Riemannian Optimization with its Application to Averaging Positive Definite Matrices

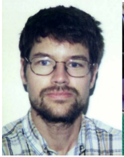
Wen Huang

Xiamen University
Malaysia

December 4, 2018

Joint work with:

- Pierre-Antoine Absil, Professor of Mathematical Engineering, *Université catholique de Louvain*
- Kyle A. Gallivan, Professor of Mathematics, *Florida State University*
- Xinru Yuan, Data scientist, *Esurance*

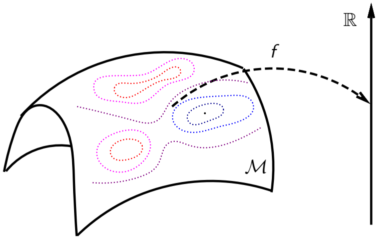


Riemannian Optimization

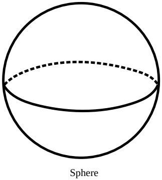
Problem: Given $f(x) : \mathcal{M} \to \mathbb{R}$, solve

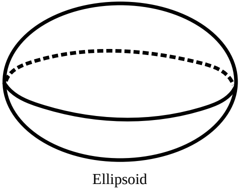
$$\min_{x \in \mathcal{M}} f(x)$$

where $\mathcal{M}$ is a Riemannian manifold.



Examples of Manifolds

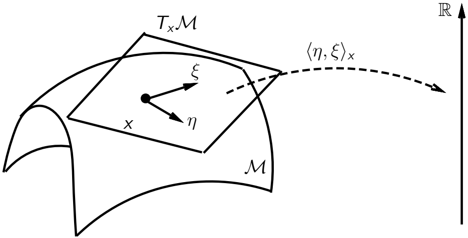




- Stiefel manifold: $\mathrm{St}(p, n) = \{X \in \mathbb{R}^{n \times p} | X^T X = I_p\}$
- Grassmann manifold: Set of all p-dimensional subspaces of $\mathbb{R}^n$
- Set of fixed rank m-by-n matrices
- And many more

Riemannian Manifolds

Roughly, a Riemannian manifold $\mathcal{M}$ is a smooth set with a smoothly-varying inner product on the tangent spaces.

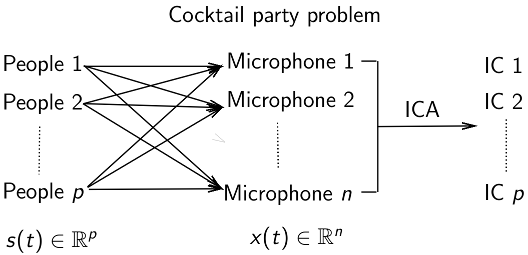


Applications

Three applications are used to demonstrate the importance of the Riemannian optimization:

- Independent component analysis [CS93]
- Matrix completion problem [Van13, HAGH16]
- Elastic shape analysis of curves [SKJJ11, HGSA15]

Application: Independent Component Analysis



- Observed signal is $x(t) = As(t)$
- One approach:
  - Assumption: $E\{s(t)s(t+\tau)\}$ is diagonal for all τ
  - $C_\tau(x) := E\{x(t)x(x+\tau)^T\} = AE\{s(t)s(t+\tau)^T\}A^T$

Application: Independent Component Analysis

- Minimize joint diagonalization cost function on the Stiefel manifold [TI06]:

$$f : \mathrm{St}(p, n) \to \mathbb{R} : V \mapsto \sum_{i=1}^{N} \|V^T C_i V - \mathrm{diag}(V^T C_i V)\|_F^2.$$

- $C_1, \dots, C_N$ are covariance matrices and $\mathrm{St}(p, n) = \{X \in \mathbb{R}^{n \times p} | X^T X = I_p\}$.

Application: Matrix Completion Problem

Matrix completion problem

$$
\begin{array}{c}
 \\
\text{User } 1 \\
\text{User } 2 \\
 \\
\text{User } m
\end{array}
\begin{array}{c}
\begin{array}{cccccc} \text{Movie } 1 & \text{Movie } 2 & & & & \text{Movie } n \end{array} \\
\begin{pmatrix}
 & 1 & & & 4 & \\
3 & 5 & & & 4 & \\
 & & 5 & 1 & & \\
 & 2 & & & 5 & 3
\end{pmatrix}
\end{array}
$$

Rate matrix M

- The matrix M is sparse
- The goal: complete the matrix M

Application: Matrix Completion Problem

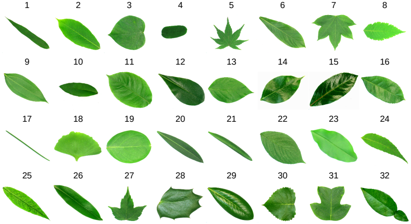
$$\begin{array}{c} \text{movies} \\ \begin{pmatrix} a_{11} & & & a_{14} \\ & & & a_{24} \\ & & a_{33} & \\ a_{41} & & & \\ & a_{52} & a_{53} & \end{pmatrix} \end{array} = \begin{array}{c} \text{meta-user} \\ \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \\ b_{41} & b_{42} \\ b_{51} & b_{52} \end{pmatrix} \end{array} \begin{array}{c} \text{meta-movie} \\ \begin{pmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{21} & c_{22} & c_{23} & c_{24} \end{pmatrix} \end{array}$$

- Minimize the cost function

$$f : \mathbb{R}_r^{m \times n} \to \mathbb{R} : X \mapsto f(X) = \|P_\Omega M - P_\Omega X\|_F^2.$$

- $\mathbb{R}_r^{m \times n}$ is the set of m-by-n matrices with rank r. It is known to be a Riemannian manifold.

Application: Elastic Shape Analysis of Curves



- Classification [LKS+12, HGSA15]
- Face recognition [DBS+13]

Application: Elastic Shape Analysis of Curves

- Elastic shape analysis invariants:
  - Rescaling
  - Translation
  - Rotation
  - Reparametrization
- The shape space is a quotient space

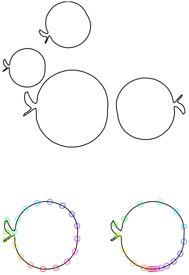
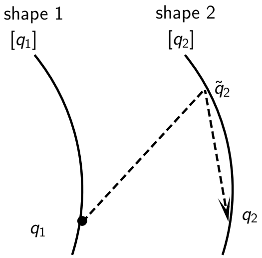


Figure: All are the same shape.

Application: Elastic Shape Analysis of Curves



- Optimization problem $\min_{q_2 \in [q_2]} \mathrm{dist}(q_1, q_2)$ is defined on a Riemannian manifold
- Computation of a geodesic between two shapes
- Computation of Karcher mean of a population of shapes

More Applications

- Role model extraction [MHB$^+$16]
- Computations on SPD matrices [YHAG17]
- Phase retrieval problem [HGZ17]
- Blind deconvolution [HH17]
- Synchronization of rotations [Hua13]
- Computations on low-rank tensor
- Low-rank approximate solution for Lyapunov equation

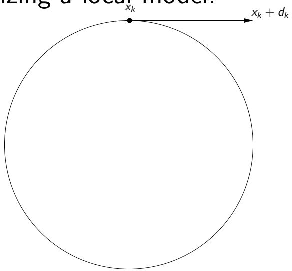
Iterations on the Manifold

Consider the following generic update for an iterative Euclidean optimization algorithm:

$$x_{k+1} = x_k + \Delta x_k = x_k + \alpha_k s_k \ .$$

This iteration is implemented in numerous ways, e.g.:

- Steepest descent: $x_{k+1} = x_k - \alpha_k \nabla f(x_k)$
- Newton's method: $x_{k+1} = x_k - \left[\nabla^2 f(x_k)\right]^{-1} \nabla f(x_k)$
- Trust region method: Δx_k is set by optimizing a local model.



Riemannian Manifolds Provide

- Riemannian concepts describing directions and movement on the manifold
- Riemannian analogues for gradient and Hessian

Riemannian gradient and Riemannian Hessian

Definition

The Riemannian gradient of f at x is the unique tangent vector in $T_x M$ satisfying $\forall \eta \in T_x M$, the directional derivative

$$\mathrm{D}\, f(x)[\eta] = \langle \operatorname{grad} f(x), \eta \rangle$$

and $\operatorname{grad} f(x)$ is the direction of steepest ascent.

Definition

The Riemannian Hessian of f at x is a symmetric linear operator from $T_x M$ to $T_x M$ defined as

$$\operatorname{Hess} f(x) : T_x M \to T_x M : \eta \to \nabla_\eta \operatorname{grad} f,$$

where ∇ is the affine connection.

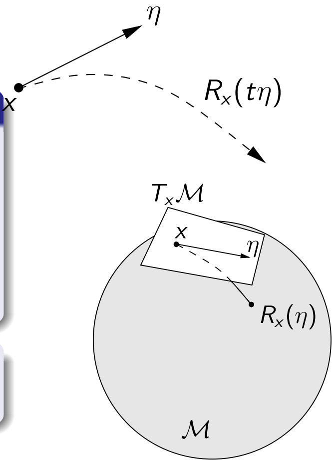
Retractions

Euclidean	Riemannian
$x_{k+1} = x_k + \alpha_k d_k$	$x_{k+1} = R_{x_k}(\alpha_k \eta_k)$

Definition

A retraction is a mapping R from TM to M satisfying the following:

- R is continuously differentiable
- $R_x(0) = x$
- $\mathrm{D}\, R_x(0)[\eta] = \eta$



- maps tangent vectors back to the manifold
- defines curves in a direction

Categories of Riemannian optimization methods

Retraction-based: local information only

Line search-based: use local tangent vector and $R_x(t\eta)$ to define line

- Steepest decent
- Newton

Local model-based: series of flat space problems

- Riemannian trust region Newton (RTR)
- Riemannian adaptive cubic overestimation (RACO)

Categories of Riemannian optimization methods

Retraction and transport-based: information from multiple tangent spaces

- Nonlinear conjugate gradient: multiple tangent vectors
- Quasi-Newton e.g. Riemannian BFGS: transport operators between tangent spaces

Additional element required for optimizing a cost function (M, g):

- formulas for combining information from multiple tangent spaces.

Vector Transports

Vector Transport

- Vector transport: Transport a tangent vector from one tangent space to another
- $\mathcal{T}_{\eta_x}\xi_x$, denotes transport of ξ_x to tangent space of $R_x(\eta_x)$. R is a retraction associated with $\mathcal{T}$

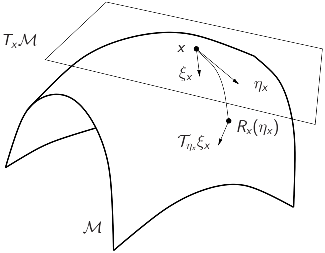


Figure: Vector transport.

Retraction/Transport-based Riemannian Optimization

Given a retraction and a vector transport, we can generalize many Euclidean methods to the Riemannian setting. Do the Riemannian versions of the methods work well?

Retraction/Transport-based Riemannian Optimization

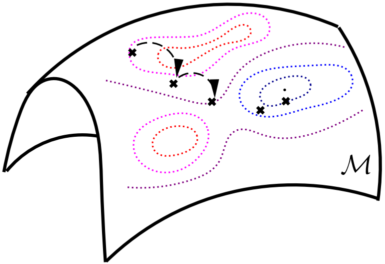
Given a retraction and a vector transport, we can generalize many Euclidean methods to the Riemannian setting. Do the Riemannian versions of the methods work well?

No

- Lose many theoretical results and important properties;
- Impose restrictions on retraction/vector transport;

Comparison with Constrained Optimization

- All iterates on the manifold
- Convergence properties of unconstrained optimization algorithms
- No need to consider Lagrange multipliers or penalty functions
- Exploit the structure of the constrained set



Some History of Optimization On Manifolds (I)

Luenberger (1973), *Introduction to linear and nonlinear programming*. Luenberger mentions the idea of performing line search along geodesics, "which we would use if it were computationally feasible (which it definitely is not)". Rosen (1961) essentially anticipated this but was not explicit in his Gradient Projection Algorithm.

Gabay (1982), *Minimizing a differentiable function over a differential manifold*. Steepest descent along geodesics; Newton's method along geodesics; Quasi-Newton methods along geodesics. On Riemannian submanifolds of $\mathbb{R}^n$.

Smith (1993-94), *Optimization techniques on Riemannian manifolds*. Levi-Civita connection ∇; Riemannian exponential mapping; parallel translation.

Some History of Optimization On Manifolds (II)

The “pragmatic era” begins:

Manton (2002), *Optimization algorithms exploiting unitary constraints* “The present paper breaks with tradition by not moving along geodesics”. The geodesic update $\mathrm{Exp}_x \eta$ is replaced by a projective update $\pi(x + \eta)$, the *projection* of the point $x + \eta$ onto the manifold.

Adler, Dedieu, Shub, et al. (2002), *Newton's method on Riemannian manifolds and a geometric model for the human spine*. The exponential update is relaxed to the general notion of *retraction*. The geodesic can be replaced by any (smoothly prescribed) curve tangent to the search direction.

Some History of Optimization On Manifolds (III)

Theory, efficiency, and library design improve dramatically:

Absil, Baker, Gallivan (2004-07), Theory and implementations of Riemannian Trust Region method. Retraction-based approach. Matrix manifold problems, software repository:
`http://www.math.fsu.edu/~cbaker/GenRTR`
Anasazi Eigenproblem package in Trilinos Library at Sandia National Laboratory

Ring and With (2012), combination of differentiated retraction and isometric vector transport for convergence analysis of RBFGS

Absil, Gallivan, Huang (2009-2017), Complete theory of Riemannian Quasi-Newton and related transport/retraction conditions, Riemannian SR1 with trust-region, RBFGS on partly smooth problems, A C++ library: `http://www.math.fsu.edu/~whuang2/ROPTLIB`

Some History of Optimization On Manifolds (IV)

Ring and With (2012), Global convergence analysis for Fletcher-Reeves Riemannian nonlinear CG method with the strong wolfe conditions under a strong assumption.

Sato, Iwai (2013-2015), Global convergence analysis for Fletcher-Reeves type Riemannian nonlinear CG method with the strong wolfe conditions under a mild assumption; and global convergence for Dai-Yuan type Riemannian nonlinear CG method with the weak wolfe conditions under mild assumptions.

Zhu (2017), Global convergence for Riemannian version of Dai's nonmonotone nonlinear CG method.

Some History of Optimization On Manifolds (IV)

Bonnabel (2011) Riemannian stochastic gradient descent method.

Sato, Kasai, Mishra(2017) Riemannian stochastic gradient descent method using variance reduction or quasi-Newton.

Becigneul, Ganea(2018) Riemannian versions of ADAM, ADAGRAD, and AMSGRAD for geodesically convex functions.

Zhang, Sra(2016-2018) Riemannian first-order methods for geodesically convex optimization.

Bento, Ferreira, Melo(2017) Chen, Ma, So, Zhang(2018) Riemannian proximal gradient method.

Many people Application interests increase noticeably

Riemannian Optimization Libraries

Four Riemannian optimization libraries for general problems:

- Boumal, Mishra, Absil, Sepulchre(2014)
  Manopt (Matlab library)
- Townsend, Koep, Weichwald (2016)
  Pymanopt (Python version of manopt)
- Huang, Absil, Gallivan, Hand (2018)
  ROPTLIB (C++ library, interfaces to Matlab and Julia)
- Martin, Raim, Huang, Adragni(2018)
  ManifoldOptim (R wrapper of ROPTLIB)

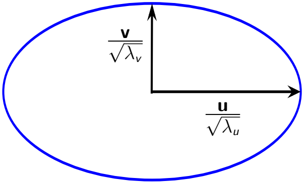
Application: Averaging Symmetric Positive Definite matrices

Definition

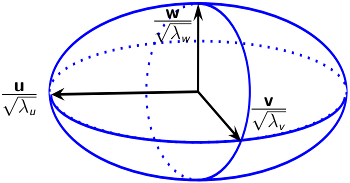
A symmetric matrix A is called positive definite $A \succ 0$ iff all its eigenvalues are positive.

$$\mathcal{S}_{++}^{n} = \{A \in \mathbb{R}^{n \times n} : A = A^T, A \succ 0\}$$

2×2 SPD matrix

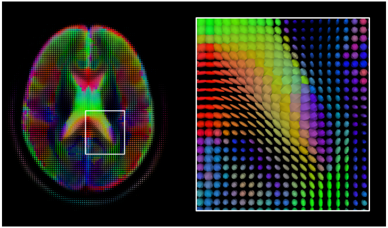


3×3 SPD matrix



Motivation of Averaging SPD Matrices

- Possible applications of SPD matrices
  - Diffusion tensors in medical imaging [CSV12, FJ07, RTM07]
  - Describing images and video [LWM13, SFD02, ASF^{+}05, TPM06, HWSC15]



- Motivation of averaging SPD matrices
  - Aggregate several noisy measurements of the same object
  - Subtask in interpolation methods, segmentation, and clustering

Averaging Schemes: from Scalars to Matrices

Let $A_1, \ldots, A_K$ be SPD matrices.

- Generalized arithmetic mean: $\frac{1}{K}\sum_{i=1}^{K} A_i$

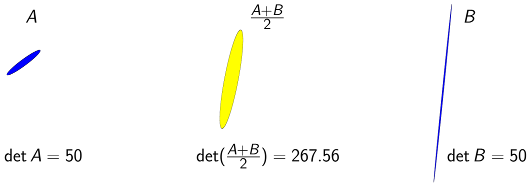
  $\rightarrow$ Not appropriate in many practical applications

Averaging Schemes: from Scalars to Matrices

Let $A_1, \ldots, A_K$ be SPD matrices.

- Generalized arithmetic mean: $\frac{1}{K}\sum_{i=1}^{K} A_i$

  $\rightarrow$ Not appropriate in many practical applications

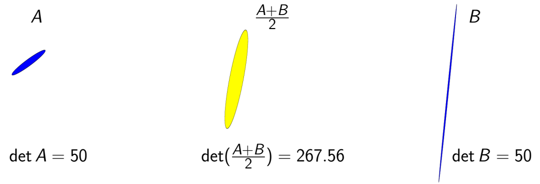


Averaging Schemes: from Scalars to Matrices

Let $A_1, \ldots, A_K$ be SPD matrices.

- Generalized arithmetic mean: $\frac{1}{K}\sum_{i=1}^{K} A_i$

  $\rightarrow$ Not appropriate in many practical applications



- Generalized geometric mean: $(A_1 \cdots A_K)^{1/K}$

  $\rightarrow$ Not appropriate due to non-commutativity

  $\rightarrow$ How to define a matrix geometric mean?

Desired Properties of a Matrix Geometric Mean

The desired properties are given in the ALM list[1], some of which are:

- $G(A_{\pi(1)}, \dots, A_{\pi(K)}) = G(A_1, \dots, A_K)$ with π a permutation of $(1, \dots, K)$
- if $A_1, \dots, A_K$ commute, then $G(A_1, \dots, A_K) = (A_1, \dots, A_K)^{1/K}$
- $G(A_1, \dots, A_K)^{-1} = G(A_1^{-1}, \dots, A_K^{-1})$
- $\det(G(A_1, \dots, A_K)) = (\det(A_1) \cdots \det(A_K))^{1/K}$

[1] T. Ando, C.-K. Li, and R. Mathias, *Geometric means*, Linear Algebra and Its Applications, 385:305-334, 2004

Geometric Mean of SPD Matrices

- A well-known mean on the manifold of SPD matrices is the **Karcher mean** [Kar77]:

$$G(A_1, \ldots, A_K) = \arg \min_{X \in \mathcal{S}_{++}^n} \frac{1}{2K} \sum_{i=1}^{K} \delta^2(X, A_i), \qquad (1)$$

where $\delta(X, Y) = \| \log(X^{-1/2} Y X^{-1/2}) \|_F$ is the geodesic distance under the affine-invariant metric

$$g(\eta_X, \xi_X) = \mathrm{trace}(\eta_X X^{-1} \xi_X X^{-1})$$

- The Karcher mean defined in (1) satisfies all the geometric properties in the ALM list [LL11]

Algorithms

$$G(A_1, \dots, A_k) = \operatorname*{argmin}_{X \in \mathcal{S}^n_{++}} \frac{1}{2K} \sum_{i=1}^{K} \delta^2(X, A_i),$$

- Riemannian steepest descent [RA11, Ren13]
- Riemannian Barzilai-Borwein method [IP15]
- Riemannian Newton method [RA11]
- Richardson-like iteration [BI13]
- Riemannian steepest descent, conjugate gradient, BFGS, and trust region Newton methods [JVV12]
- Limited-memory Riemannian BFGS method [YHAG16]

Remarks

Previous work:

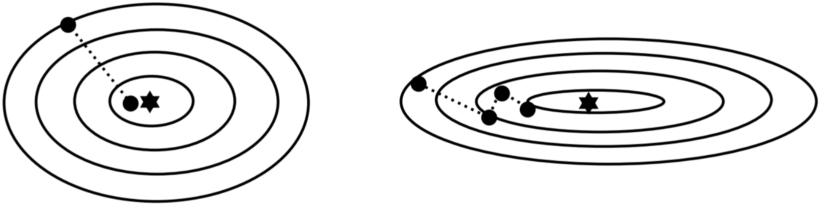
- Riemannian steepest descent and Riemannian CG methods are preferred in terms of computational time
- High rate of convergence of Riemannian Newton method does not make up for extra complexity

New results:

- Explain the preference for the first order methods
- More options of retractions and vector transports
- More efficient implementation
- Limited-memory Riemannian BFGS method

Conditioning of the Objective Function

Hemstitching phenomenon
for steepest descent



well-conditioned Hessian

ill-conditioned Hessian

- Small condition number $\Rightarrow$ fast convergence
- Large condition number $\Rightarrow$ slow convergence

Conditioning of the Karcher Mean Objective Function

- **Riemannian metric**:

$$g_X(\xi, \eta) = \mathrm{trace}(\xi X^{-1} \eta X^{-1})$$

- **Euclidean metric**:

$$g_X(\xi, \eta) = \mathrm{trace}(\xi \eta)$$

Condition number κ of Hessian at the minimizer μ:

- Hessian of Riemannian metric:
  - $\kappa(H^R) \leq 1 + \frac{\ln(\max \kappa_i)}{2}$, where $\kappa_i = \kappa(\mu^{-1/2} A_i \mu^{-1/2})$
  - $\kappa(H^R) \leq 20$ if $\max(\kappa_i) = 10^{16}$
- Hessian of Euclidean metric:
  - $\frac{\kappa^2(\mu)}{\kappa(H^{\mathrm{R}})} \leq \kappa(H^{\mathrm{E}}) \leq \kappa(H^{\mathrm{R}}) \kappa^2(\mu)$
  - $\kappa(H^E) \geq \kappa^2(\mu)/20$

Implementations

- Retraction
  - Exponential mapping: $\mathrm{Exp}_X(\xi) = X^{1/2} \exp(X^{-1/2} \xi X^{-1/2}) X^{1/2}$
  - Second order approximation retraction [JVV12]:
    $$R_X(\xi) = X + \xi + \frac{1}{2} \xi X^{-1} \xi$$
- Vector transport
  - Parallel translation: $\mathcal{T}_{p_\eta}(\xi) = Q \xi Q^T$, with $Q = X^{\frac{1}{2}} \exp(\frac{X^{-\frac{1}{2}} \eta X^{-\frac{1}{2}}}{2}) X^{-\frac{1}{2}}$
  - Vector transport by parallelization [HAG16]: essentially an identity
  - Requires orthogonal basis for tangent spaces

Implementations

- Cholesky $X_k = L_k L_k^T$ assumed to be computed on each step
- B_X of $\mathrm{T}_X \mathcal{S}^n_{++}$, the orthonormal basis of $\mathrm{T}_X \mathcal{S}^n_{++}$

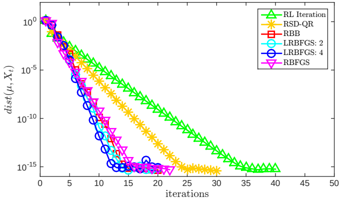
$$B_X = \{L e_i e_i^T L^T : i = 1, \dots, n\} \cup \{\frac{1}{\sqrt{2}} L(e_i e_j^T + e_j e_i^T) L^T,$$
$$i < j,\ i = 1, \dots, n,\ j = 1, \dots, n\},$$

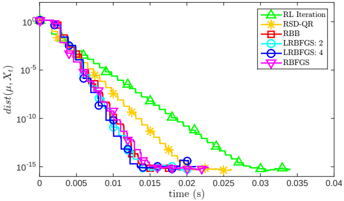
!where $\{e_i, \dots, e_n\}$ is the standard basis of n-dimensional Euclidean space.

- orthonormal under $g_X(\xi_X, \eta_X)$.
- $\xi_X = B_X \hat{\xi}_X \leftrightarrow \xi_X = LSL^T$, where S is symmetric and constains scale coefficients.
- intrinsic representation of tangent vectors is easily maintained.

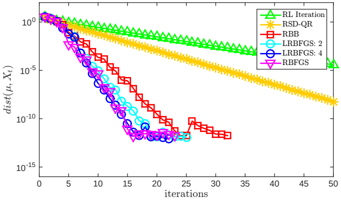
Numerical Results: $K = 100$, size $= 3 \times 3$, $d = 6$

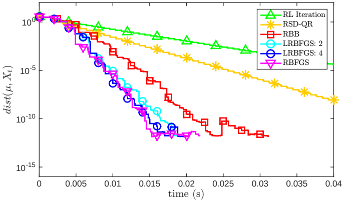
- $1 \le \kappa(A_i) \le 200$





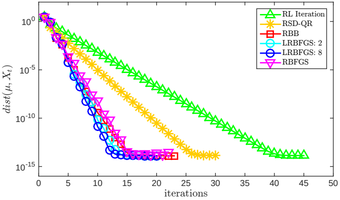
- $10^3 \le \kappa(A_i) \le 10^7$

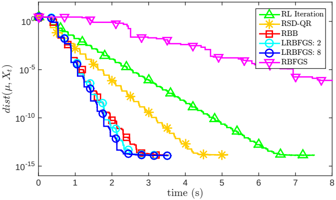




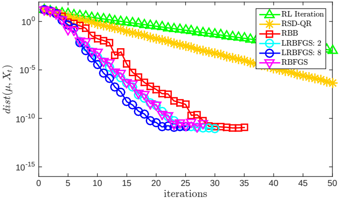
Numerical Results: $K = 30$, size $= 100 \times 100$, $d = 5050$

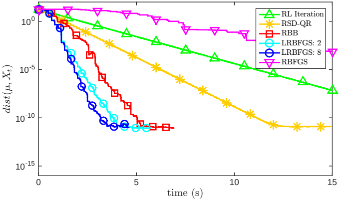
- $1 \leq \kappa(A_i) \leq 20$





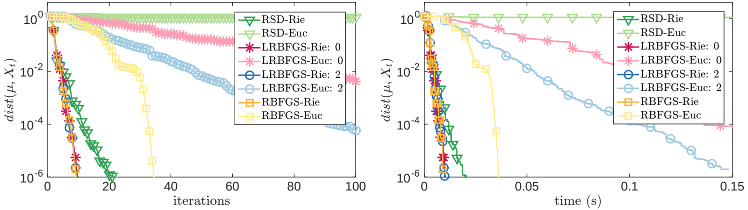
- $10^4 \leq \kappa(A_i) \leq 10^7$



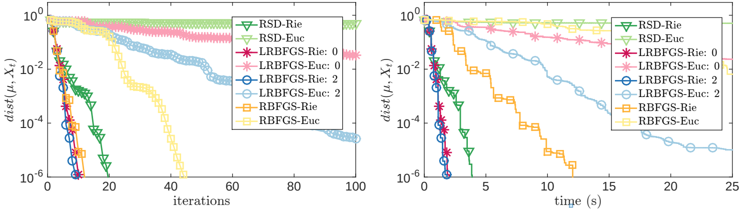


Numerical Results: Riemannian vs. Euclidean Metrics

- $K = 100$, $n = 3$, and $1 \leq \kappa(A_i) \leq 10^6$.



- $K = 30$, $n = 100$, and $1 \leq \kappa(A_i) \leq 10^5$.



Other types of mean

Sacrifice a few properties in the ALM list to gain robustness/efficiency etc

- Log-Euclidean mean
- arithmetic-harmonic mean
- Divergence-based means
- Inductive means
- Minimax centers
- Medians

Other types of mean

Sacrifice a few properties in the ALM list to gain robustness/efficiency etc

- Log-Euclidean mean
- arithmetic-harmonic mean
- Divergence-based means
- Inductive means
- Minimax centers
- Medians
  The median based on Riemannian distance

Motivations

- The mean of a set of points is sensitive to outliers
- The median is robust to outliers

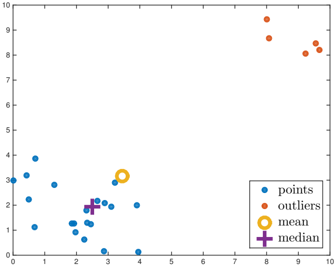


Figure: The geometric mean and median in $\mathbb{R}^2$ space.

Riemannian Median of SPD Matrices

The Riemannian median of a set of SPD matrices is defined as:

$$M(A_1, \dots, A_K) = \operatorname*{argmin}_{X \in \mathcal{S}_{++}^n} \frac{1}{2K} \sum_{i=1}^{K} \delta(A_i, X),$$

where $\delta(X, Y)$ is a distance function.

- The cost function is non-smooth at $X = A_i$
- The median is unique

Algorithms

$$M(A_1, \ldots, A_K) = \operatorname*{argmin}_{X \in \mathcal{S}^n_{++}} \frac{1}{2K} \sum_{i=1}^{K} \delta(A_i, X)$$

- Riemannian Weiszfeld's algorithm [FVJ09]
- Our approach: Riemannian quasi-Newton algorithms
  - Smooth RBFGS [HAG18]
  - Modified RBFGS [Hua13]
  - Nonsmooth RBFGS [HHY18]
  - Limited-memory versions of the above three [HAGH16]

Numerical Results for L^1 Median Computation on $\mathcal{S}^n_{++}$: Comparison of Different Algorithms

- $K = 100$, size $= 3 \times 3$
- well-conditioned A_i

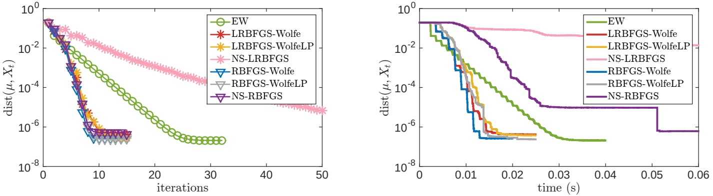


Figure: Evolution of averaged distance between current iterate and the exact Riemannian median with respect to time and iterations.

Numerical Results for L^1 Median Computation on $\mathcal{S}^n_{++}$: Comparison of Different Algorithms

- $K = 100$, size $= 3 \times 3$
- well-conditioned A_i + 5% ill-conditioned outliers

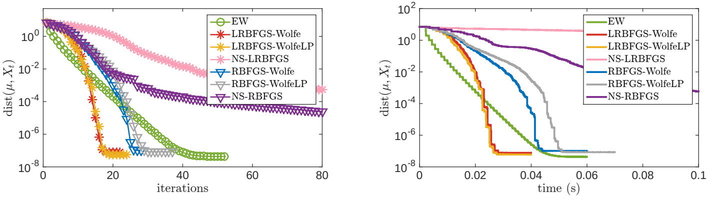
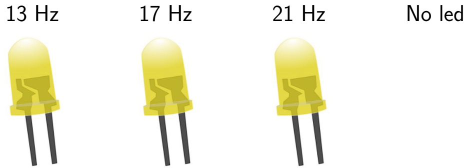


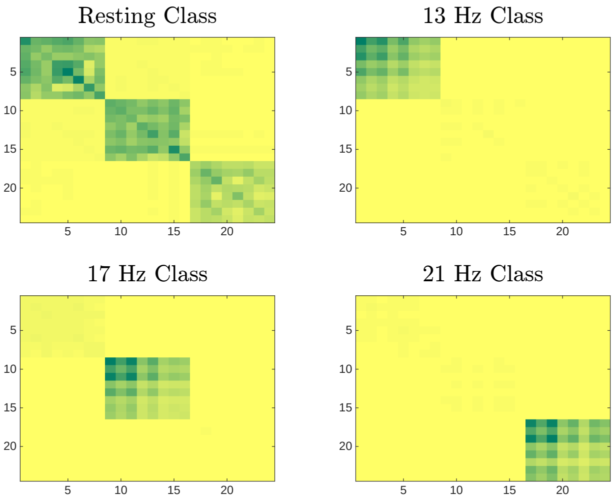
Figure: Evolution of averaged distance between current iterate and the exact Riemannian median with respect to time and iterations.

Application: Electroencephalography (EEG) Classification



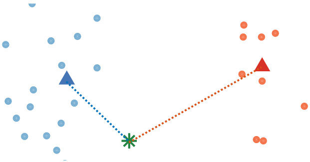
- The subject is either asked to focus on one specific blinking LED or a location without LED
- EEG system is used to record brain signals
- Covariance matrices of size 24×24 are used to represent EEG recordings [KCB$^+$15, MC17]

EEG Classification: Examples of Covariance Matrices



EEG Classification: Minimum Distance to Mean classier

Goal: classify new covariance matrix using Minimum Distance to Mean Classifier

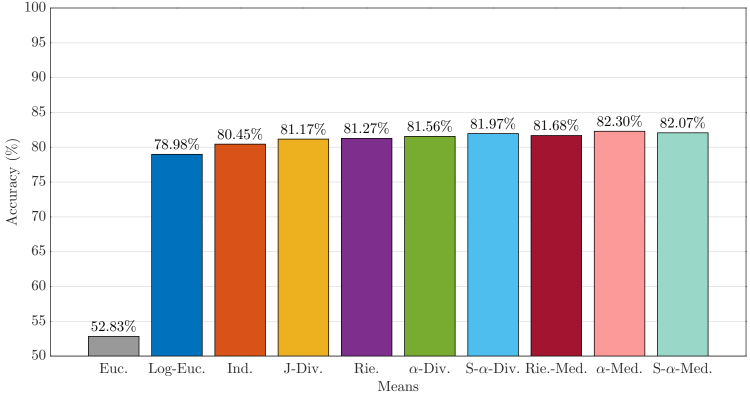


- For each class $k = 1, \ldots, K$, compute the center μ_k of the covariance matrices in the training set that belong to class k
- Classify a new covariance matrix X according to

$$\hat{k} = \operatorname*{argmin}_{1 \leq k \leq K} \delta(X, \mu_k)$$

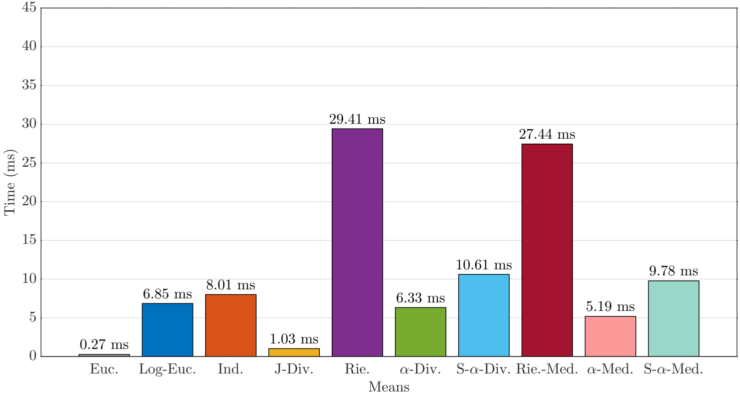
EEG Classfification: Accuracy

- Accuracy comparison



EEG Classfification: Computation Time

- Computation time comparison



Summary

- Introduced the framework of Riemannian optimization
- Used applications to show the importance of Riemannian optimization
- Briefly reviewed the history of Riemannian optimization
- Introduced the mean of SPD matrices
- Demonstrated the performance of the Riemannian methods

Thank you

Thank you!

References I

Ognjen Arandjelovic, Gregory Shakhnarovich, John Fisher, Roberto Cipolla, and Trevor Darrell.
Face recognition with image sets using manifold density divergence.
In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, volume 1, pages 581–588. IEEE, 2005.

D. A. Bini and B. Iannazzo.
Computing the Karcher mean of symmetric positive definite matrices.
Linear Algebra and its Applications, 438(4):1700–1710, 2013.

J. F. Cardoso and A. Souloumiac.
Blind beamforming for non-gaussian signals.
IEE Proceedings F Radar and Signal Processing, 140(6):362, 1993.

Guang Cheng, Hesamoddin Salehian, and Baba Vemuri.
Efficient recursive algorithms for computing the mean diffusion tensor and applications to DTI segmentation.
Computer Vision–ECCV 2012, pages 390–401, 2012.

H. Drira, B. Ben Amor, A. Srivastava, M. Daoudi, and R. Slama.
3D face recognition under expressions, occlusions, and pose variations.
Pattern Analysis and Machine Intelligence, IEEE Transactions on, 35(9):2270–2283, 2013.

P. T. Fletcher and S. Joshi.
Riemannian geometry for the statistical analysis of diffusion tensor data.
Signal Processing, 87(2):250–262, 2007.

P Thomas Fletcher, Suresh Venkatasubramanian, and Sarang Joshi.
The geometric median on Riemannian manifolds with application to robust atlas estimation.
NeuroImage, 45(1):S143–S152, 2009.

References II

W. Huang, P.-A. Absil, and K. A. Gallivan.
Intrinsic representation of tangent vectors and vector transport on matrix manifolds.
Numerische Mathematik, 136(2):523–543, 2016.

Wen Huang, P.-A. Absil, and K. A. Gallivan.
A Riemannian BFGS Method without Differentiated Retraction for Nonconvex Optimization Problems.
SIAM Journal on Optimization, 28(1):470–495, 2018.

Wen Huang, P.-A. Absil, K. A. Gallivan, and Paul Hand.
ROPTLIB: an object-oriented C++ library for optimization on Riemannian manifolds.
Technical Report FSU16-14, Florida State University, 2016.

W. Huang, K. A. Gallivan, Anuj Srivastava, and P.-A. Absil.
Riemannian optimization for registration of curves in elastic shape analysis.
Journal of Mathematical Imaging and Vision, 54(3):320–343, 2015.
DOI:10.1007/s10851-015-0606-8.

Wen Huang, K. A. Gallivan, and Xiangxiong Zhang.
Solving PhaseLift by low rank Riemannian optimization methods for complex semidefinite constraints.
SIAM Journal on Scientific Computing, 39(5):B840–B859, 2017.

W. Huang and P. Hand.
Blind deconvolution by a steepest descent algorithm on a quotient manifold, 2017.
In preparation.

S. Hosseini, Wen Huang, and R. Yousefpour.
Line search algorithms for locally Lipschitz functions on Riemannian manifolds.
SIAM Journal on Optimization, 28(1):596–619, 2018.

References III

W. Huang.
Optimization algorithms on Riemannian manifolds with applications.
PhD thesis, Florida State University, Department of Mathematics, 2013.

Zhiwu Huang, Ruiping Wang, Shiguang Shan, and Xilin Chen.
Face recognition on large-scale video in the wild with hybrid Euclidean-and-Riemannian metric learning.
Pattern Recognition, 48(10):3113–3124, 2015.

Bruno Iannazzo and Margherita Porcelli.
The Riemannian Barzilai-Borwein method with nonmonotone line-search and the Karcher mean computation.
Optimization online, December, 2015.

B. Jeuris, R. Vandebril, and B. Vandereycken.
A survey and comparison of contemporary algorithms for computing the matrix geometric mean.
Electronic Transactions on Numerical Analysis, 39:379–402, 2012.

H. Karcher.
Riemannian center of mass and mollifier smoothing.
Communications on Pure and Applied Mathematics, 1977.

Emmanuel K Kalunga, Sylvain Chevallier, Quentin Barthélemy, Karim Djouani, Yskandar Hamam, and Eric Monacelli.
From Euclidean to Riemannian means: Information geometry for SSVEP classification.
In *International Conference on Networked Geometric Science of Information*, pages 595–604. Springer, 2015.

H. Laga, S. Kurtek, A. Srivastava, M. Golzarian, and S. J. Miklavcic.
A Riemannian elastic metric for shape-based plant leaf classification.
2012 International Conference on Digital Image Computing Techniques and Applications (DICTA), pages 1–7, December 2012.
doi:10.1109/DICTA.2012.6411702.

References IV

J. Lawson and Y. Lim.
Monotonic properties of the least squares mean.
Mathematische Annalen, 351(2):267–279, 2011.

Jiwen Lu, Gang Wang, and Pierre Moulin.
Image set classification using holistic multiple order statistics features and localized multi-kernel metric learning.
In *Proceedings of the IEEE International Conference on Computer Vision*, pages 329–336, 2013.

Estelle M Massart and Sylvain Chevallier.
Inductive means and sequences applied to online classication of EEG.
Technical report, ICTEAM Institute, Universite Catholiqué de Louvain, 2017.

Melissa Marchand, Wen Huang, Arnaud Browet, Paul Van Dooren, and Kyle A. Gallivan.
A riemannian optimization approach for role model extraction.
In *Proceedings of the 22nd International Symposium on Mathematical Theory of Networks and Systems*, pages 58–64, 2016.

Q. Rentmeesters and P.-A. Absil.
Algorithm comparison for Karcher mean computation of rotation matrices and diffusion tensors.
In *19th European Signal Processing Conference*, pages 2229–2233, Aug 2011.

Q. Rentmeesters.
Algorithms for data fitting on some common homogeneous spaces.
PhD thesis, Universite catholiqué de Louvain, 2013.

Y. Rathi, A. Tannenbaum, and O. Michailovich.
Segmenting images on the tensor manifold.
In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2007.

Gregory Shakhnarovich, John W Fisher, and Trevor Darrell.
Face recognition from long-term observations.
In *European Conference on Computer Vision*, pages 851–865. Springer, 2002.

References V

A. Srivastava, E. Klassen, S. H. Joshi, and I. H. Jermyn.
Shape analysis of elastic curves in Euclidean spaces.
IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(7):1415–1428, September 2011.
doi:10.1109/TPAMI.2010.184.

F. J. Theis and Y. Inouye.
On the use of joint diagonalization in blind signal processing.
2006 IEEE International Symposium on Circuits and Systems, (2):7–10, 2006.

Oncel Tuzel, Fatih Porikli, and Peter Meer.
Region covariance: A fast descriptor for detection and classification.
In *European conference on computer vision*, pages 589–600. Springer, 2006.

B. Vandereycken.
Low-rank matrix completion by Riemannian optimization—extended version.
SIAM Journal on Optimization, 23(2):1214–1236, 2013.

Xinru Yuan, Wen Huang, P.-A. Absil, and Kyle A. Gallivan.
A Riemannian limited-memory BFGS algorithm for computing the matrix geometric mean.
Procedia Computer Science, 80:2147–2157, 2016.

Xinru Yuan, Wen Huang, P.-A. Absil, and K. A. Gallivan.
A Riemannian quasi-newton method for computing the Karcher mean of symmetric positive definite matrices.
Technical Report FSU17-02, Florida State University, 2017.
www.math.fsu.edu/ whuang2/papers/RMKMSPDM.htm.