Complexity of a Riemannian direct-search algorithm

Bastien Cavarretta, *F. Goyens , C. Royer and F. Yger*

EUROPT - June 30^{th}, 2025

Introduction

- **Derivative Free Optimization:** Only function values.
  Goal: use as little evaluations as possible $\rightarrow$ scales badly with dimension.

Introduction

- **Derivative Free Optimization:** Only function values.
  Goal: use as little evaluations as possible $\rightarrow$ scales badly with dimension.
- **Optimization on manifolds:** [Boumal, 2023], [Zhang and Sra, 2016]

$$\min_{x \in \mathcal{M}} f(x), \qquad \text{where } \mathcal{M} \text{ is a Riemannian manifold.}$$

Introduction

- **Derivative Free Optimization:** Only function values.
  Goal: use as little evaluations as possible $\rightarrow$ scales badly with dimension.
- **Optimization on manifolds:** [Boumal, 2023], [Zhang and Sra, 2016]

$$\min_{x \in \mathcal{M}} f(x), \qquad \text{where } \mathcal{M} \text{ is a Riemannian manifold.}$$

- V. Kungurtsev, F. Rinaldi, D. Zeffiro (2024), *Retraction-Based Direct Search Methods for Derivative-Free Riemannian Optimization*

Introduction

- **Derivative Free Optimization:** Only function values.
  Goal: use as little evaluations as possible $\rightarrow$ scales badly with dimension.
- **Optimization on manifolds:** [Boumal, 2023], [Zhang and Sra, 2016]

$$\min_{x \in \mathcal{M}} f(x), \qquad \text{where } \mathcal{M} \text{ is a Riemannian manifold.}$$

- V. Kungurtsev, F. Rinaldi, D. Zeffiro (2024), *Retraction-Based Direct Search Methods for Derivative-Free Riemannian Optimization*
- **Our work:**
  - Complexity bound of this Riemannian direct-search $\mathcal{O}(\epsilon^{-2})$
  - Compare performance of two versions (projection?)

Table of content

Table of content

Positive spanning sets (PSSs)

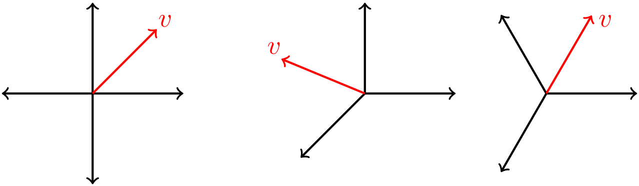
Definition

A *Positive Spanning Set* (PSS) of $\mathbb{R}^n$ is a set $D \subset \mathbb{R}^n$ s.t

$$\mathrm{cone}(D) = \mathbb{R}^n.$$

The *cosine measure* of D is

$$\mathrm{cm}(D) := \min_{v \in \mathbb{R}^n \setminus \{0\}} \max_{d \in D} \frac{\langle v, d \rangle}{\|v\| \|d\|}$$



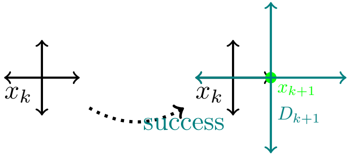
$\{I_2, -I_2\}$ $\{e_1, e_2, -e_1 - e_2\}$ unif. angles

PSS in $\mathbb{R}^2$.

Directional direct-search

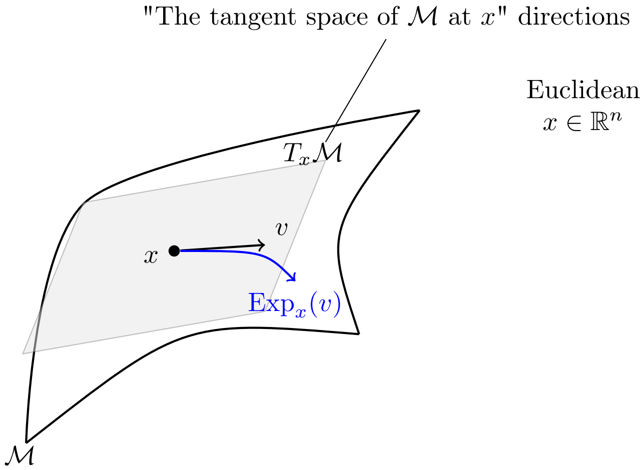
Start: $x_0 \in \mathbb{R}^n$. At each step k:

- Take a PSS $D_k \subset \mathbb{R}^n$
- If there exists $d \in D_k$ s.t. $f(x_k + \alpha_k d) \le f(x_k) - c\alpha_k^2 \|d\|^2$ ($\rightarrow$ *successful step*):
$$x_{k+1} = x_k + \alpha_k d, \ \alpha_{k+1} = \min(\Gamma\alpha_k, \alpha_{\max}).$$
- Else ($\rightarrow$ *unsuccessful step*):
$$x_{k+1} = x_k, \alpha_{k+1} = \gamma\alpha_k$$



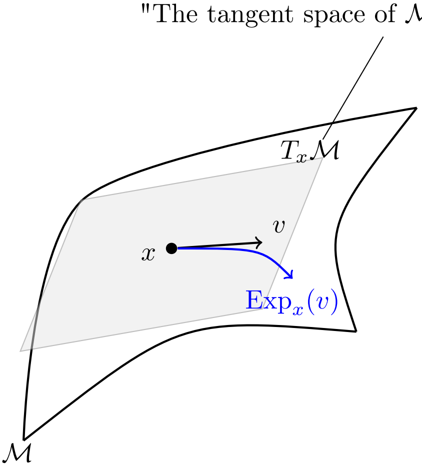
$\Gamma > 1 > \gamma > 0$ and $c > 0$

Moving along a manifold



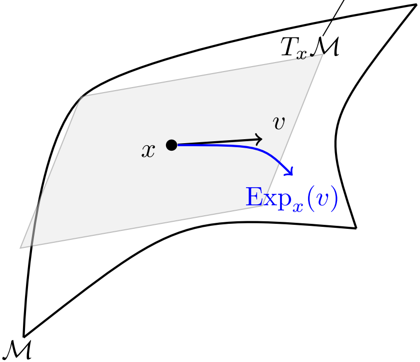
Euclidean	Riemannian
$x \in \mathbb{R}^n$	$x \in \mathcal{M}$

Moving along a manifold



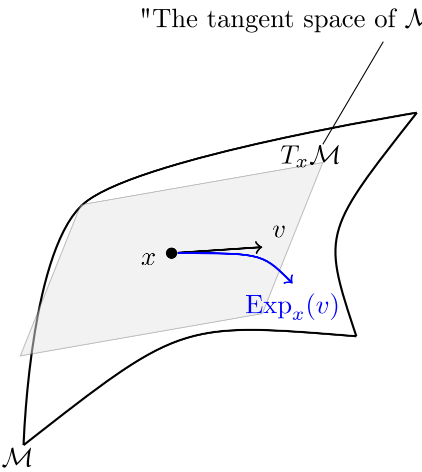
Euclidean	Riemannian
$x \in \mathbb{R}^n$	$x \in \mathcal{M}$
$v \in \mathbb{R}^n$	$v \in \mathrm{T}_x\mathcal{M}$

"The tangent space of $\mathcal{M}$ at x" directions



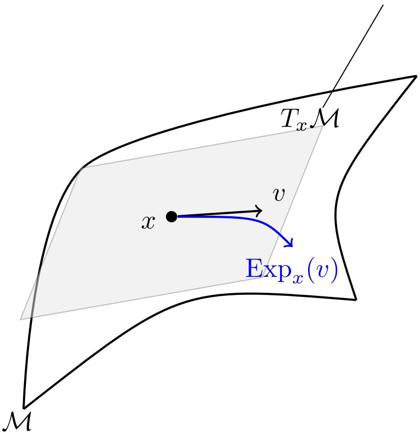
Euclidean	Riemannian
$x \in \mathbb{R}^n$	$x \in \mathcal{M}$
$v \in \mathbb{R}^n$	$v \in \mathrm{T}_x\mathcal{M}$
$\langle u, v\rangle = u^{\mathrm{T}} v$	$\langle u, v\rangle_x$

Moving along a manifold



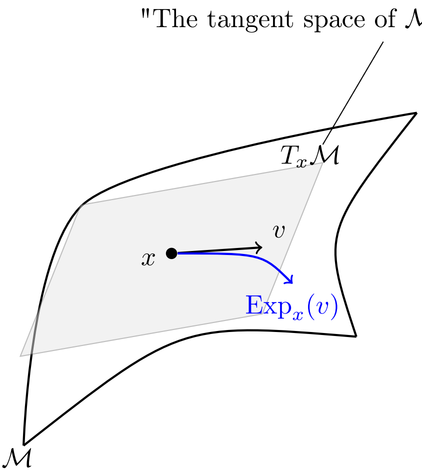
Euclidean	Riemannian
$x \in \mathbb{R}^n$	$x \in \mathcal{M}$
$v \in \mathbb{R}^n$	$v \in \mathrm{T}_x\mathcal{M}$
$\langle u, v \rangle = u^{\mathrm{T}} v$	$\langle u, v \rangle_x$
$y = x + v$	$y = \mathrm{Exp}_x(v)$

Moving along a manifold



Euclidean	Riemannian
$x \in \mathbb{R}^n$	$x \in \mathcal{M}$
$v \in \mathbb{R}^n$	$v \in \mathrm{T}_x\mathcal{M}$
$\langle u, v \rangle = u^{\mathrm{T}} v$	$\langle u, v \rangle_x$
$y = x + v$	$y = \mathrm{Exp}_x(v)$
$f : \mathbb{R}^n \to \mathbb{R}$	$f : \mathcal{M} \to \mathbb{R}$

Moving along a manifold



Euclidean	Riemannian
$x \in \mathbb{R}^n$	$x \in \mathcal{M}$
$v \in \mathbb{R}^n$	$v \in \mathrm{T}_x\mathcal{M}$
$\langle u, v \rangle = u^{\mathrm{T}} v$	$\langle u, v \rangle_x$
$y = x + v$	$y = \mathrm{Exp}_x(v)$
$f : \mathbb{R}^n \to \mathbb{R}$	$f : \mathcal{M} \to \mathbb{R}$
$\nabla f(x) \in \mathbb{R}^n$	$\mathrm{grad} f(x) \in \mathrm{T}_x\mathcal{M}$

Positive spanning sets (PSSs) in $\mathrm{T}_x\mathcal{M}$

Definition

A *Positive Spanning Set* (PSS) of $\mathbb{R}^n$ is a set $D \subset \mathbb{R}^n$ s.t

$$\operatorname{cone}(D) = \mathbb{R}^n.$$

The *cosine measure* of D is

$$\operatorname{cm}(D) := \min_{v \in \mathbb{R}^n \setminus \{0\}} \max_{d \in D} \frac{\langle v, d \rangle}{\|v\| \|d\|}$$

Positive spanning sets (PSSs) in $\mathrm{T}_x\mathcal{M}$

Definition

A *Positive Spanning Set* (PSS) of $\mathrm{T}_x\mathcal{M}$ is a set $D \subset \mathrm{T}_x\mathcal{M}$ s.t

$$\mathrm{cone}(D) = \mathrm{T}_x\mathcal{M}.$$

The *cosine measure* of D is

$$\mathrm{cm}(D) := \min_{v \in \mathrm{T}_x\mathcal{M} \setminus \{0\}} \max_{d \in D} \frac{\langle v, d \rangle_x}{\|v\| \|d\|}$$

Riemannian directional direct-search [1]

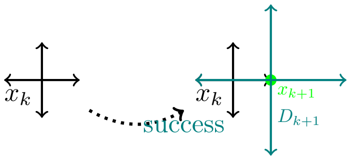
Start: $x_0 \in \mathbb{R}^n$. At each step k:

- Take a PSS $D_k \subset \mathbb{R}^n$
- If there exists $d \in D_k$ s.t. $f(x_k + \alpha_k d) \leq f(x_k) - c\alpha_k^2 \|d\|^2$ ($\rightarrow$ *successful step*):

$$x_{k+1} = x_k + \alpha_k d, \ \alpha_{k+1} = \min(\Gamma \alpha_k, \alpha_{\max}).$$

- Else ($\rightarrow$ *unsuccessful step*):

$$x_{k+1} = x_k, \alpha_{k+1} = \gamma \alpha_k$$



$\Gamma > 1 > \gamma > 0$ and $c > 0$

[1] [Kungurstev et al., 2024]

Riemannian directional direct-search [1]

Start: $x_0 \in \mathcal{M}$. At each step k:

- Take a PSS $D_k \subset \mathrm{T}_{x_k}\mathcal{M}$
- If there exists $d \in D_k$ s.t. $f(\mathrm{Exp}_{x_k}(\alpha_k d)) \leq f(x_k) - c\alpha_k^2 \|d\|_{x_k}^2$ ($\rightarrow$ *successful step*):

$$x_{k+1} = \mathrm{Exp}_{x_k}(\alpha_k d),\ \alpha_{k+1} = \min(\Gamma\alpha_k, \alpha_{\max}).$$

- Else ($\rightarrow$ *unsuccessful step*):

$$x_{k+1} = x_k, \alpha_{k+1} = \gamma\alpha_k$$

$\Gamma > 1 > \gamma > 0$ and $c > 0$

[1] [Kungurstev et al., 2024]

Riemannian directional direct-search [1]

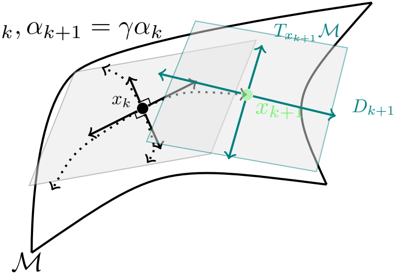
Start: $x_0 \in \mathcal{M}$. At each step k:

- Take a PSS $D_k \subset \mathrm{T}_{x_k}\mathcal{M}$
- If there exists $d \in D_k$ s.t. $f(\mathrm{Exp}_{x_k}(\alpha_k d)) \leq f(x_k) - c\alpha_k^2 \|d\|_{x_k}^2$ ($\rightarrow$ *successful step*):

$$x_{k+1} = \mathrm{Exp}_{x_k}(\alpha_k d),\ \alpha_{k+1} = \min(\Gamma\alpha_k, \alpha_{\max}).$$

- Else ($\rightarrow$ *unsuccessful step*):

$$x_{k+1} = x_k, \alpha_{k+1} = \gamma\alpha_k$$



$\Gamma > 1 > \gamma > 0$ and $c > 0$

[1][Kungurstev et al., 2024]

Table of content

Assumptions:

1. $f \geq f_{low}$ and
$$Lev_{\leq f(x_0)} := \{x \in \mathcal{M} \mid f(x) \leq f(x_0)\}$$
is bounded
2. For all k, for all $d \in D_k$:
$$\mathrm{cm}_{\mathrm{T}_{x_k}\mathcal{M}}(D_k) \geq \kappa,$$
$$d_{\min} \leq \|d\|_{x_k} \leq d_{\max}.$$
3. $\mathcal{M}$ connected Riemannian manifold with $\mathrm{inj}(\mathcal{M}) > d_{\max}\alpha_{max}$
4. f is geodesically L_g-smooth.

Complexity bound

Goal: $\|\text{grad} f(x_k)\|_{x_k} \leq \epsilon$ with k small.

Theorem

Reach $\|\text{grad} f\| \leq \epsilon$ *in* [2]

$$\mathcal{O}(\kappa^{-2}\epsilon^{-2}) \qquad \textit{iterations}$$

$$\mathcal{O}(\ell\kappa^{-2}\epsilon^{-2}) \qquad \textit{function evaluations}$$

where

- κ*: a lower bound on* $\text{cm}_{\mathrm{T}_{x_k}\mathcal{M}}(D_k) := \min_{v \in \mathrm{T}_{x_k}\mathcal{M}\setminus\{0\}} \max_{d \in D} \dfrac{\langle v, d\rangle}{\|v\|\|d\|}$.
- ℓ*: upper bound on the cardinal* $|D_k|$ *of* D_k.

[2] where constants depend on $f(x_0), f_{low}, d_{\min}, d_{\max}, c, \gamma, \Gamma, L_g, \alpha_0, \alpha_{\max}$

Table of content

Premliminary lemma

Lemma ([Kungurstev et al., 2024])

If iteration k is unsuccessful then:

$$\|\mathrm{grad} f(x_k)\|_{x_k} \leq \frac{d_{\max}(c + L_g/2)}{\kappa} \alpha_k$$

Premliminary lemma

Lemma ([Kungurstev et al., 2024])

If iteration k is unsuccessful then:

$$\|\mathrm{grad} f(x_k)\|_{x_k} \leq \frac{d_{\max}(c + L_g/2)}{\kappa}\alpha_k$$

Sketch of proof

$$\begin{aligned}\langle -\mathrm{grad} f(x_k), \alpha_k d\rangle_{x_k} - \frac{L_g}{2}\|\alpha_k d\|_{x_k}^2 &\overbrace{\leq}^{L_g\text{-smoothness}} f(x_k) - f(\mathrm{Exp}_{x_k}(\alpha_k d)) \\ &\underbrace{\leq}_{\text{failure}} c\alpha_k^2 \|d\|_{x_k}^2 .\end{aligned}$$

Premliminary lemma

Lemma ([Kungurstev et al., 2024])

If iteration k is unsuccessful then:

$$\|\text{grad} f(x_k)\|_{x_k} \leq \frac{d_{\max}(c + L_g/2)}{\kappa} \alpha_k$$

Sketch of proof

$$\begin{aligned}\langle -\text{grad} f(x_k), \alpha_k d\rangle_{x_k} - \frac{L_g}{2} \|\alpha_k d\|_{x_k}^2 &\overbrace{\leq}^{L_g\text{-smoothness}} f(x_k) - f(\text{Exp}_{x_k}(\alpha_k d)) \\ &\underbrace{\leq}_{\text{failure}} c\alpha_k^2 \|d\|_{x_k}^2 .\end{aligned}$$

$$\max_{d \in D_k} \frac{\langle -\text{grad} f(x_k), d\rangle_{x_k}}{\|d\|_{x_k}} \leq \left(c + \frac{L_g}{2}\right) \alpha_k d_{\max}$$

Premliminary lemma

Lemma ([Kungurstev et al., 2024])

If iteration k is unsuccessful then:

$$\|\mathrm{grad} f(x_k)\|_{x_k} \leq \frac{d_{\max}(c + L_g/2)}{\kappa}\alpha_k$$

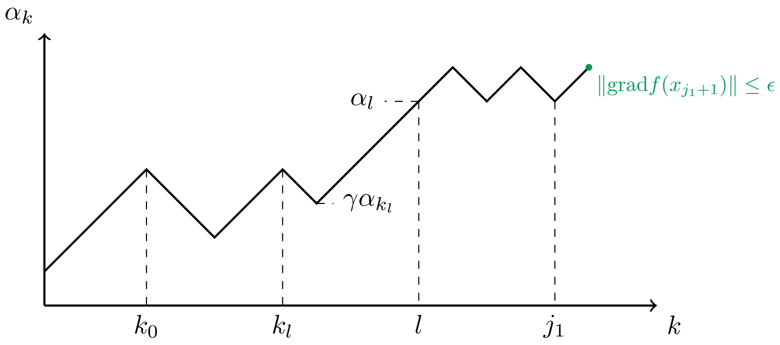
Sketch of proof

$$\begin{aligned}\langle -\mathrm{grad} f(x_k), \alpha_k d\rangle_{x_k} - \frac{L_g}{2}\|\alpha_k d\|_{x_k}^2 &\overbrace{\leq}^{L_g\text{-smoothness}} f(x_k) - f(\mathrm{Exp}_{x_k}(\alpha_k d)) \\ &\underbrace{\leq}_{\text{failure}} c\alpha_k^2\|d\|_{x_k}^2 .\end{aligned}$$

$$\kappa\|\mathrm{grad} f(x_k)\|_{x_k} \leq \max_{d \in D_k} \frac{\langle -\mathrm{grad} f(x_k), d\rangle_{x_k}}{\|d\|_{x_k}} \leq \left(c + \frac{L_g}{2}\right)\alpha_k d_{\max}$$

Notations

k_0	index of first failure
j_1	first index after k_0 s.t. $\\|\mathrm{grad} f(x_{j_1+1})\\| \leq \epsilon$
$\lvert \mathcal{U}_{k_0}^{j_1} \rvert$	Number of failures in $[\![k_0, j_1]\!]$
$\lvert \mathcal{S}_{k_0}^{j_1} \rvert$	Number of successes in $[\![k_0, j_1]\!]$
k_l	last index of failure before index l



Upper bounds [Vicente, 2013]

Lemma

- *The number of successes* $|\mathcal{S}_{k_0}^{j_1}|$ *is upper bounded:*

$$|\mathcal{S}_{k_0}^{j_1}| \leq \mathbf{cst}\epsilon^{-2}$$

- *The number of failures* $|\mathcal{U}_{k_0}^{j_1}|$ *is upper bounded:*

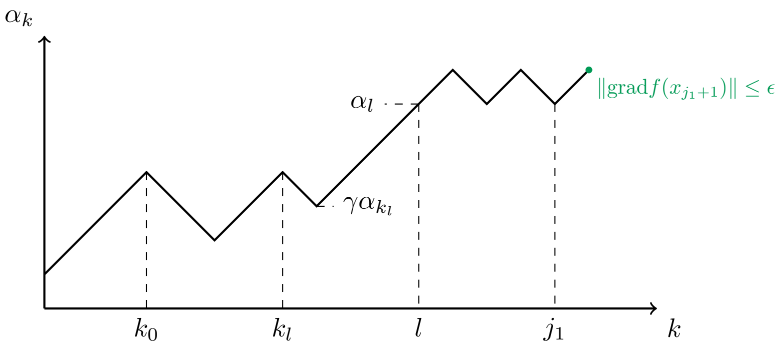
$$|\mathcal{U}_{k_0}^{j_1}| \leq \mathbf{cst}|\mathcal{S}_{k_0}^{j_1}| + \mathbf{cst}\log(\epsilon) + \mathbf{cst}$$

- *The index of first failure* k_0 *is upper bounded:*

$$k_0 \leq \mathbf{cst}$$

Notations

k_0	index of first failure
j_1	first index after k_0 s.t. $\|\mathrm{grad} f(x_{j_1+1})\| \leq \epsilon$
$\lvert\mathcal{U}_{k_0}^{j_1}\rvert$	Number of failures in $[\![k_0, j_1]\!]$
$\lvert\mathcal{S}_{k_0}^{j_1}\rvert$	Number of successes in $[\![k_0, j_1]\!]$
k_l	last index of failure before index l



Proof on the upper bound of #success

Goal: $$|\mathcal{S}_{\mathbf{k_0}}^{\mathbf{j_1}}| \leq \mathbf{cst}\epsilon^{-\mathbf{2}}$$

Proof on the upper bound of #success

Goal: $$|\mathcal{S}_{\mathbf{k_0}}^{\mathbf{j_1}}| \leq \mathbf{cst}\epsilon^{-\mathbf{2}}$$

Before j_1 $$\|\mathrm{grad} f(x_l)\|_{x_l} > \epsilon$$

Proof on the upper bound of #success

Goal: $$|\mathcal{S}_{\mathbf{k_0}}^{\mathbf{j_1}}| \leq \mathbf{cst}\epsilon^{-\mathbf{2}}$$

Before j_1 $$\|\mathrm{grad} f(x_l)\|_{x_l} > \epsilon$$

$l =$ success, k_l previous failure $$\alpha_{k_l} \geq \mathbf{cst} \, \|\mathrm{grad} f(x_{k_l})\|_{x_{k_l}}$$

Proof on the upper bound of #success

Goal: $|\mathcal{S}_{\mathbf{k_0}}^{\mathbf{j_1}}| \leq \mathbf{cst}\epsilon^{\mathbf{-2}}$

Before j_1 $\|\mathrm{grad} f(x_l)\|_{x_l} > \epsilon$

l = success, k_l previous failure $\alpha_{k_l} \geq \mathbf{cst}\, \|\mathrm{grad} f(x_{k_l})\|_{x_{k_l}}$

Only successes until l $\alpha_l \geq \gamma \alpha_{k_l}$

Proof on the upper bound of #success

Goal: $|\mathcal{S}_{\mathbf{k_0}}^{\mathbf{j_1}}| \leq \mathbf{cst}\epsilon^{-\mathbf{2}}$

Before j_1 $\|\mathrm{grad} f(x_l)\|_{x_l} > \epsilon$

$l =$ success, k_l previous failure $\alpha_{k_l} \geq \mathbf{cst} \, \|\mathrm{grad} f(x_{k_l})\|_{x_{k_l}}$

Only successes until l $\alpha_l \geq \gamma \alpha_{k_l}$

So by success $f(x_l) - f(x_{l+1}) \geq cst.\alpha_l^2 \geq \mathbf{cst}\epsilon^2$

Proof on the upper bound of #success

Goal:	$\lvert \mathcal{S}_{\mathbf{k_0}}^{\mathbf{j_1}} \rvert \leq \mathbf{cst}\epsilon^{-\mathbf{2}}$
Before j_1	$\lVert \mathrm{grad} f(x_l) \rVert_{x_l} > \epsilon$
$l =$ success, k_l previous failure	$\alpha_{k_l} \geq \mathbf{cst} \lVert \mathrm{grad} f(x_{k_l}) \rVert_{x_{k_l}}$
Only successes until l	$\alpha_l \geq \gamma \alpha_{k_l}$
So by success	$f(x_l) - f(x_{l+1}) \geq cst.\alpha_l^2 \geq \mathbf{cst}\epsilon^2$
Sum for all success in $[\![k_0, j_1]\!]$	$f(x_0) - f_{low} \geq \mathbf{cst}\epsilon^2 \lvert \mathcal{S}_{k_0}^{j_1} \rvert$

Table of content

Cardinal/cosine measure trade-off: Euclidean case

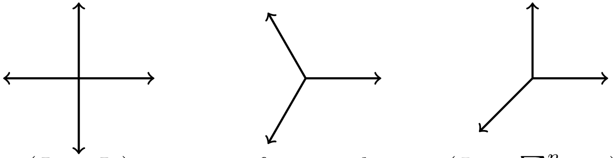
- **Complexity:** $\mathcal{O}(\ell\kappa^{-2}\epsilon^{-2})$

Cardinal/cosine measure trade-off: Euclidean case

- **Complexity:** $\mathcal{O}(\ell\kappa^{-2}\epsilon^{-2})$
- **Goal:** Low $\ell\kappa^{-2}$ $\rightarrow$ adapt PSS

Cardinal/cosine measure trade-off: Euclidean case

- **Complexity:** $\mathcal{O}(\ell\kappa^{-2}\epsilon^{-2})$
- **Goal:** Low $\ell\kappa^{-2}$ $\to$ adapt PSS

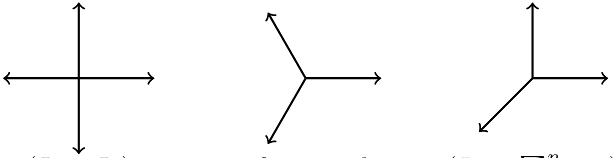


PSS	$(I_n, -I_n)$	uniform angles	$(I_n, -\sum_{i=1}^n e_i)$
size ℓ	$2n$	$n+1$	$n+1$
cm κ	$\frac{1}{\sqrt{n}}$	$\frac{1}{n}$	$\frac{1}{\sqrt{n^2+2(n-1)\sqrt{n}}}$
$\ell\kappa^{-2}$	$2n^2$	n^3+n^2	$n^3+2(n^2-1)\sqrt{n}+n^2$

Popular PSSs in $\mathbb{R}^n$

Cardinal/cosine measure trade-off: Euclidean case

- **Complexity:** $\mathcal{O}(\ell\kappa^{-2}\epsilon^{-2})$
- **Goal:** Low $\ell\kappa^{-2}$ $\rightarrow$ adapt PSS
- No absolute answer

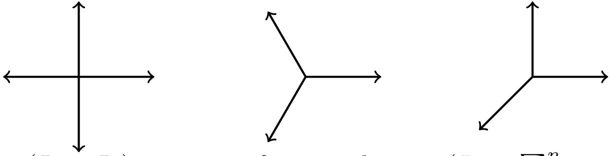


PSS	$(I_n, -I_n)$	uniform angles	$(I_n, -\sum_{i=1}^{n} e_i)$
size ℓ	$2n$	$n+1$	$n+1$
cm κ	$\frac{1}{\sqrt{n}}$	$\frac{1}{n}$	$\frac{1}{\sqrt{n^2+2(n-1)\sqrt{n}}}$
$\ell\kappa^{-2}$	$2n^2$	n^3+n^2	$n^3+2(n^2-1)\sqrt{n}+n^2$

Popular PSSs in $\mathbb{R}^n$

Cardinal/cosine measure trade-off: Euclidean case

- **Complexity:** $\mathcal{O}(\ell\kappa^{-2}\epsilon^{-2})$
- **Goal:** Low $\ell\kappa^{-2}$ → adapt PSS
- No absolute answer

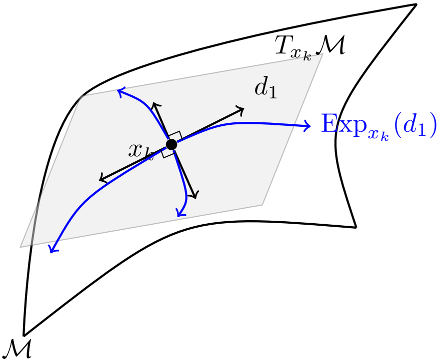


PSS	$(I_n, -I_n)$	uniform angles	$(I_n, -\sum_{i=1}^n e_i)$
size ℓ	$2n$	$n+1$	$n+1$
cm κ	$\frac{1}{\sqrt{n}}$	$\frac{1}{n}$	$\frac{1}{\sqrt{n^2+2(n-1)\sqrt{n}}}$
$\ell\kappa^{-2}$	$2n^2$	n^3+n^2	$n^3+2(n^2-1)\sqrt{n}+n^2$

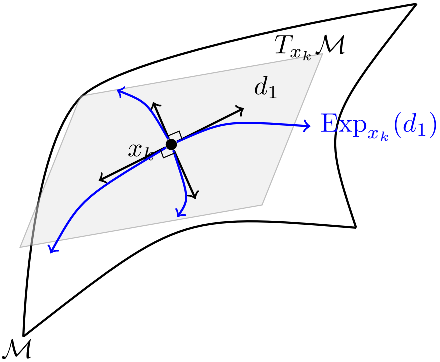
Popular PSSs in $\mathbb{R}^n$

$\longrightarrow$ for Riemannian Manifolds ?

Method 1: Designing PSS in the tangent space

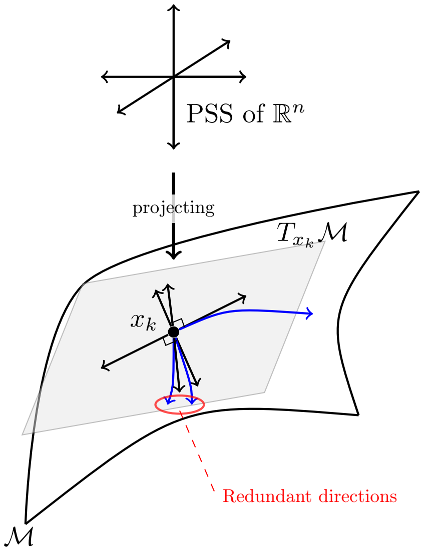


Method 1: Designing PSS in the tangent space

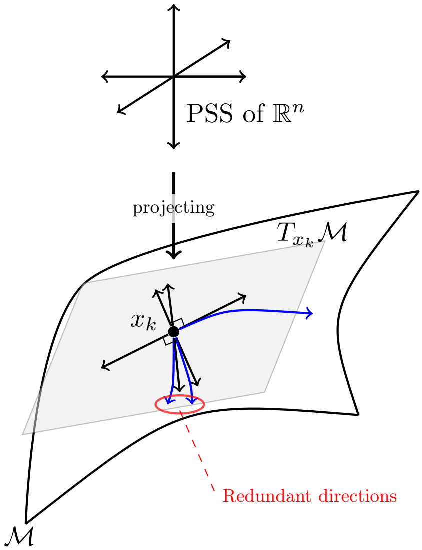


- ✗ Requires orthogonal bases of $\mathrm{T}_{x_k}\mathcal{M}$
- ✓ Computable cosine measure
- ✓ Can be reused via parallel transport (consistant κ)

Method 2: Projecting PSS to the tangent space



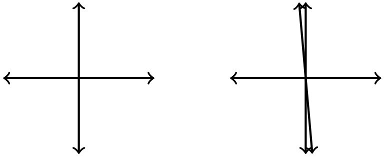
Method 2: Projecting PSS to the tangent space



- ✓ Easily obtained
- ✗ Requires the orthogonal projection
- ✗ Only applies to embedded manifolds
- ✗ Redudancy

Cardinal/cosine measure trade-off: Riemannian case

Complexity: $\mathcal{O}(\ell\kappa^{-2}\epsilon^{-2})$

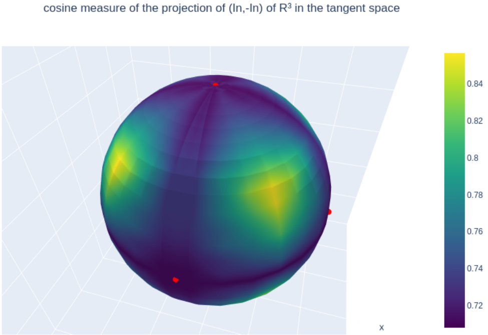


PSS D_k	$(I_m^k, -I_m^k)$	$proj_{\mathrm{T}_{x_k}\mathcal{M}}(I_n, -I_n)$
size ℓ	$2m$	$2n$
cm κ	$\frac{1}{\sqrt{m}}$	?
$\ell\kappa^{-2}$	$2m^2$	?

Which[3] PSS in $\mathrm{T}_{x_k}\mathcal{M}$

[3] m = Manifold dimension

A toy case: the sphere $\mathcal{M} = \mathbb{S}^{n-1} := \{x \in \mathbb{R}^n | \|x\|^2 = 1\}$



PSS D_k	$(I_m^k, -I_m^k)$	$proj_{\mathrm{T}_{x_k}\mathcal{M}}(I_n, -I_n)$
size ℓ	$2(n-1)$	$2n$
cm κ	$\frac{1}{\sqrt{n-1}}$	$\frac{1}{\sqrt{n-1}}$?
$\ell\kappa^{-2}$	$2(n-1)^2$	$2n(n-1)$

Table of content

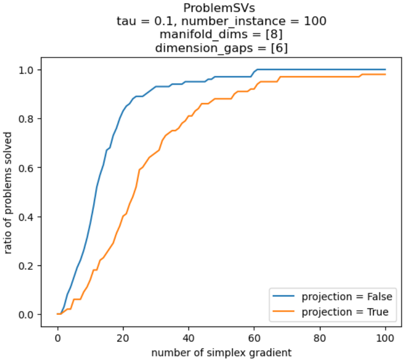
Problem instances [Boumal, 2023][Kungurstev et al., 2024]

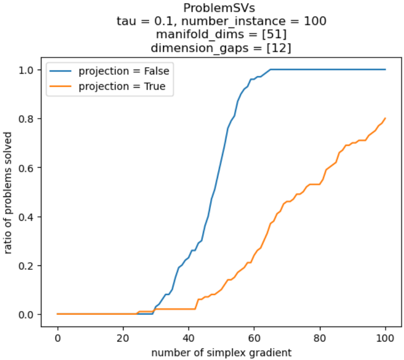
Data profiles:

- Budget: 100 gradient simplex
- Solved *iif.* $\frac{f(x)-f_{\text{low,dfo}}}{f(x_0)-f_{\text{low,dfo}}} \leq \tau$ where $\tau > 0$.
- Random initialization and parameters

Problem	**Data**	**Manifold**	**Cost function**
Lowest r Sing. Val.	$A \in \mathbb{R}^{m \times n}$ $D \in \text{diag}^{+}(r)$	$St(m,r)$ $\times St(n,r)$	$f(U,V)=$ $\sum_j \langle (UD)_{\cdot j}, (AV)_{\cdot j} \rangle$
Dict. Learning	$Y \in \mathbb{R}^{m \times n}$ $\lambda > 0,\ \epsilon > 0$	$Ob(m,r)$ $\times \mathbb{R}^{r \times n}$	$f(D,C)=\Vert Y-DC \Vert_{\text{Fro}}{}^2$ $+\lambda \sum_{i,j} \sqrt{C_{i,j}^2+\epsilon^2}$

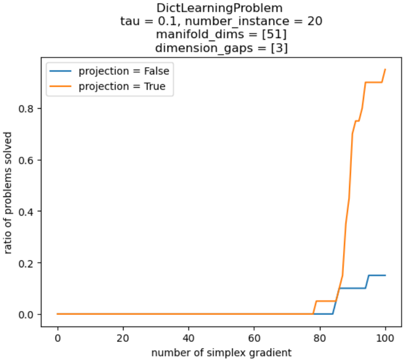
Singular values problem

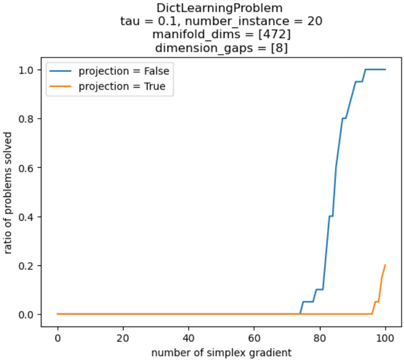




Left : $m = 8$, Right : $m = 51$

Dictionary Learning





Left : $m = 51$, Right : $m = 472$

Higher dimension favors Method 1

Conclusion

Achieved :

- Complexity bound generalized to Riemannian manifolds.
- Preliminary experiments favor PSSs in tangent space.

Further work:

- Test other PSSs (than orthogonal)
- Compute cosine measure κ of projected PSSs
- Random subspace directions.

Bibliography

Boumal, N. (2023).

An introduction to optimization on smooth manifolds.

Kungurstev, V., Rinaldi, F., and Zeffiro, D. (2024).

Retraction-based direct search methods for derivative-free Riemannian optimization.

J. Optim. Theory Appl.

Vicente, L. (2013).

Worst case complexity of direct search.

Euro. J. Comput. Optim.

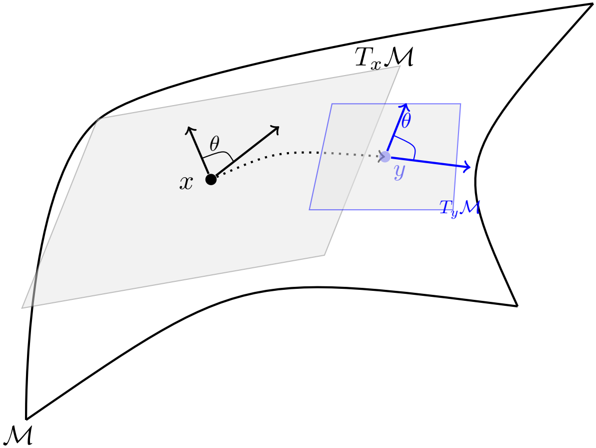
Zhang, H. and Sra, S. (2016).

First-order methods for geodesically convex optimization.

In *Conf. Learn. Theory*. PMLR.

Thank you for your time !
Questions ?

Parallel transport



Caption

Geodesic smoothness

Definition (geodesic L_g-smoothness [Zhang and Sra, 2016])

A function $f : \mathcal{M} \to \mathbb{R}$ of class $\mathcal{C}^1$ has *Lipschitz gradient* or is *geodesically L_g-smooth* if for any $a, b \in \mathcal{M}$ such that $\operatorname{dist}(a, b) < \operatorname{inj}(\mathcal{M})$,

$$\left\|\operatorname{grad} f(a) - \operatorname{grad} f(b)\right\|_a \leq L_g \operatorname{dist}(a, b)$$

where $L_g > 0$ and Γ_a^b is the parallel transport along $t \mapsto \operatorname{Exp}_a(t \operatorname{Log}_a b)$.

This implies, respecting the injectivity radius:

$$f(\operatorname{Exp}_a(v)) \leq f(a) + \left\langle \operatorname{grad} f(a), v \right\rangle_a + \frac{L_g}{2} \left\|v\right\|_a^2 .$$

Definition (geodesic L_g-smoothness [Zhang and Sra, 2016])

A function $f : \mathcal{M} \to \mathbb{R}$ of class $\mathcal{C}^1$ has *Lipschitz gradient* or is *geodesically L_g-smooth* if for any $a, b \in \mathcal{M}$ such that $\operatorname{dist}(a, b) < \operatorname{inj}(\mathcal{M})$,

$$\|\operatorname{grad} f(a) - \Gamma_b^a \operatorname{grad} f(b)\|_a \leq L_g \operatorname{dist}(a, b)$$

where $L_g > 0$ and Γ_a^b is the parallel transport along $t \mapsto \operatorname{Exp}_a(t \operatorname{Log}_a b)$.

This implies, respecting the injectivity radius:

$$f(\operatorname{Exp}_a(v)) \leq f(a) + \langle \operatorname{grad} f(a), v \rangle_a + \frac{L_g}{2} \|v\|_a^2 .$$

A more detailed complexity bound

$$
\begin{aligned}
j_1 \leq{}& \frac{f(x_0) - f_{low}}{c} \left(\frac{d_{max}(c + L_g/2)}{\gamma b} \right)^2 \left(1 - \frac{\log(\Gamma)}{\log(\gamma)} \right) \kappa^{-2} \epsilon^{-2} \\
&+ \frac{1}{\log(\gamma)} \log \left(\frac{\gamma}{d_{\max}(c + L_g/2)} \kappa \epsilon \right) + \frac{f(x_0) - f_{low}}{c \alpha_0^2} - \frac{\log \alpha_{max}}{\log \gamma} - 1
\end{aligned}
$$