NTNU

Norwegian University of Science and Technology

# The Convex Bundle Method on Hadamard Manifolds

Hajg Jasa

joint work with

Ronny Bergmann and Roland Herzog

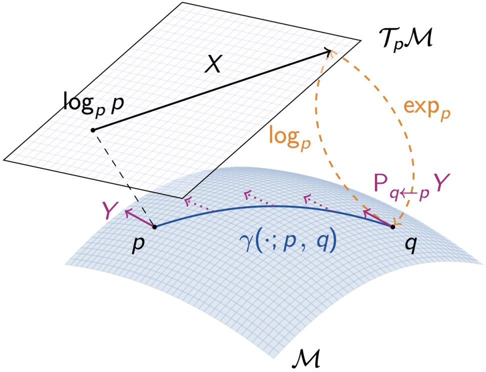
Ph.D. Seminar, NTNU, Trondheim

March 23rd, 2023

# Riemannian Geometry

Notation

- Smooth Riemannian manifold $\mathcal{M}$
- Tangent space $\mathcal{T}_p\mathcal{M}$ at the point $p \in \mathcal{M}$
- Inner product $\langle\cdot,\cdot\rangle_p\colon \mathcal{T}_p\mathcal{M} \times \mathcal{T}_p\mathcal{M} \to \mathbb{R}$
- Exponential map $\exp_p X_p = \gamma_{pq}(1) = q$
- Logarithmic map $\log_p q = \exp_p^{-1} q = X_p$
- Parallel transport $\mathrm{P}_{q\leftarrow p}\colon \mathcal{T}_p\mathcal{M} \to \mathcal{T}_q\mathcal{M}$



# The Problem

Consider the following minimization problem

$$\operatorname*{arg\,min}_{p \in \mathcal{M}} f(p)$$

where

# The Problem

Consider the following minimization problem

$$\operatorname*{arg\,min}_{p \in \mathcal{M}} f(p)$$

where

- $\mathcal{M}$ is a Hadamard manifold:

# The Problem

Consider the following minimization problem

$$\arg\min_{p \in \mathcal{M}} f(p)$$

where

- $\mathcal{M}$ is a Hadamard manifold:
  - complete
  - simply connected
  - non-positive sectional curvature everywhere

# The Problem

Consider the following minimization problem

$$\operatorname*{arg\,min}_{p \in \mathcal{M}} f(p)$$

where

- $\mathcal{M}$ is a Hadamard manifold:
  - complete
  - simply connected
  - non-positive sectional curvature everywhere
- $f\colon \mathcal{M} \to \mathbb{R}$ is

# The Problem

Consider the following minimization problem

$$\operatorname*{arg\,min}_{p \in \mathcal{M}} f(p)$$

where

- $\mathcal{M}$ is a Hadamard manifold:
  - complete
  - simply connected
  - non-positive sectional curvature everywhere
- $f \colon \mathcal{M} \to \mathbb{R}$ is
  - geodesically convex: $f \circ \gamma : [0, 1] \to \mathbb{R}$ is convex in the usual sense
  - lower semi-continuous: $\liminf_{q \to p} f(q) \geq f(p)$

# The Problem

Consider the following minimization problem

$$\operatorname*{arg\,min}_{p \in \mathcal{M}} f(p)$$

where

- $\mathcal{M}$ is a Hadamard manifold:
  - complete
  - simply connected
  - non-positive sectional curvature everywhere
- $f\colon \mathcal{M} \to \mathbb{R}$ is
  - geodesically convex: $f \circ \gamma : [0, 1] \to \mathbb{R}$ is convex in the usual sense
  - lower semi-continuous: $\liminf_{q \to p} f(q) \geq f(p)$

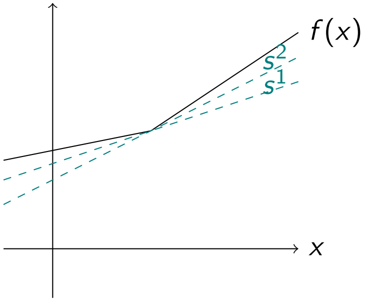
**Goal**. Solve this non-smooth optimization problem with the bundle method

# The Subdifferentials

For a convex function, the subdifferential is defined as

$$\partial f(x) = \left\{ s \in \mathbb{R}^n \,\middle|\, f(y) \geq f(x) + (s)^T(y - x) \text{ for all } y \in \mathbb{R}^n \right\}$$

and it is a non-empty, closed and convex subset.



# The Subdifferentials

For a convex function, the subdifferential is defined as

$$\partial f(x) = \left\{ s \in \mathbb{R}^n \,\middle|\, f(y) \geq f(x) + (s)^T (y - x) \text{ for all } y \in \mathbb{R}^n \right\}$$

and it is a non-empty, closed and convex subset. Let $\varepsilon > 0$. The $\varepsilon-$subdifferential is

$$\partial_\varepsilon f(x) = \left\{ s \in \mathbb{R}^n \,\middle|\, f(y) \geq f(x) + (s)^T (y - x) - \varepsilon \text{ for all } y \in \mathbb{R}^n \right\}$$

and

$$\partial f(x) \subseteq \partial_\varepsilon f(x)$$

# The Subdifferentials

For a geodesically convex function, the subdifferential is defined as

$$\partial f(p) = \left\{ X_p \in \mathcal{T}_p\mathcal{M} \,\middle|\, f(q) \geq f(p) + \langle X_p, \log_p q \rangle \text{ for all } q \in \mathcal{U} \subseteq \mathcal{M} \right\}$$

and it is a non-empty, closed and convex subset. Let $\varepsilon > 0$. The $\varepsilon-$subdifferential is

$$\partial_\varepsilon f(p) = \left\{ X_p \in \mathcal{T}_p\mathcal{M} \,\middle|\, f(q) \geq f(p) + \langle X_p, \log_p q \rangle - \varepsilon \text{ for all } q \in \mathcal{U} \subseteq \mathcal{M} \right\}$$

and

$$\partial f(p) \subseteq \partial_\varepsilon f(p)$$

# The Bundle Method

Bundle methods are about descent as well as stability.

# The Bundle Method

Bundle methods are about descent as well as stability.

Core characteristics of a general bundle algorithm (Bonnans et al. 2006):

# The Bundle Method

Bundle methods are about descent as well as stability.

Core characteristics of a general bundle algorithm (Bonnans et al. 2006):

- keeps memory of the last “best” points

# The Bundle Method

Bundle methods are about descent as well as stability.

Core characteristics of a general bundle algorithm (Bonnans et al. 2006):

- keeps memory of the last “best” points
- solves a stabilization subproblem

# The Bundle Method

Bundle methods are about descent as well as stability.

Core characteristics of a general bundle algorithm (Bonnans et al. 2006):

- keeps memory of the last "best" points
- solves a stabilization subproblem
  - we employ a dual approach with the $\varepsilon-$subdifferential as in Geiger and Kanzow 2002

# The Bundle Method

Bundle methods are about descent as well as stability.

Core characteristics of a general bundle algorithm (Bonnans et al. 2006):

- keeps memory of the last “best” points
- solves a stabilization subproblem
  - we employ a dual approach with the $\varepsilon-$subdifferential as in Geiger and Kanzow 2002
- generates sequences of *candidate* points and *stability centers*

# Approximating the $\varepsilon-$subdifferential

Let $\varepsilon > 0$ and $k \in \mathbb{N}$. Given $x_1, \ldots, x_k \in \mathbb{R}^n$, and $s^j \in \partial f(x_j)$, then

# Approximating the $\varepsilon-$subdifferential

Let $\varepsilon > 0$ and $k \in \mathbb{N}$. Given $x_1, \ldots, x_k \in \mathbb{R}^n$, and $s^j \in \partial f(x_j)$, then $s^j \in \partial_\varepsilon f(x_k)$

# Approximating the $\varepsilon-$subdifferential

Let $\varepsilon > 0$ and $k \in \mathbb{N}$. Given $x_1, \dots, x_k \in \mathbb{R}^n$, and $s^j \in \partial f(x_j)$, then

$$s^j \in \partial_\varepsilon f(x_k) \quad \text{if} \quad \varepsilon \geq e_j^k = f(x_k) - f(x_j) - (s^j)^T(x_k - x_j) \quad \text{for all} \quad j \in \{1, \dots, k\}$$

# Approximating the $\varepsilon-$subdifferential

Let $\varepsilon > 0$ and $k \in \mathbb{N}$. Given $x_1, \dots, x_k \in \mathbb{R}^n$, and $s^j \in \partial f(x_j)$, then

$$s^j \in \partial_\varepsilon f(x_k) \quad \text{if} \quad \varepsilon \geq e_j^k = f(x_k) - f(x_j) - (s^j)^T(x_k - x_j) \quad \text{for all} \quad j \in \{1, \dots, k\}$$

- Main challenge on manifolds: given $p_1, \dots, p_k \in \mathcal{M}$, and $X_{p_j} \in \partial f(p_j)$, then

# Approximating the $\varepsilon-$subdifferential

Let $\varepsilon > 0$ and $k \in \mathbb{N}$. Given $x_1, \dots, x_k \in \mathbb{R}^n$, and $s^j \in \partial f(x_j)$, then

$$s^j \in \partial_\varepsilon f(x_k) \quad \text{if} \quad \varepsilon \geq e_j^k = f(x_k) - f(x_j) - (s^j)^T(x_k - x_j) \quad \text{for all} \quad j \in \{1, \dots, k\}$$

- Main challenge on manifolds: given $p_1, \dots, p_k \in \mathcal{M}$, and $X_{p_j} \in \partial f(p_j)$, then

$$\mathrm{P}_{p_k \leftarrow p_j} X_{p_j} \in \partial_\varepsilon f(p_k)?$$

# Approximating the $\varepsilon-$subdifferential

Let $\varepsilon > 0$ and $k \in \mathbb{N}$. Given $x_1, \ldots, x_k \in \mathbb{R}^n$, and $s^j \in \partial f(x_j)$, then

$$s^j \in \partial_\varepsilon f(x_k) \quad \text{if} \quad \varepsilon \geq e_j^k = f(x_k) - f(x_j) - (s^j)^T(x_k - x_j) \quad \text{for all} \quad j \in \{1, \ldots, k\}$$

- Main challenge on manifolds: given $p_1, \ldots, p_k \in \mathcal{M}$, and $X_{p_j} \in \partial f(p_j)$, then
$$\mathrm{P}_{p_k \leftarrow p_j} X_{p_j} \in \partial_\varepsilon f(p_k)?$$
- For all $\sigma > 0$ and all $p_j, p_{j+1} \in \mathcal{M}$, there exists $\delta > 0$ such that
$$\|\mathrm{P}_{p_{j+1} \leftarrow p_j} X_{p_j} - \mathrm{P}_{p_{j+1} \leftarrow p_{j+2}} \mathrm{P}_{p_{j+2} \leftarrow p_j} X_{p_j}\| < \sigma \|X_{p_j}\|$$
for all $p_{j+2} \in B_\delta(p_j) \cup B_\delta(p_{j+1})$ and all $X_{p_j} \in \mathcal{T}_{p_j}\mathcal{M}$ (Azagra and Ferrera 2005, p. 169). We localize by choosing a set $\mathcal{U}$ with $\varrho = \sqrt{2}\,\mathrm{diam}(\mathcal{U}) < +\infty$.

# Approximating the $\varepsilon-$subdifferential

Let $\varepsilon > 0$ and $k \in \mathbb{N}$. Given $x_1, \ldots, x_k \in \mathbb{R}^n$, and $s^j \in \partial f(x_j)$, then

$$s^j \in \partial_\varepsilon f(x_k) \quad \text{if} \quad \varepsilon \geq e_j^k = f(x_k) - f(x_j) - (s^j)^T(x_k - x_j) \quad \text{for all} \quad j \in \{1, \ldots, k\}$$

- Main challenge on manifolds: given $p_1, \ldots, p_k \in \mathcal{M}$, and $X_{p_j} \in \partial f(p_j)$, then
$$\mathrm{P}_{p_k \leftarrow p_j} X_{p_j} \in \partial_\varepsilon f(p_k)?$$
- For all $\sigma > 0$ and all $p_j, p_{j+1} \in \mathcal{M}$, there exists $\delta > 0$ such that
$$\|\mathrm{P}_{p_{j+1} \leftarrow p_j} X_{p_j} - \mathrm{P}_{p_{j+1} \leftarrow p_{j+2}} \mathrm{P}_{p_{j+2} \leftarrow p_j} X_{p_j}\| < \sigma \|X_{p_j}\|$$
for all $p_{j+2} \in B_\delta(p_j) \cup B_\delta(p_{j+1})$ and all $X_{p_j} \in \mathcal{T}_{p_j}\mathcal{M}$ (Azagra and Ferrera 2005, p. 169). We localize by choosing a set $\mathcal{U}$ with $\varrho = \sqrt{2}\,\mathrm{diam}(\mathcal{U}) < +\infty$.
- Enlarge the $\varepsilon-$subdifferential by adding $\varrho\sqrt{\sigma}\|X_{p_j}\|$:
$$c_j^k = f(p_k) - f(p_j) - \langle X_{p_j}, \log_{p_j} p_k \rangle + \varrho\sqrt{\sigma}\|X_{p_j}\|$$

# The Euclidean Subproblem

Let $k \in \mathbb{N}$ and $j \in \{1, \ldots, k\} = J_k$ and let $\lambda_j$ be convex coefficients.

# The Euclidean Subproblem

Let $k \in \mathbb{N}$ and $j \in \{1, \ldots, k\} = J_k$ and let $\lambda_j$ be convex coefficients. In the Euclidean case:

$$s^j \in \partial f(x_j) \Rightarrow \sum_{j=1}^{k} \lambda_j s^j \in \partial_\varepsilon f(x_k)$$

# The Euclidean Subproblem

Let $k \in \mathbb{N}$ and $j \in \{1, \ldots, k\} = J_k$ and let $\lambda_j$ be convex coefficients. In the Euclidean case:

$$s^j \in \partial f(x_j) \Rightarrow \sum_{j=1}^{k} \lambda_j s^j \in \partial_\varepsilon f(x_k) \quad \text{if} \quad \varepsilon \geq \sum_{j=1}^{k} \lambda_j e_j^k$$

# The Euclidean Subproblem

Let $k \in \mathbb{N}$ and $j \in \{1, \ldots, k\} = J_k$ and let $\lambda_j$ be convex coefficients. In the Euclidean case:

$$s^j \in \partial f(x_j) \Rightarrow \sum_{j=1}^{k} \lambda_j s^j \in \partial_\varepsilon f(x_k) \quad \text{if} \quad \varepsilon \geq \sum_{j=1}^{k} \lambda_j e_j^k$$

Then the search direction is

$$d^k = -\sum_{j=1}^{k} \lambda_j s^j$$

# The Euclidean Subproblem

Let $k \in \mathbb{N}$ and $j \in \{1, \dots, k\} = J_k$ and let $\lambda_j$ be convex coefficients. In the Euclidean case:

$$s^j \in \partial f(x_j) \Rightarrow \sum_{j=1}^{k} \lambda_j s^j \in \partial_\varepsilon f(x_k) \quad \text{if} \quad \varepsilon \geq \sum_{j=1}^{k} \lambda_j e_j^k$$

Then the search direction is

$$d^k = -\sum_{j=1}^{k} \lambda_j s^j$$

where the coefficients $\lambda_j$ are solution to

# The Euclidean Subproblem

Let $k \in \mathbb{N}$ and $j \in \{1, \ldots, k\} = J_k$ and let $\lambda_j$ be convex coefficients. In the Euclidean case:

$$s^j \in \partial f(x_j) \Rightarrow \sum_{j=1}^{k} \lambda_j s^j \in \partial_\varepsilon f(x_k) \quad \text{if} \quad \varepsilon \geq \sum_{j=1}^{k} \lambda_j e_j^k$$

Then the search direction is

$$d^k = -\sum_{j=1}^{k} \lambda_j s^j$$

where the coefficients $\lambda_j$ are solution to

$$\begin{aligned} \operatorname*{arg\,min}_{\lambda \in \mathbb{R}^{|J_k|}} \quad & \frac{1}{2} \left\| \sum_{j \in J_k} \lambda_j s^j \right\|^2 + \sum_{j \in J_k} \lambda_j e_j^k \\ \text{s.t.} \quad & \sum_{j \in J_k} \lambda_j = 1, \quad \lambda_j \geq 0 \qquad \text{for all } j \in J_k \end{aligned}$$

# The Hadamard Subproblem

Let $k \in \mathbb{N}$ and $j \in \{1, \dots, k\} = J_k$ and let $\lambda_j$ be convex coefficients. In the Hadamard case:

$$s^j \in \partial f(x_j) \Rightarrow \sum_{j=1}^{k} \lambda_j s^j \in \partial_\varepsilon f(x_k) \quad \text{if} \quad \varepsilon \geq \sum_{j=1}^{k} \lambda_j e_j^k$$

Then the search direction is

$$d^k = -\sum_{j=1}^{k} \lambda_j s^j$$

where the coefficients $\lambda_j$ are solution to

$$\begin{aligned} \operatorname*{arg\,min}_{\lambda \in \mathbb{R}^{|J_k|}} \quad & \frac{1}{2} \left\| \sum_{j \in J_k} \lambda_j s^j \right\|^2 + \sum_{j \in J_k} \lambda_j e_j^k \\ \text{s.t.} \quad & \sum_{j \in J_k} \lambda_j = 1, \quad \lambda_j \geq 0 \qquad \text{for all } j \in J_k \end{aligned}$$

# The Hadamard Subproblem

Let $k \in \mathbb{N}$ and $j \in \{1, \ldots, k\} = J_k$ and let $\lambda_j$ be convex coefficients. In the Hadamard case:

$$X_{q_j} \in \partial f(q_j) \Rightarrow \sum_{j=1}^{k} \lambda_j \mathrm{P}_{p_k \leftarrow q_j} X_{q_j} \in \partial_\varepsilon f(p_k) \quad \text{if} \quad \varepsilon \geq \sum_{j=1}^{k} \lambda_j c_j^k$$

Then the search direction is

$$d^k = -\sum_{j=1}^{k} \lambda_j s^j$$

where the coefficients $\lambda_j$ are solution to

$$\begin{aligned} \operatorname*{arg\,min}_{\lambda \in \mathbb{R}^{|J_k|}} \quad & \frac{1}{2} \left\| \sum_{j \in J_k} \lambda_j s^j \right\|^2 + \sum_{j \in J_k} \lambda_j e_j^k \\ \text{s.t.} \quad & \sum_{j \in J_k} \lambda_j = 1, \quad \lambda_j \geq 0 \qquad \text{for all } j \in J_k \end{aligned}$$

# The Hadamard Subproblem

Let $k \in \mathbb{N}$ and $j \in \{1, \dots, k\} = J_k$ and let $\lambda_j$ be convex coefficients. In the Hadamard case:

$$X_{q_j} \in \partial f(q_j) \Rightarrow \sum_{j=1}^{k} \lambda_j \mathrm{P}_{p_k \leftarrow q_j} X_{q_j} \in \partial_\varepsilon f(p_k) \quad \text{if} \quad \varepsilon \geq \sum_{j=1}^{k} \lambda_j c_j^k$$

Then the search direction is

$$d^k = -\sum_{j=1}^{k} \lambda_j \mathrm{P}_{p_k \leftarrow q_j} X_{q_j}$$

where the coefficients $\lambda_j$ are solution to

$$\begin{aligned} \operatorname*{arg\,min}_{\lambda \in \mathbb{R}^{|J_k|}} \quad & \frac{1}{2} \left\| \sum_{j \in J_k} \lambda_j s^j \right\|^2 + \sum_{j \in J_k} \lambda_j e_j^k \\ \text{s.t.} \quad & \sum_{j \in J_k} \lambda_j = 1, \quad \lambda_j \geq 0 \qquad \text{for all } j \in J_k \end{aligned}$$

# The Hadamard Subproblem

Let $k \in \mathbb{N}$ and $j \in \{1, \dots, k\} = J_k$ and let $\lambda_j$ be convex coefficients. In the Hadamard case:

$$X_{q_j} \in \partial f(q_j) \Rightarrow \sum_{j=1}^{k} \lambda_j \mathrm{P}_{p_k \leftarrow q_j} X_{q_j} \in \partial_\varepsilon f(p_k) \quad \text{if} \quad \varepsilon \geq \sum_{j=1}^{k} \lambda_j c_j^k$$

Then the search direction is

$$d^k = -\sum_{j=1}^{k} \lambda_j \mathrm{P}_{p_k \leftarrow q_j} X_{q_j}$$

where the coefficients $\lambda_j$ are solution to

$$\begin{aligned} \operatorname*{arg\,min}_{\lambda \in \mathbb{R}^{|J_k|}} \quad & \frac{1}{2} \left\| \sum_{j \in J_k} \lambda_j \mathrm{P}_{p_k \leftarrow q_j} X_{q_j} \right\|^2 + \sum_{j \in J_k} \lambda_j c_j^k \\ \text{s.t.} \quad & \sum_{j \in J_k} \lambda_j = 1, \quad \lambda_j \geq 0 \qquad \text{for all } j \in J_k \end{aligned}$$

# The Euclidean Bundle Method

**Data:** $x_1 = y_1 \in \mathbb{R}^n$, $g^0 = s^1 \in \partial f(x_1)$, $m \in (0,1)$, $\varepsilon^0 = e_1^1 = 0$, $k = 1$, $J_k = \{1\}$

Compute a solution $\lambda^k \in \mathbb{R}^{|J_k|}$ of the stabilizing subproblem and set

$$g^k = \sum_{j \in J_k} \lambda_j^k s^j, \quad \varepsilon^k = \sum_{j \in J_k} \lambda_j^k e_j^k, \quad d^k = -g^k, \quad \xi^k = -\|g^k\|^2 - \varepsilon^k$$

**while** *not converged* **do**

Set $y_{k+1} = x_k + d^k$ and compute $s^{k+1} \in \partial f(y_{k+1})$

If $f(y_{k+1}) \leq f(x_k) + m\xi^k$, then set $x_{k+1} = y_{k+1}$, else set $x_{k+1} = x_k$

Update the index set to $J_{k+1}$, set $k \leftarrow k+1$, and set

$$e_j^{k+1} = f(x_{k+1}) - f(y_j) - (s^j)^T (x_{k+1} - y_j) \text{ for all } j \in J_{k+1}$$

**end**

**Result:** $\{x_k\}_{k \in \mathbb{N}}$, $\{y_k\}_{k \in \mathbb{N}}$

# The Convex Hadamard Bundle Method

**Data:** $p_1 = q_1 \in \mathcal{M}$, $g^0 = X_1 \in \partial f(p_1)$, $m \in (0,1)$, $\varepsilon^0 = c_1^1 = 0$, $k = 1$, $J_k = \{1\}$

Compute a solution $\lambda^k \in \mathbb{R}^{|J_k|}$ of the stabilizing subproblem and set

$$g^k = \sum_{j \in J_k} \lambda_j^k s^j, \quad \varepsilon^k = \sum_{j \in J_k} \lambda_j^k e_j^k, \quad d^k = -g^k, \quad \xi^k = -\|g^k\|^2 - \varepsilon^k$$

**while** *not converged* **do**

Set $y_{k+1} = x_k + d^k$ and compute $s^{k+1} \in \partial f(y_{k+1})$

If $f(y_{k+1}) \leq f(x_k) + m\xi^k$, then set $x_{k+1} = y_{k+1}$, else set $x_{k+1} = x_k$

Update the index set to $J_{k+1}$, set $k \leftarrow k+1$, and set

$$e_j^{k+1} = f(x_{k+1}) - f(y_j) - (s^j)^T (x_{k+1} - y_j) \text{ for all } j \in J_{k+1}$$

**end**

**Result:** $\{x_k\}_{k \in \mathbb{N}}$, $\{y_k\}_{k \in \mathbb{N}}$

# The Convex Hadamard Bundle Method

**Data:** $p_1 = q_1 \in \mathcal{M}$, $g^0 = X_1 \in \partial f(p_1)$, $m \in (0,1)$, $\varepsilon^0 = c_1^1 = 0$, $k = 1$, $J_k = \{1\}$

Compute a solution $\lambda^k \in \mathbb{R}^{|J_k|}$ of the stabilizing subproblem and set

$$g^k = \sum_{j \in J_k} \lambda_j^k \mathrm{P}_{p_k \leftarrow q_j} X_{q_j}, \quad \varepsilon^k = \sum_{j \in J_k} \lambda_j^k c_j^k, \quad d^k = -g^k, \quad \xi^k = -\|g^k\|^2 - \varepsilon^k$$

**while** *not converged* **do**

Set $y_{k+1} = x_k + d^k$ and compute $s^{k+1} \in \partial f(y_{k+1})$

If $f(y_{k+1}) \leq f(x_k) + m\xi^k$, then set $x_{k+1} = y_{k+1}$, else set $x_{k+1} = x_k$

Update the index set to $J_{k+1}$, set $k \leftarrow k+1$, and set

$$e_j^{k+1} = f(x_{k+1}) - f(y_j) - (s^j)^T (x_{k+1} - y_j) \text{ for all } j \in J_{k+1}$$

**end**

**Result:** $\{x_k\}_{k \in \mathbb{N}}$, $\{y_k\}_{k \in \mathbb{N}}$

# The Convex Hadamard Bundle Method

**Data:** $p_1 = q_1 \in \mathcal{M}$, $g^0 = X_1 \in \partial f(p_1)$, $m \in (0,1)$, $\varepsilon^0 = c_1^1 = 0$, $k = 1$, $J_k = \{1\}$

Compute a solution $\lambda^k \in \mathbb{R}^{|J_k|}$ of the stabilizing subproblem and set

$$g^k = \sum_{j \in J_k} \lambda_j^k \mathrm{P}_{p_k \leftarrow q_j} X_{q_j}, \quad \varepsilon^k = \sum_{j \in J_k} \lambda_j^k c_j^k, \quad d^k = -g^k, \quad \xi^k = -\|g^k\|^2 - \varepsilon^k$$

**while** *not converged* **do**

Set $q_{k+1} = \exp_{p_k}(d^k)$ and compute $X_{q_{k+1}} \in \partial f(q_{k+1})$

If $f(y_{k+1}) \leq f(x_k) + m\xi^k$, then set $x_{k+1} = y_{k+1}$, else set $x_{k+1} = x_k$

Update the index set to $J_{k+1}$, set $k \leftarrow k+1$, and set

$$e_j^{k+1} = f(x_{k+1}) - f(y_j) - (s^j)^T (x_{k+1} - y_j) \text{ for all } j \in J_{k+1}$$

**end**

**Result:** $\{x_k\}_{k \in \mathbb{N}}$, $\{y_k\}_{k \in \mathbb{N}}$

# The Convex Hadamard Bundle Method

**Data:** $p_1 = q_1 \in \mathcal{M}$, $g^0 = X_1 \in \partial f(p_1)$, $m \in (0,1)$, $\varepsilon^0 = c_1^1 = 0$, $k = 1$, $J_k = \{1\}$

Compute a solution $\lambda^k \in \mathbb{R}^{|J_k|}$ of the stabilizing subproblem and set

$$g^k = \sum_{j \in J_k} \lambda_j^k \mathrm{P}_{p_k \leftarrow q_j} X_{q_j}, \quad \varepsilon^k = \sum_{j \in J_k} \lambda_j^k c_j^k, \quad d^k = -g^k, \quad \xi^k = -\|g^k\|^2 - \varepsilon^k$$

**while** *not converged* **do**

  Set $q_{k+1} = \exp_{p_k}(d^k)$ and compute $X_{q_{k+1}} \in \partial f(q_{k+1})$

  If $f(q_{k+1}) \leq f(p_k) + m\xi^k$, then set $p_{k+1} = q_{k+1}$, else set $p_{k+1} = p_k$

  Update the index set to $J_{k+1}$, set $k \leftarrow k+1$, and set

$$e_j^{k+1} = f(x_{k+1}) - f(y_j) - (s^j)^T (x_{k+1} - y_j) \text{ for all } j \in J_{k+1}$$

**end**

**Result:** $\{x_k\}_{k \in \mathbb{N}}$, $\{y_k\}_{k \in \mathbb{N}}$

# The Convex Hadamard Bundle Method

**Data:** $p_1 = q_1 \in \mathcal{M}$, $g^0 = X_1 \in \partial f(p_1)$, $m \in (0,1)$, $\varepsilon^0 = c_1^1 = 0$, $k = 1$, $J_k = \{1\}$

Compute a solution $\lambda^k \in \mathbb{R}^{|J_k|}$ of the stabilizing subproblem and set

$$g^k = \sum_{j \in J_k} \lambda_j^k \mathrm{P}_{p_k \leftarrow q_j} X_{q_j}, \quad \varepsilon^k = \sum_{j \in J_k} \lambda_j^k c_j^k, \quad d^k = -g^k, \quad \xi^k = -\|g^k\|^2 - \varepsilon^k$$

**while** *not converged* **do**

Set $q_{k+1} = \exp_{p_k}(d^k)$ and compute $X_{q_{k+1}} \in \partial f(q_{k+1})$

If $f(q_{k+1}) \leq f(p_k) + m\xi^k$, then set $p_{k+1} = q_{k+1}$, else set $p_{k+1} = p_k$

Update the index set to $J_{k+1}$, set $k \leftarrow k+1$, and set

$$c_j^{k+1} = f(p_{k+1}) - f(q_j) - \langle X_{q_j}, \log_{q_j} p_{k+1} \rangle \text{ for all } j \in J_{k+1}$$

**end**

**Result:** $\{x_k\}_{k \in \mathbb{N}}$, $\{y_k\}_{k \in \mathbb{N}}$

# The Convex Hadamard Bundle Method

**Data:** $p_1 = q_1 \in \mathcal{M}$, $g^0 = X_1 \in \partial f(p_1)$, $m \in (0,1)$, $\varepsilon^0 = c_1^1 = 0$, $k = 1$, $J_k = \{1\}$

Compute a solution $\lambda^k \in \mathbb{R}^{|J_k|}$ of the stabilizing subproblem and set

$$g^k = \sum_{j \in J_k} \lambda_j^k \mathrm{P}_{p_k \leftarrow q_j} X_{q_j}, \quad \varepsilon^k = \sum_{j \in J_k} \lambda_j^k c_j^k, \quad d^k = -g^k, \quad \xi^k = -\|g^k\|^2 - \varepsilon^k$$

**while** *not converged* **do**

Set $q_{k+1} = \exp_{p_k}(d^k)$ and compute $X_{q_{k+1}} \in \partial f(q_{k+1})$

If $f(q_{k+1}) \leq f(p_k) + m\xi^k$, then set $p_{k+1} = q_{k+1}$, else set $p_{k+1} = p_k$

Update the index set to $J_{k+1}$, set $k \leftarrow k+1$, and set

$$c_j^{k+1} = f(p_{k+1}) - f(q_j) - \langle X_{q_j}, \log_{q_j} p_{k+1} \rangle + \varrho \sqrt{\|\log_{q_j} p_{k+1}\|} \|X_{q_j}\| \text{ for all } j \in J_{k+1}$$

**end**

**Result:** $\{x_k\}_{k \in \mathbb{N}}$, $\{y_k\}_{k \in \mathbb{N}}$

# The Convex Hadamard Bundle Method

**Data:** $p_1 = q_1 \in \mathcal{M}$, $g^0 = X_1 \in \partial f(p_1)$, $m \in (0,1)$, $\varepsilon^0 = c_1^1 = 0$, $k = 1$, $J_k = \{1\}$

Compute a solution $\lambda^k \in \mathbb{R}^{|J_k|}$ of the stabilizing subproblem and set

$$g^k = \sum_{j \in J_k} \lambda_j^k \mathrm{P}_{p_k \leftarrow q_j} X_{q_j}, \quad \varepsilon^k = \sum_{j \in J_k} \lambda_j^k c_j^k, \quad d^k = -g^k, \quad \xi^k = -\|g^k\|^2 - \varepsilon^k$$

**while** *not converged* **do**

Set $q_{k+1} = \exp_{p_k}(d^k)$ and compute $X_{q_{k+1}} \in \partial f(q_{k+1})$

If $f(q_{k+1}) \leq f(p_k) + m\xi^k$, then set $p_{k+1} = q_{k+1}$, else set $p_{k+1} = p_k$

Update the index set to $J_{k+1}$, set $k \leftarrow k+1$, and set

$$c_j^{k+1} = f(p_{k+1}) - f(q_j) - \langle X_{q_j}, \log_{q_j} p_{k+1} \rangle + \varrho \sqrt{\|\log_{q_j} p_{k+1}\|} \|X_{q_j}\| \text{ for all } j \in J_{k+1}$$

**end**

**Result:** $\{p_k\}_{k \in \mathbb{N}}$, $\{q_k\}_{k \in \mathbb{N}}$

# Convergence

- In the Euclidean case we have Geiger and Kanzow 2002, Theorem 6.80

**Theorem**
*Let the solution set* $S = \{x_* \in \mathbb{R}^n \mid f(x_*) = \inf_{x \in \mathbb{R}^n} f(x)\}$ *of the minimization problem be nonempty. Then every sequence* $\{x_k\}$ *generated by the bundle method algorithm converges to a minimum of* $f$*.*

# Convergence

- In the Euclidean case we have Geiger and Kanzow 2002, Theorem 6.80

**Theorem**
*Let the solution set* $S = \{x_* \in \mathbb{R}^n \mid f(x_*) = \inf_{x \in \mathbb{R}^n} f(x)\}$ *of the minimization problem be nonempty. Then every sequence* $\{x_k\}$ *generated by the bundle method algorithm converges to a minimum of* $f$.

- In the Hadamard case, we obtain an analogous result

# Convergence

- In the Euclidean case we have Geiger and Kanzow 2002, Theorem 6.80

**Theorem**
*Let the solution set* $S = \{x_* \in \mathbb{R}^n \mid f(x_*) = \inf_{x \in \mathbb{R}^n} f(x)\}$ *of the minimization problem be nonempty. Then every sequence* $\{x_k\}$ *generated by the bundle method algorithm converges to a minimum of* $f$.

- In the Hadamard case, we obtain an analogous result
  - achieved by enlarging the $\varepsilon-$subdifferential

# Convergence

- In the Euclidean case we have Geiger and Kanzow 2002, Theorem 6.80

### Theorem

*Let the solution set $S = \{x_* \in \mathbb{R}^n \mid f(x_*) = \inf_{x \in \mathbb{R}^n} f(x)\}$ of the minimization problem be nonempty. Then every sequence $\{x_k\}$ generated by the bundle method algorithm converges to a minimum of $f$.*

- In the Hadamard case, we obtain an analogous result
  - achieved by enlarging the $\varepsilon-$subdifferential
- Numerically, given $\mathrm{tol} > 0$, we employ the following stopping criterion

$$-\xi^k \leq \mathrm{tol}$$

# Implementation

The algorithm is implemented[1] in Julia using `Manopt.jl` (Bergmann 2022) which uses manifolds from `Manifolds.jl` (Axen et al. 2021). A solver call just looks like

```
p_star = bundle_method(M, f, ∂f, p0)
```

where

- `M` is a Hadamard manifold
- `f` is the objective function
- `∂f` is a subgradient of the objective function
- `p0` is an initial point on the manifold
- the parameter for the descent test is set at a default $m = 0.0125$

[1] not yet in a release version

# Numerical Example: $\mathcal{H}^4$ and $\mathcal{P}(3)$

Let $\mathcal{M}_1 = \mathcal{H}^4$ be the four-dimensional hyperbolic space, and let $\mathcal{M}_2 = \mathcal{P}(3)$ be the space of $3 \times 3$ symmetric positive definite matrices. Let $q_1, \ldots, q_n \in \mathcal{M}_j$ be $n = 100$ random data points for $j \in \{1, 2\}$.

# Numerical Example: $\mathcal{H}^4$ and $\mathcal{P}(3)$

Let $\mathcal{M}_1 = \mathcal{H}^4$ be the four-dimensional hyperbolic space, and let $\mathcal{M}_2 = \mathcal{P}(3)$ be the space of $3 \times 3$ symmetric positive definite matrices. Let $q_1, \ldots, q_n \in \mathcal{M}_j$ be $n = 100$ random data points for $j \in \{1, 2\}$. Let $f_{ij} \colon \mathcal{M}_j \to \mathbb{R}$ be defined by

$$f_{ij}(p) = \sum_{k=1}^{n} \operatorname{dist}^i_{\mathcal{M}_j}(p, q_k)$$

# Numerical Example: $\mathcal{H}^4$ and $\mathcal{P}(3)$

Let $\mathcal{M}_1 = \mathcal{H}^4$ be the four-dimensional hyperbolic space, and let $\mathcal{M}_2 = \mathcal{P}(3)$ be the space of $3 \times 3$ symmetric positive definite matrices. Let $q_1, \ldots, q_n \in \mathcal{M}_j$ be $n = 100$ random data points for $j \in \{1, 2\}$. Let $f_{ij} \colon \mathcal{M}_j \to \mathbb{R}$ be defined by

$$f_{ij}(p) = \sum_{k=1}^{n} \operatorname{dist}^i_{\mathcal{M}_j}(p, q_k)$$

and let $g_j \colon \mathcal{M}_j \to \mathbb{R}$ be defined by

$$g_j(p) = \max_{i \in \{1,2\}} \{f_{ij}(p)\}$$

for $i, j \in \{1, 2\}$.

NTNU

# Numerical Example: $\mathcal{H}^4$ and $\mathcal{P}(3)$

Let $\mathcal{M}_1 = \mathcal{H}^4$ be the four-dimensional hyperbolic space, and let $\mathcal{M}_2 = \mathcal{P}(3)$ be the space of $3 \times 3$ symmetric positive definite matrices. Let $q_1, \ldots, q_n \in \mathcal{M}_j$ be $n = 100$ random data points for $j \in \{1, 2\}$. Let $f_{ij} \colon \mathcal{M}_j \to \mathbb{R}$ be defined by

$$f_{ij}(p) = \sum_{k=1}^{n} \operatorname{dist}^i_{\mathcal{M}_j}(p, q_k)$$

and let $g_j \colon \mathcal{M}_j \to \mathbb{R}$ be defined by

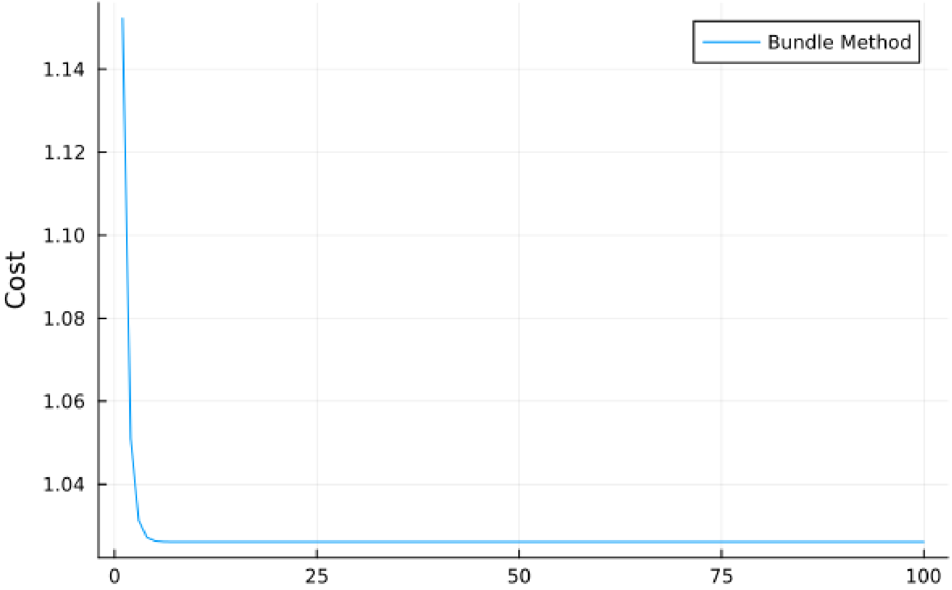
$$g_j(p) = \max_{i \in \{1,2\}} \{f_{ij}(p)\}$$

for $i, j \in \{1, 2\}$. We want to solve

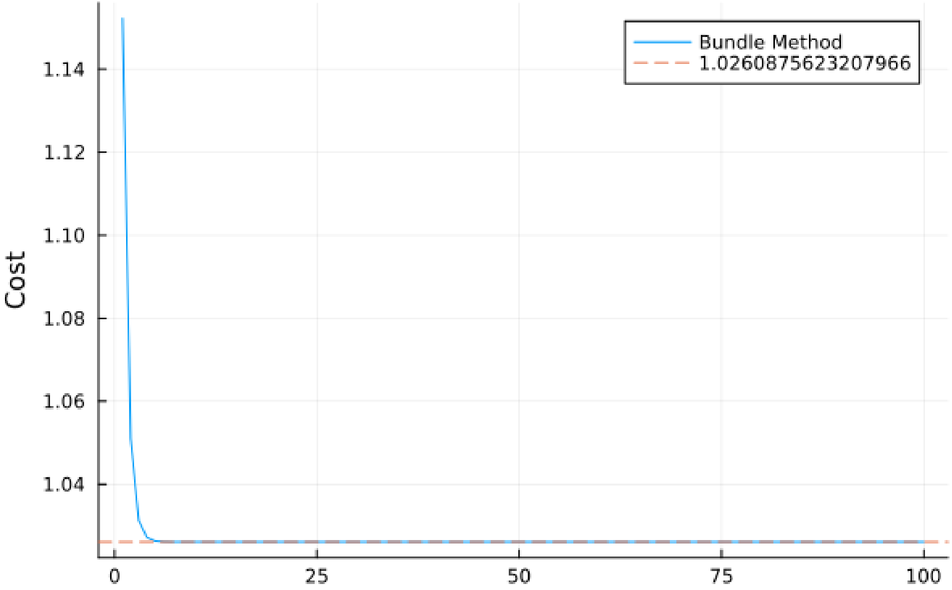
$$\operatorname*{arg\,min}_{p \in \mathcal{M}_j} g_j(p)$$

for $j \in \{1, 2\}$.

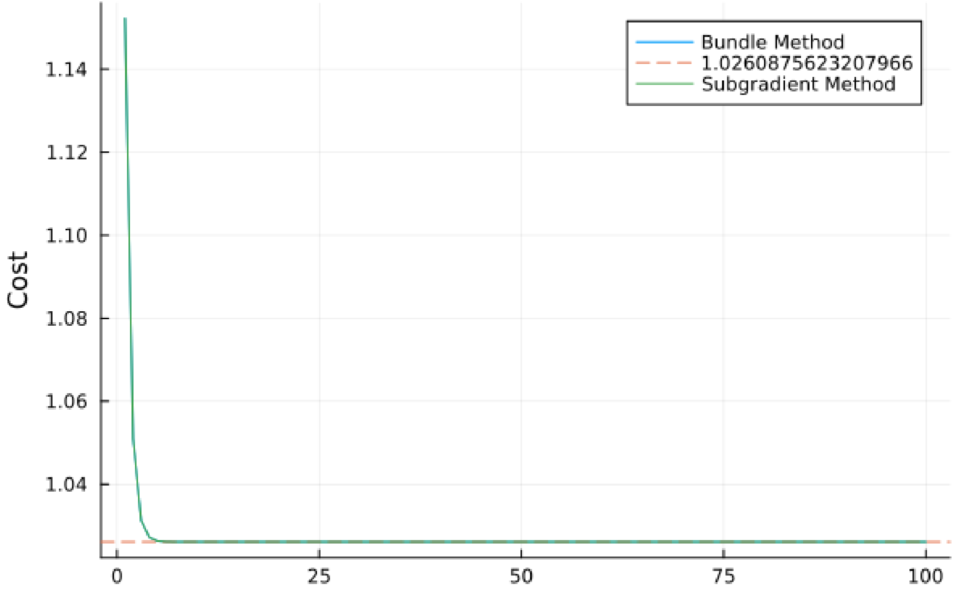
# Numerical Example: $\mathcal{H}^4$



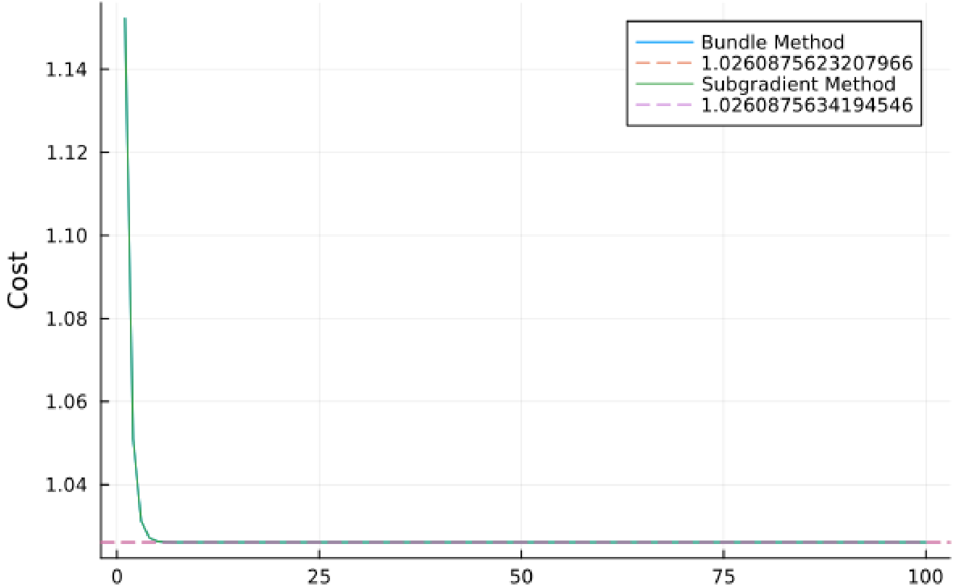
# Numerical Example: $\mathcal{H}^4$



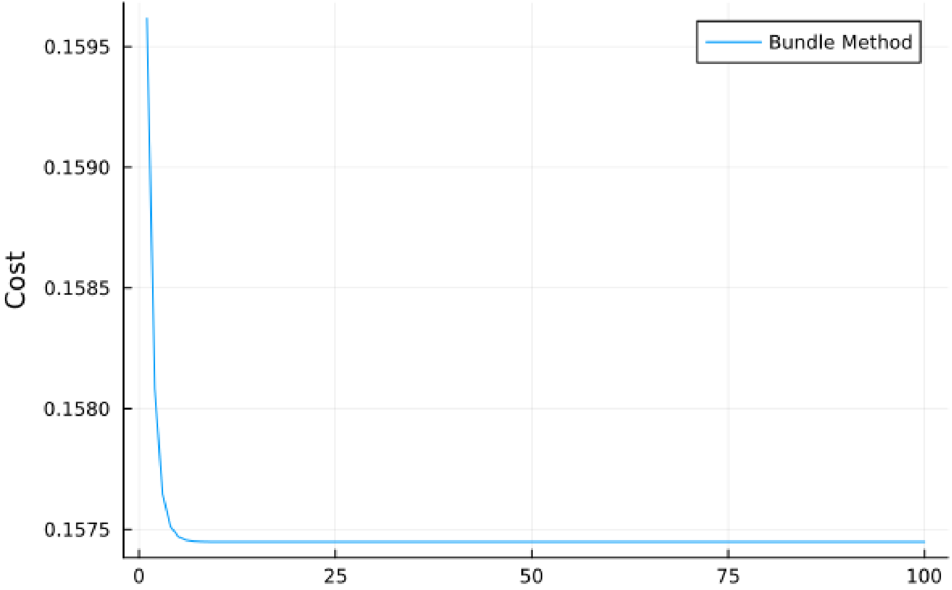
# Numerical Example: $\mathcal{H}^4$



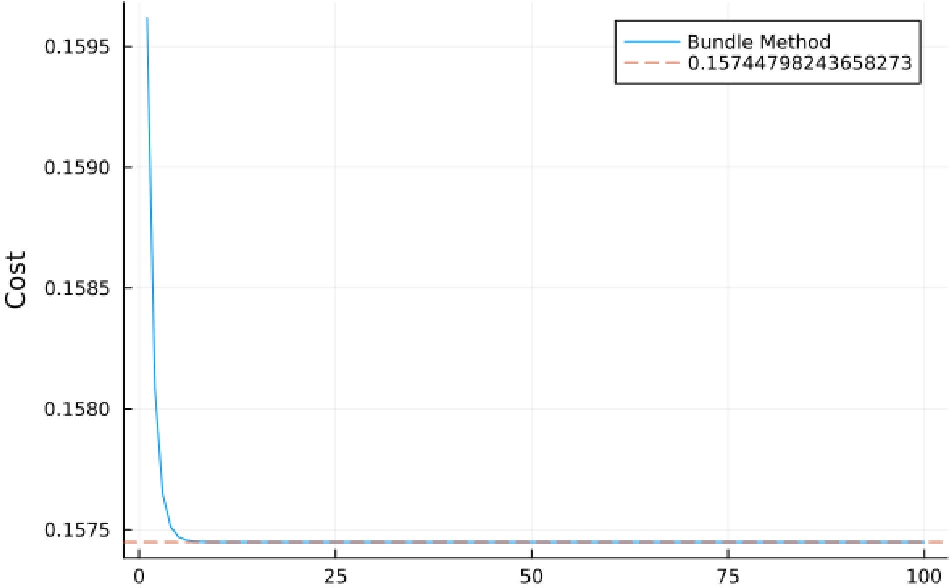
# Numerical Example: $\mathcal{H}^4$



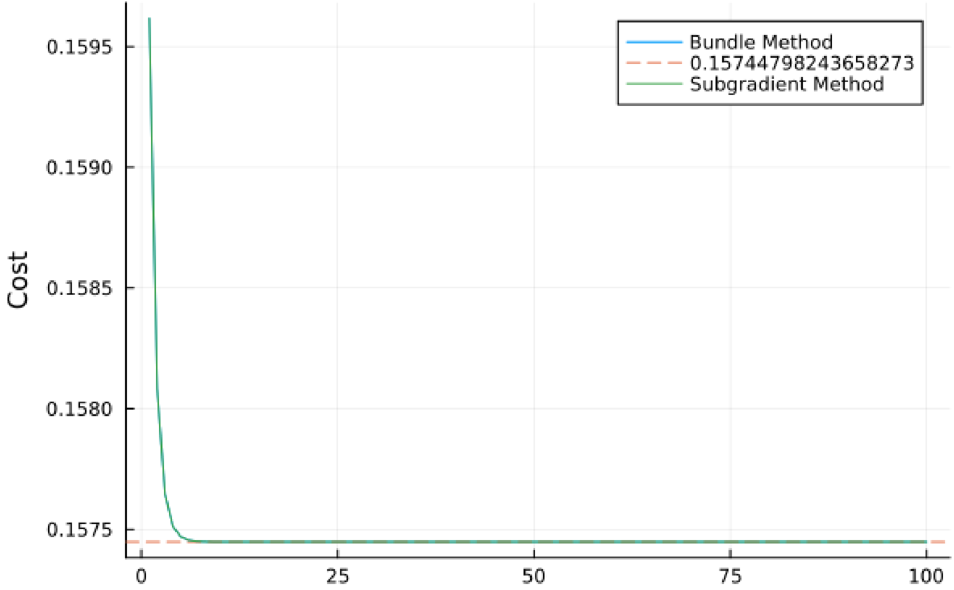
# Numerical Example: $\mathcal{P}(3)$



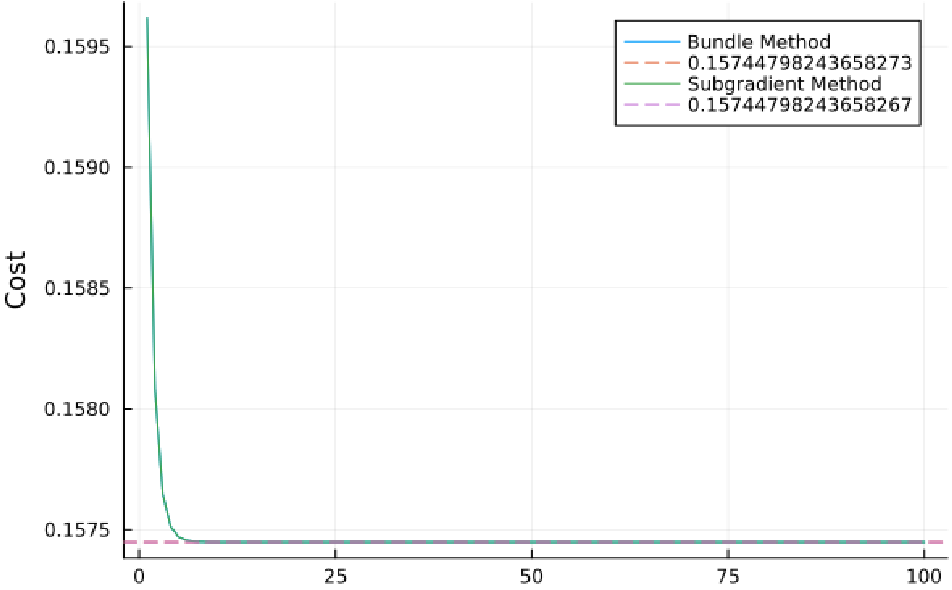
# Numerical Example: $\mathcal{P}(3)$



# Numerical Example: $\mathcal{P}(3)$



# Numerical Example: $\mathcal{P}(3)$



# Summary and Future Work

What we did:

- presented the bundle method for geodesically convex functions on Hadamard manifolds
- touched upon convergence
- showed two numerical examples

To do:

- further investigate convergence
- generic Riemannian manifolds

# Bibliography


- Axen, Seth D. et al. (2021). *Manifolds.jl: An Extensible Julia Framework for Data Analysis on Manifolds*. arXiv: 2106.08777.
- Azagra, Daniel and Juan Ferrera (2005). “Proximal calculus on Riemannian manifolds”. In: *Mediterranean Journal of Mathematics* 2.4, pp. 437–450.
- Bergmann, Ronny (2022). “Manopt.jl: Optimization on Manifolds in Julia”. In: *Journal of Open Source Software* 7.70, p. 3866. DOI: 10.21105/joss.03866.
- Bonnans, J.-F. et al. (2006). *Numerical optimization: theoretical and practical aspects*. Springer-Verlag.
- Geiger, C. and C. Kanzow (2002). *Theorie und Numerik restringierter Optimierungsaufgaben*. New York: Springer. DOI: 10.1007/978-3-642-56004-0.