j}$ are the Bernstein polynomials of degree K.

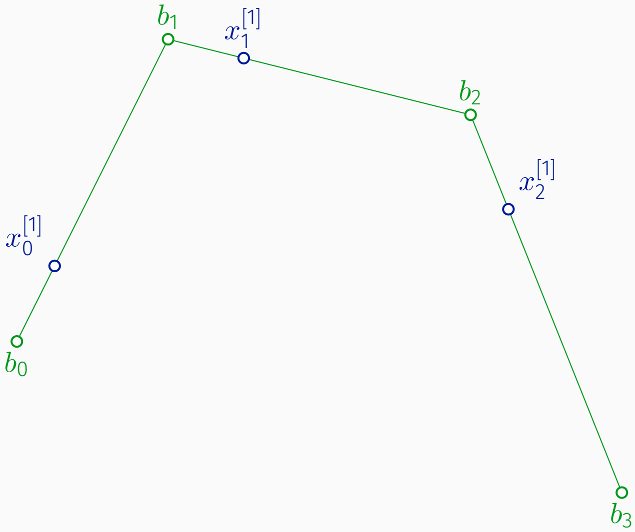
Evaluation via Casteljau's algorithm. [de Casteljau, '59]

Illustration of de Casteljau's Algorithm



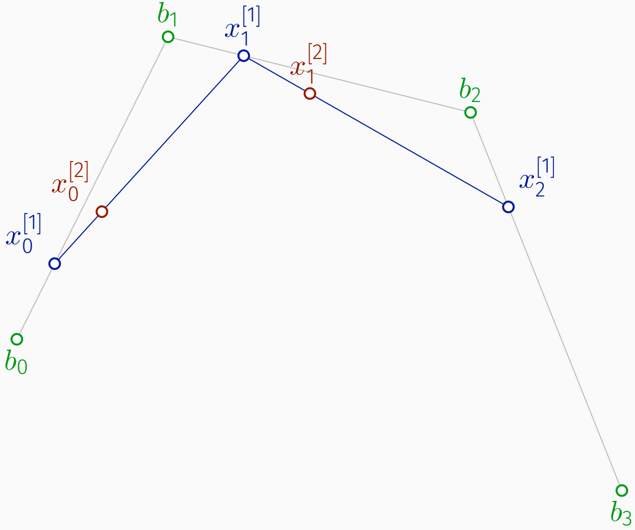
The set of control points b_0, b_1, b_2, b_3.

Illustration of de Casteljau's Algorithm



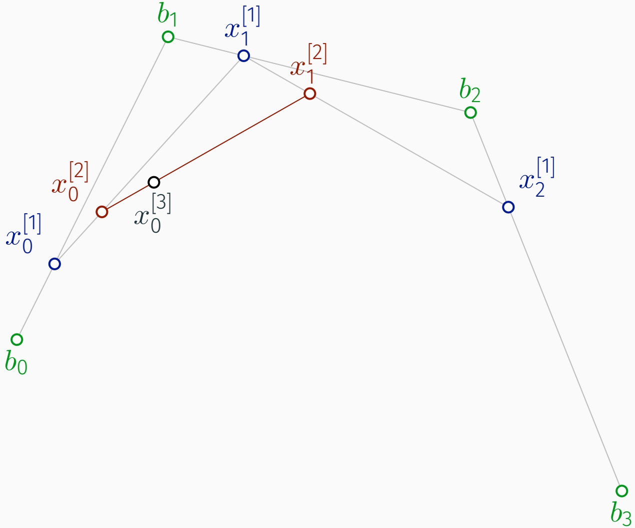
Evaluate line segments at $t = \frac{1}{4}$, obtain $x_0^{[1]}, x_1^{[1]}, x_2^{[1]}$.

Illustration of de Casteljau's Algorithm



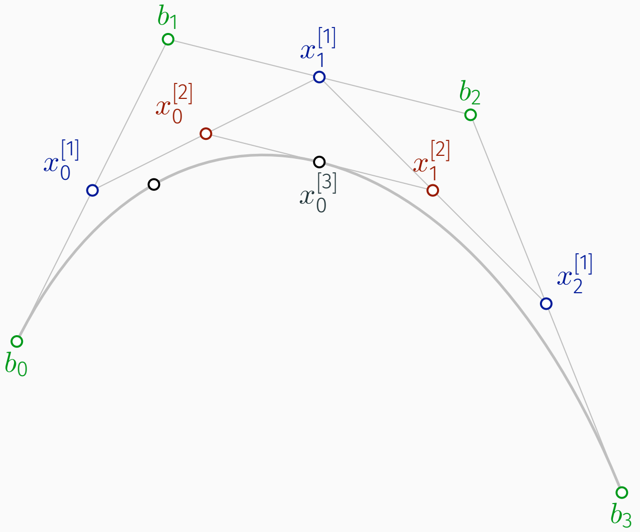
Repeat evaluation for new line segments to obtain $x_0^{[2]}, x_1^{[2]}$.

Illustration of de Casteljau's Algorithm



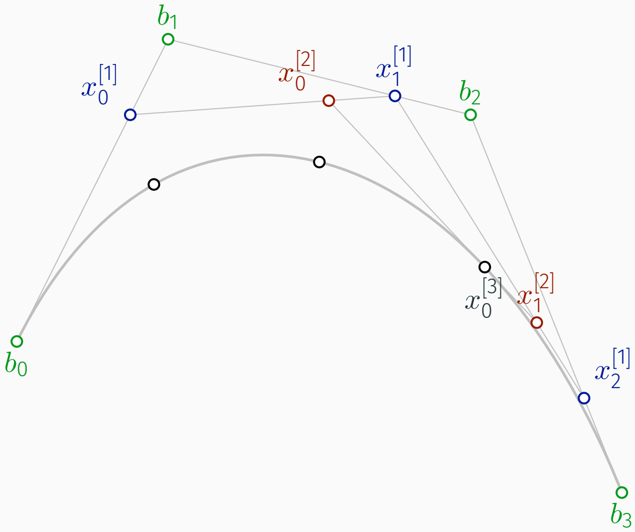
Repeat for the last segment to obtain $\beta_3(\frac{1}{4}; b_0, b_1, b_2, b_3) = x_0^{[3]}$.

Illustration of de Casteljau's Algorithm



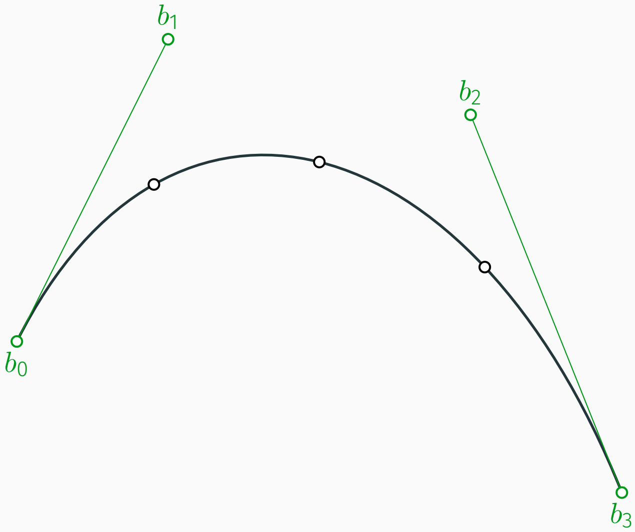
Same procedure for evaluation of $\beta_3(\frac{1}{2}; b_0, b_1, b_2, b_3)$.

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 Bezier curve $B\colon [0,n] \to \mathbb{R}^d$ is defined as

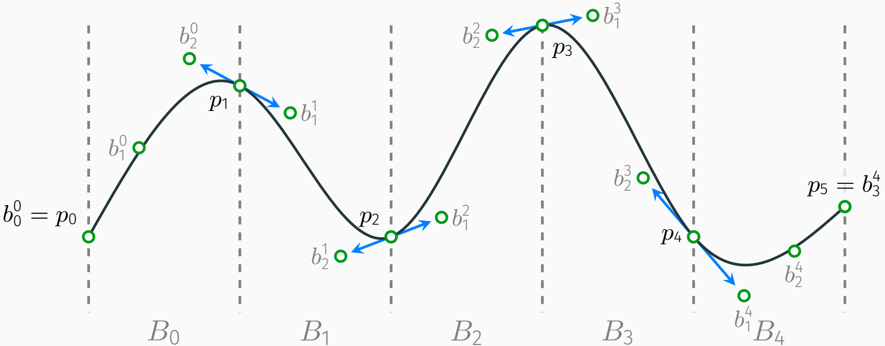
$$B(t) := \begin{cases} \beta_K(t; b_0^0, \ldots, b_K^0) & \text{if } t \in [0,1], \\ \beta_K(t-i; b_0^i, \ldots, b_K^i), & \text{if } t \in (i, i+1], \quad i = 1, \ldots, n-1. \end{cases}$$

Composite Bézier Curves

Definition
A composite Bezier curve $B\colon [0,n] \to \mathbb{R}^d$ is defined as

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

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



Composite Bézier Curves

Definition
A composite Bezier curve $B\colon [0,n] \to \mathbb{R}^d$ 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 ith 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 $b_0, \ldots, b_K \in \mathcal{M}$, $K \in \mathbb{N}$.

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

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

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 $b_0, \ldots, b_K \in \mathcal{M}$, $K \in \mathbb{N}$.

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

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

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 $B\colon [0, n] \to \mathcal{M}$ completely analogue (using geodesics for line segments)

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

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

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

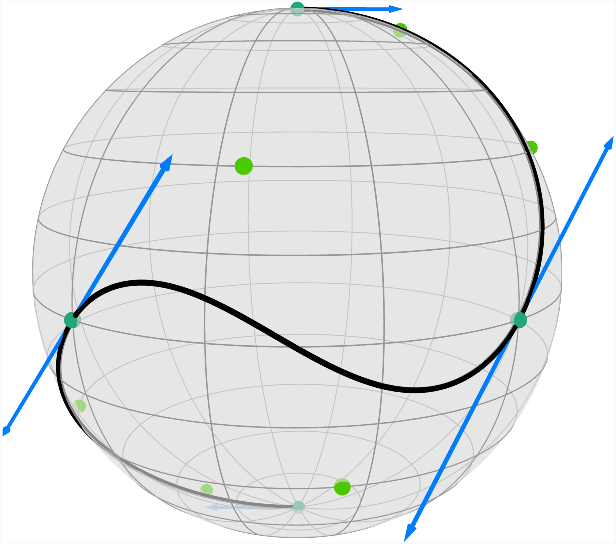
if $k > 0$, and

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

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

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

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

$$E(B) = \frac{\lambda}{2} \sum_{k=0}^{n} d^2_{\mathcal{M}}(B(k), d_k) + \int_0^n \Big\| \frac{D^2 B(t)}{\mathrm{d}t^2} \Big\|^2_{B(t)}, \mathrm{d}t \qquad \lambda > 0,$$

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^i_j, i.e. on $\mathcal{M}^L$ with
  - $L = n(K-1) + 2$
  - $\lambda \to \infty$ yields interpolation $(p_k = d_k) \Rightarrow L = n(K-2) + 1$
- On $\mathcal{M} = \mathbb{R}^m$: closed form solution, natural (cubic) splines

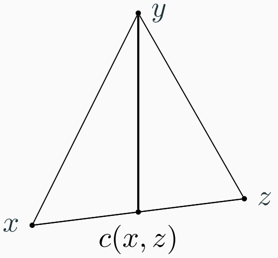
Interlude: Second Order Differences on Manifolds

Second order difference: [RB et al., 2014; RB, Weinmann, 2016; Bačák et al., 2016]

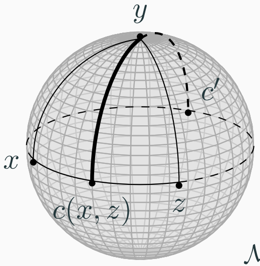
$$\mathrm{d}_2(x,y,z) \coloneqq \min_{c \in \mathcal{C}_{x,z}} \mathrm{d}_{\mathcal{M}}(c,y), \quad x,y,z \in \mathcal{M},$$

$\mathcal{C}_{x,z}$ mid point(s) of geodesic(s) $g(\cdot;x,z)$

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



$$\min_{c \in \mathcal{C}_{x,z}} d_{\mathcal{M}}(c,y)$$



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

Discretizing the Data Fitting Model

We discretize the absolute second order covariant derivative

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

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

Evaluating $E(B)$ 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
- N squared distances to obtain the second order absolute finite differences squared

Gradient and Chain Rule on a Manifold

The gradient $\nabla_{\mathcal{M}} f(x) \in T_x\mathcal{M}$ of $f\colon \mathcal{M} \to \mathbb{R}$, $x \in \mathcal{M}$, is defined as the tangent vector that fulfills

$$\langle \nabla_{\mathcal{M}} f(x), \xi \rangle_x = Df(x)[\xi] \quad \text{for all} \quad \xi \in T_x\mathcal{M}.$$

For a composition $F(x) = (g \circ h)(x) = g(h(x))$ of two functions $g, h\colon \mathcal{M} \to \mathcal{M}$ the chain rule reads for $x \in \mathcal{M}$ and $\xi \in T_x\mathcal{M}$ as

$$D_x F[\xi] = D_{h(x)} g\bigl[D_x h[\xi]\bigr],$$

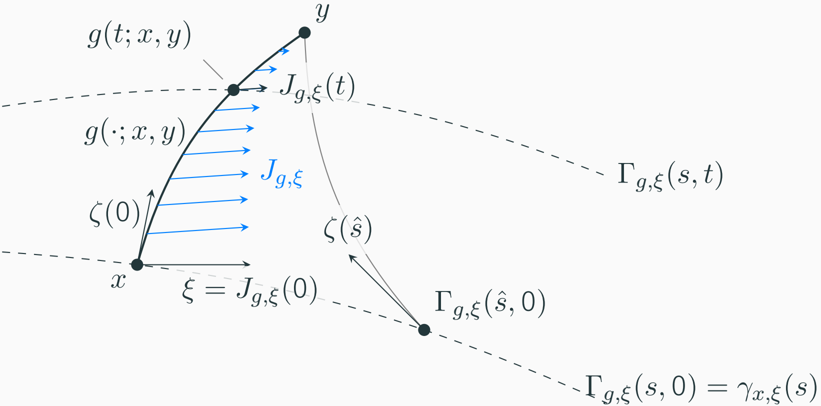
where $D_x h[\xi] \in T_{h(x)}\mathcal{M}$ and $D_x F[\xi] \in T_{F(x)}\mathcal{M}$.

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

The geodesic variation

$$\Gamma_{g,\xi}(s,t) := \exp_{\gamma_{x,\xi}(s)}(t\zeta(s)), \qquad s \in (-\varepsilon,\varepsilon),\ t \in [0,1], \varepsilon > 0.$$

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



Then the differential reads $D_x g(t,\cdot,y)[\xi] = J_{g,\xi}(t)$.

Implementing Jacobi Fields on Symmetric Spaces

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

Then

- one can diagonalize the curvature tensor R,
- let κ_ℓ denote its eigenvalues.
- let $\{\xi_1, \ldots, \xi_m\} \subseteq T_x\mathcal{M}$ be an ONB of eigenvalues with $\xi_1 = \log_x y$.
- parallel transport $\Xi_j(t) = \mathrm{PT}_{x \to g(t;x,y)}\, \xi_j$, $j = 1, \ldots, m$

Implementing Jacobi Fields on Symmetric Spaces II

Decompose $\eta = \sum_{i=1}^{m} \eta_\ell \xi \ell$. Then

$$D_x g(t; x, y)[\eta] = J_{g,\eta}(t) = \sum_{\ell=1}^{m} \eta_\ell J_{g,\xi_\ell}(t),$$

with

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

Implementing the Gradient using adjoint Jacobi Fields.

The adjoint Jacobi fields

$$J^*_{g,\cdot}(t)\colon T_{g(t)}\mathcal{M} \to T_x\mathcal{M}$$

are characterized by

$$\langle J_{g,\xi}(t), \eta\rangle_{g(t)} = \langle \xi, J^*_{g,\eta}(t)\rangle_x, \quad \text{for all } \xi \in T_x\mathcal{M}, \eta \in T_{g(t;x,y)}\mathcal{M}.$$

- can be computed without extra efforts, i.e. the same ODEs occur.

$\Rightarrow$ can be used to calculate the gradient

- the gradient of iterated evaluations of geodesics can be computed by composition of (adjoint) Jacobi fields

Gradient Descent on a Manifold

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

Input.

- $f\colon \mathcal{N} \to \mathbb{R}$,
- its gradient $\nabla_{\mathcal{N}} f$,
- initial data $x^{(0)} = b \in \mathcal{N}$
- step sizes $s_k > 0, k \in \mathbb{N}$.

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

$k \leftarrow 0$

repeat

$\quad x^{(k+1)} \leftarrow \exp_{x^{(k)}}\bigl(-s_k \nabla_{\mathcal{N}} f(x^{(k)})\bigr)$

$\quad k \leftarrow k + 1$

until a stopping criterion is reached

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

Armijo Step Size Rule

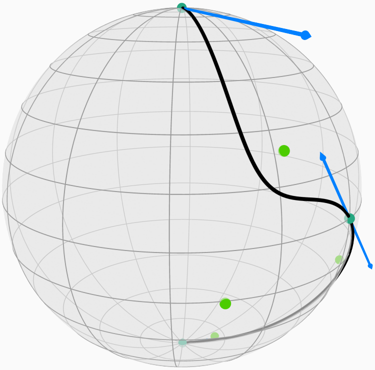
Let $x = x^{(k)}$ be an iterate from the gradient descent algorithm, $\beta, \sigma \in (0, 1), \alpha > 0$.

Let m be the smallest positive integer such that

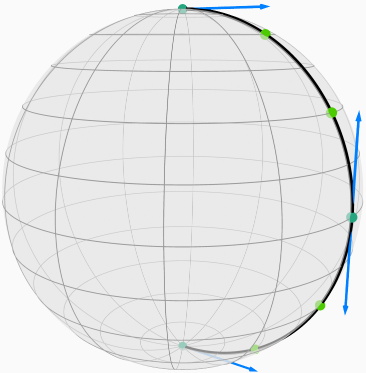
$$f(x) - f\big(\exp_x(-\beta^m \alpha \nabla_{\mathcal{N}} f(x))\big) \geq \sigma \beta^m \alpha \|\nabla_{\mathcal{N}} f(x)\|_x$$

Set the step size $s_k \coloneqq \beta^m \alpha$.

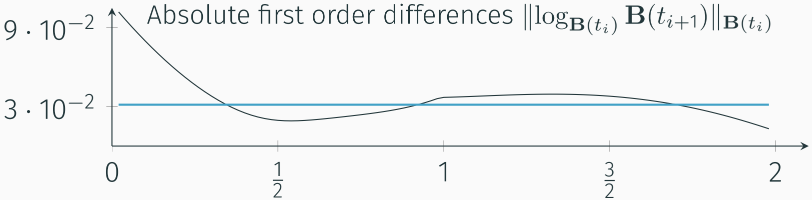
Minimizing with Known Minimizer



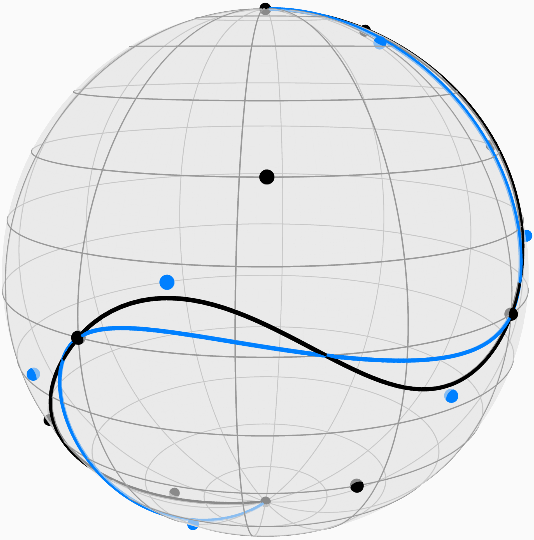
Original



Minimized



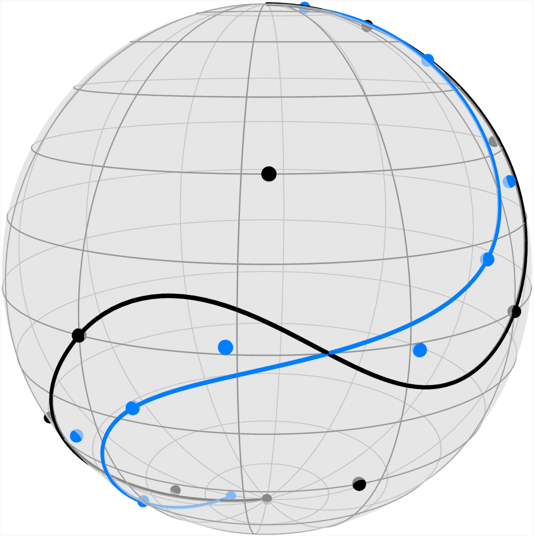
Interpolation by Bézier Curves with Minimal Acceleration.



A comp. Bezier curve (black) and its mnimizer (blue).

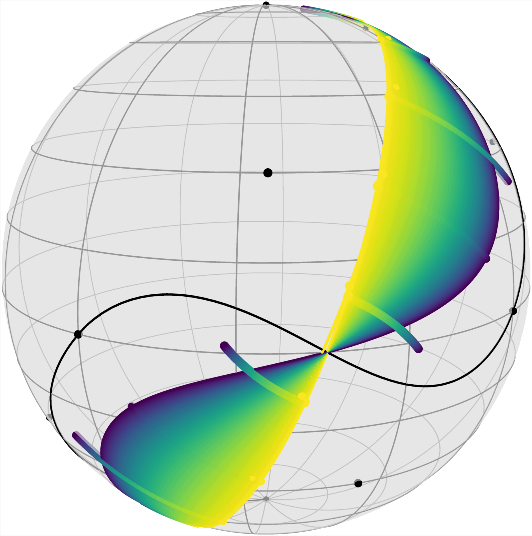
Approximation by Bézier Curves with Minimal Acceleration.

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



The initial setting, $\lambda = 10$.

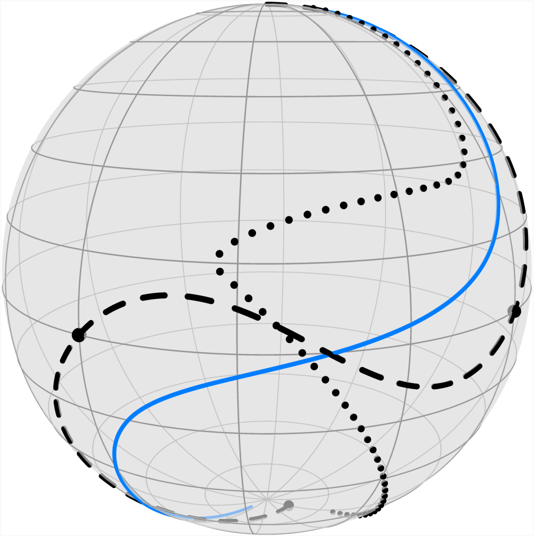
In the following video λ is slowly decreased from 10 to 0.



Summary of the video.

Comparison to Previous Approach

[Gousenbourger, Massart, Absil, 2018]

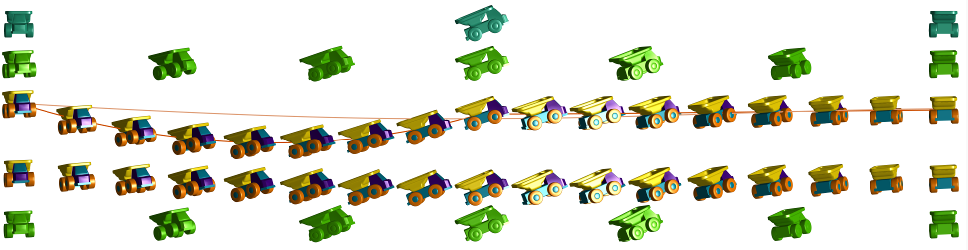


This curve (dashed) is “too global” to be solved in a tangent space (dotted) correctly, while our method (blue) still works.

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

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.

Summary

- Data fitting on manifolds with Bézier curves minimizing their acceleration
- computed the gradient with respect to control points
- employed Jacobi fields and their adjoints.
- implemented within the MVIR toolbox (available soon)
  ronnybergmann.net/mvirt/
- a Julia implementation in preparation (Manopt.jl)

Literature

- M. Bačák, R. Bergmann, G. Steidl, and A. Weinmann. "A Second Order Non-Smooth Variational Model for Restoring Manifold-Valued Images". In: *SIAM Journal on Scientific Computing* 38.1 (2016), A567–A597. DOI: 10.1137/15M101988X.
- R. Bergmann and P.-Y. Gousenbourger. "A variational model for data fitting on manifolds by minimizing the acceleration of a Bézier curve". In: *Frontiers in Applied Mathematics and Statistics* (2018). DOI: 10.3389/fams.2018.00059. arXiv: 1807.10090.
- P.-A. Absil, R. Mahony, and R. Sepulchre. *Optimization Algorithms on Matrix Manifolds*. Princeton University Press, 2008. DOI: 10.1515/9781400830244.
- P.-Y. Gousenbourger, E. Massart, and P.-A. Absil. "Data fitting on manifolds with composite Bézier-like curves and blended cubic splines". In: *Journal of Mathematical Imaging and Vision* (2018). accepted. DOI: 10.1007/s10851-018-0865-2.