Incremental gradient method for Karcher mean on symmetric cones

Sangho Kum and Sangwoon Yun

Department of Mathematics Education
Chungbuk National University
shkum@cbnu.ac.kr

2016 MAO (Workshop on Matrices and Operator)

July 5, 2016

Suites Hotel, Jeju

A brief history

A brief history

Definition 1. (Riemannian center of mass)

- (M, d): *an n-dimensional complete Riemannian manifold with distance* d *induced by the Riemannian structure.*

A brief history

Definition 1. (Riemannian center of mass)

- (M, d): *an n-dimensional complete Riemannian manifold with distance* d *induced by the Riemannian structure.*
- ν : *a probability measure on* M*.*

A brief history

Definition 1. (Riemannian center of mass)

- (M, d)*: an n-dimensional complete Riemannian manifold with distance d induced by the Riemannian structure.*
- ν : *a probability measure on* M.
- $$f_2(x) = \frac{1}{2} \int d^2(x, s) d\nu(s).$$

A brief history

Definition 1. (Riemannian center of mass)

- (M, d): *an n-dimensional complete Riemannian manifold with distance d induced by the Riemannian structure.*
- ν : *a probability measure on* M.
- $$f_2(x) = \frac{1}{2} \int d^2(x, s) d\nu(s).$$
- *Any minimizer of* f_2 *is called a Riemannian* L^2 *center of mass with respect to* ν.

1. Riemannian center of mass or Karcher mean

- E. Cartan : In 1920s, the Riemannian L^2 center of mass in an Hadamard manifold (the first one in the context of Riemannian geometry)

1. Riemannian center of mass or Karcher mean

- E. Cartan : In 1920s, the Riemannian L^2 center of mass in an Hadamard manifold (the first one in the context of Riemannian geometry)

- Existence and Uniqueness.

1. Riemannian center of mass or Karcher mean

- E. Cartan : In 1920s, the Riemannian L^2 center of mass in an Hadamard manifold (the first one in the context of Riemannian geometry)
- Existence and Uniqueness.
- Any compact subgroup of the isometry group of an Hadamard manifold has a fixed point.

1. Riemannian center of mass or Karcher mean

- E. Cartan : In 1920s, the Riemannian L^2 center of mass in an Hadamard manifold (the first one in the context of Riemannian geometry)
- Existence and Uniqueness.
- Any compact subgroup of the isometry group of an Hadamard manifold has a fixed point.

1. Riemannian center of mass or Karcher mean

- H. Karcher : The Riemannian L^2 center of mass in general Riemannian manifolds but for probability measures with support in small enough balls

1. Riemannian center of mass or Karcher mean

- H. Karcher : The Riemannian L^2 center of mass in general Riemannian manifolds but for probability measures with support in small enough balls
- He enlarged the domain of existence and uniqueness and considered new applications.

1. Riemannian center of mass or Karcher mean

- H. Karcher : The Riemannian L^2 center of mass in general Riemannian manifolds but for probability measures with support in small enough balls
- He enlarged the domain of existence and uniqueness and considered new applications.
- More recently, the Karcher mean has found applications in many applied fields:

1. Riemannian center of mass or Karcher mean

- H. Karcher : The Riemannian L^2 center of mass in general Riemannian manifolds but for probability measures with support in small enough balls
- He enlarged the domain of existence and uniqueness and considered new applications.
- More recently, the Karcher mean has found applications in many applied fields:
- computer vision, statistical analysis of shapes, medical imaging, sensor networks, data analysis applications, and so an.

1. Riemannian center of mass or Karcher mean

- H. Karcher : The Riemannian L^2 center of mass in general Riemannian manifolds but for probability measures with support in small enough balls
- He enlarged the domain of existence and uniqueness and considered new applications.
- More recently, the Karcher mean has found applications in many applied fields:
- computer vision, statistical analysis of shapes, medical imaging, sensor networks, data analysis applications, and so an.

Existing methodologies

Existing methodologies

- In these applied settings, an important problem is to numerically approximate or compute the Karcher mean.

Existing methodologies

- In these applied settings, an important problem is to numerically approximate or compute the Karcher mean.

- Gradient descent methods

Existing methodologies

- In these applied settings, an important problem is to numerically approximate or compute the Karcher mean.
- Gradient descent methods
- Proximal point methods(or incremental proximal methods)

Existing methodologies

- In these applied settings, an important problem is to numerically approximate or compute the Karcher mean.
- Gradient descent methods
- Proximal point methods(or incremental proximal methods)
- Newton methods

Existing methodologies

- In these applied settings, an important problem is to numerically approximate or compute the Karcher mean.
- Gradient descent methods
- Proximal point methods(or incremental proximal methods)
- Newton methods

Points to consider

Points to consider

- But, even though numerical algorithms were developed in general Riemannian manifolds circumstances or more, some of them are not numerically implementable in a practical sense.

Points to consider

- But, even though numerical algorithms were developed in general Riemannian manifolds circumstances or more, some of them are not numerically implementable in a practical sense.
- It is quite often that algorithms on Riemannian manifolds seem to be conceptual when we consider that applications are mainly concentrated on matrix cases.

Points to consider

- But, even though numerical algorithms were developed in general Riemannian manifolds circumstances or more, some of them are not numerically implementable in a practical sense.
- It is quite often that algorithms on Riemannian manifolds seem to be conceptual when we consider that applications are mainly concentrated on matrix cases.
- In comparison to this, to develop algorithms is more tangible in symmetric cone settings as in the case of the positive semidefinite cone.

Points to consider

- But, even though numerical algorithms were developed in general Riemannian manifolds circumstances or more, some of them are not numerically implementable in a practical sense.
- It is quite often that algorithms on Riemannian manifolds seem to be conceptual when we consider that applications are mainly concentrated on matrix cases.
- In comparison to this, to develop algorithms is more tangible in symmetric cone settings as in the case of the positive semidefinite cone.
- This is a main reason why we work under the framework of symmetric cones.

Jordan algebra

Jordan algebra

- A *Jordan algebra* V over $\mathbb{R}$ is a (non-associative) commutative algebra satisfying $x^2(xy) = x(x^2y)$ for all $x, y \in V$.

Jordan algebra

- A *Jordan algebra* V over $\mathbb{R}$ is a (non-associative) commutative algebra satisfying $x^2(xy) = x(x^2y)$ for all $x, y \in V$.
- For $x \in V$, let $L(x)$ be the linear operator defined by $L(x)y = xy$, and let $P(x) = 2L(x)^2 - L(x^2)$. The map P is called the quadratic representation of V.

Jordan algebra

- A *Jordan algebra* V over $\mathbb{R}$ is a (non-associative) commutative algebra satisfying $x^2(xy) = x(x^2y)$ for all $x, y \in V$.
- For $x \in V$, let $L(x)$ be the linear operator defined by $L(x)y = xy$, and let $P(x) = 2L(x)^2 - L(x^2)$. The map P is called the quadratic representation of V.
- An element $x \in V$ is said to be invertible if there exists an element y (denoted by $y = x^{-1}$) in the subalgebra generated by x and e (the Jordan identity) such that $xy = e$.

Euclidean Jordan algebra (finite dim′l symmetric cone)

Euclidean Jordan algebra (finite dim'l symmetric cone)

- A finite dimensional Jordan algebra V is said to be *Euclidean* if there exists an inner product $\langle \cdot, \cdot \rangle$ such that

$$\langle xy, z \rangle = \langle y, xz \rangle \tag{1.1}$$

for all $x, y, z \in V$.

Euclidean Jordan algebra (finite dim'l symmetric cone)

- A finite dimensional Jordan algebra V is said to be *Euclidean* if there exists an inner product $\langle \cdot, \cdot \rangle$ such that

$$\langle xy, z \rangle = \langle y, xz \rangle \tag{1.1}$$

for all $x, y, z \in V$.

- An element $c \in V$ is called an idempotent if $c^2 = c$. We say that $c_1, \ldots, c_k$ is a complete system of orthogonal idempotents if $c_i^2 = c_i, c_i c_j = 0, i \neq j, c_1 + \cdots + c_k = e$. An idempotent is primitive if it is non-zero and cannot be written as the sum of two non-zero idempotents. A Jordan frame is a complete system of primitive idempotents.

- Let Q be the set of all square elements of V. Then Q is a closed convex cone of V with $Q \cap -Q = \{0\}$, and is the set of element $x \in V$ such that $L(x)$ is positive semi-definite.

- Let Q be the set of all square elements of V. Then Q is a closed convex cone of V with $Q \cap -Q = \{0\}$, and is the set of element $x \in V$ such that $L(x)$ is positive semi-definite.

- It turns out that Q has non-empty interior Ω, and Ω is a symmetric cone, that is, the group

$$G(\Omega) = \{g \in \mathrm{GL}(V) | g(\Omega) = \Omega\}$$

acts transitively on it and Ω is a self-dual cone with respect to the inner product $\langle \cdot | \cdot \rangle$.

- Let Q be the set of all square elements of V. Then Q is a closed convex cone of V with $Q \cap -Q = \{0\}$, and is the set of element $x \in V$ such that $L(x)$ is positive semi-definite.
- It turns out that Q has non-empty interior Ω, and Ω is a symmetric cone, that is, the group

$$G(\Omega) = \{g \in \mathrm{GL}(V) | g(\Omega) = \Omega\}$$

acts transitively on it and Ω is a self-dual cone with respect to the inner product $\langle \cdot | \cdot \rangle$.

- Furthermore, for any a in Ω, $P(a) \in G(\Omega)$ and is positive definite.

Two typical examples

Two typical examples

- Second-order cone(SOC) is the closed convex cone

$$\mathcal{K} := \left\{(x_1, x_2) \in \mathbb{R} \times \mathbb{R}^{n-1} \mid \|x_2\| \leq x_1\right\}.$$

Two typical examples

- Second-order cone(SOC) is the closed convex cone

$$\mathcal{K} := \left\{(x_1, x_2) \in \mathbb{R} \times \mathbb{R}^{n-1} \mid \|x_2\| \leq x_1\right\}.$$

- The Euclidean space $\mathbb{R}^n$ with the Jordan product defined by

$$x \circ y = (\langle x, y\rangle,\ x_1 y_2 + y_1 x_2)$$

is a Euclidean Jordan algebra equipped with the standard inner product $\langle \cdot, \cdot \rangle$ where $x = (x_1, x_2),\ y = (y_1, y_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$.

Two typical examples

- Second-order cone(SOC) is the closed convex cone

$$\mathcal{K} := \left\{ (x_1, x_2) \in \mathbb{R} \times \mathbb{R}^{n-1} \mid \|x_2\| \le x_1 \right\}.$$

- The Euclidean space $\mathbb{R}^n$ with the Jordan product defined by

$$x \circ y = (\langle x, y \rangle,\ x_1 y_2 + y_1 x_2)$$

  is a Euclidean Jordan algebra equipped with the standard inner product $\langle \cdot, \cdot \rangle$ where $x = (x_1, x_2),\ y = (y_1, y_2) \in \mathbb{R} \times \mathbb{R}^{n-1}$.

- $\mathcal{K}$ is the corresponding symmetric cone of the Euclidean Jordan algebra $\mathbb{R}^n$.

- Let $\mathbb{S}^n$ be the algebra of $n \times n$ real symmetric matrices with the Jordan product defined by

$$X \circ Y = \frac{XY + YX}{2}$$

where XY is the usual matrix multiplication of X and Y.

1. Riemannian center of mass or Karcher mean

- Let $\mathbb{S}^n$ be the algebra of $n \times n$ real symmetric matrices with the Jordan product defined by

$$X \circ Y = \frac{XY + YX}{2}$$

where XY is the usual matrix multiplication of X and Y.

- Then $\mathbb{S}^n$ is a Euclidean Jordan algebra equipped with the trace inner product

$$\langle X, Y \rangle = \mathrm{tr}(XY), \; P(X)Y = XYX.$$

1. Riemannian center of mass or Karcher mean

- Let $\mathbb{S}^n$ be the algebra of $n \times n$ real symmetric matrices with the Jordan product defined by

$$X \circ Y = \frac{XY + YX}{2}$$

where XY is the usual matrix multiplication of X and Y.

- Then $\mathbb{S}^n$ is a Euclidean Jordan algebra equipped with the trace inner product

$$\langle X, Y \rangle = \mathrm{tr}(XY), \;\; P(X)Y = XYX.$$

- PD is the corresponding symmetric cone of the Euclidean Jordan algebra $\mathbb{S}^n$.

Symmetric cone setting

Symmetric cone setting

- The symmetric cone Ω admits a Riemannian metric defined by

$$\langle u, v\rangle_x = \langle P(x)^{-1}u, v\rangle, x \in \Omega, u, v \in V.$$

Symmetric cone setting

- The symmetric cone Ω admits a Riemannian metric defined by

$$\langle u, v\rangle_x = \langle P(x)^{-1}u, v\rangle, x \in \Omega, u, v \in V.$$

- The Riemannian distance $\delta(a, b)$ is given by

$$\delta(a, b) == \| \log P(a^{-\frac{1}{2}})b\| = \left(\sum_{i=1}^{r} \log^2 \lambda_i(P(a^{-1/2})b)\right)^{1/2}.$$

Symmetric cone setting

- The symmetric cone Ω admits a Riemannian metric defined by

$$\langle u, v\rangle_x = \langle P(x)^{-1}u, v\rangle, x \in \Omega, u, v \in V.$$

- The Riemannian distance $\delta(a, b)$ is given by

$$\delta(a, b) == \| \log P(a^{-\frac{1}{2}})b\| = \left(\sum_{i=1}^{r} \log^2 \lambda_i(P(a^{-1/2})b)\right)^{1/2}.$$

- The unique geodesic curve joining a and b is

$$t \mapsto a\#_t b := P(a^{1/2})(P(a^{-1/2})b)^t.$$

1. Riemannian center of mass or Karcher mean

- An important property of the metric δ is the semiparallelogram law

$$\delta^2(z, x\#y) \leq \frac{1}{2}\delta^2(z,x) + \frac{1}{2}\delta^2(z,y) - \frac{1}{4}\delta^2(x,y)$$

and its general form for any $t \in [0,1]$

$$\delta^2(z, x\#_t y) \leq (1-t)\delta^2(z,x) + t\delta^2(z,y) - t(1-t)\delta^2(x,y).$$

- An important property of the metric δ is the semiparallelogram law

$$\delta^2(z, x\#y) \leq \frac{1}{2}\delta^2(z,x) + \frac{1}{2}\delta^2(z,y) - \frac{1}{4}\delta^2(x,y)$$

and its general form for any $t \in [0,1]$

$$\delta^2(z, x\#_t y) \leq (1-t)\delta^2(z,x) + t\delta^2(z,y) - t(1-t)\delta^2(x,y).$$

- The Riemannian manifold (Ω, δ) is an important example of an Hadamard manifold.

- In this circumstances, the Karcher mean reduces to the following:

1. Riemannian center of mass or Karcher mean

- In this circumstances, the Karcher mean reduces to the following:

Definition 2. (Karcher mean in symmetric cones)

The Karcher mean of $a_1, \ldots, a_n \in \Omega$ is defined to be the unique minimizer of the sum of squares of the Riemannian distances to each of the a_i, i.e.,

$$\Lambda(a_1, \ldots, a_n) = \underset{x \in \Omega}{\arg\min} \frac{1}{2} \sum_{i=1}^{n} \delta^2(x, a_i).$$

- In this circumstances, the Karcher mean reduces to the following:

Definition 2. (Karcher mean in symmetric cones)

The Karcher mean of $a_1, \ldots, a_n \in \Omega$ *is defined to be the unique minimizer of the sum of squares of the Riemannian distances to each of the* a_i*, i.e.,*

$$\Lambda(a_1, \ldots, a_n) = \underset{x \in \Omega}{\arg\min} \frac{1}{2} \sum_{i=1}^{n} \delta^2(x, a_i).$$

Motivations

Motivations

- The objective function of the aforementioned minimization problem is the sum of many functions, i.e., the squares of Riemannian distance functions with given data $a_1, \ldots, a_n \in \Omega$:

Motivations

- The objective function of the aforementioned minimization problem is the sum of many functions, i.e., the squares of Riemannian distance functions with given data $a_1, \ldots, a_n \in \Omega$:

$$\min_{x \in \Omega} f(x) := \sum_{i=1}^{m} f_i(x), \tag{2.1}$$

where $f_i(x) = \frac{1}{2}\delta(x, a_i)^2$ with a_i's and x in Ω.

Motivations

- The objective function of the aforementioned minimization problem is the sum of many functions, i.e., the squares of Riemannian distance functions with given data $a_1, \ldots, a_n \in \Omega$:

$$\min_{x \in \Omega} f(x) := \sum_{i=1}^{m} f_i(x), \tag{2.1}$$

where $f_i(x) = \frac{1}{2}\delta(x, a_i)^2$ with a_i's and x in Ω.

- It is observed that the solution of the problem belongs to a bounded set $\mathcal{D} = \{x \in \Omega \mid \alpha e \leq x \leq \beta e\}$, where $0 < \alpha \leq \beta$, that contains $\{a_1, \ldots, a_n\}$.

Problem formulation

Problem formulation

- Thus, we consider the following bound constained minimization problem formulation of :

$$\min_{x \in \mathcal{D}} \ f(x). \tag{2.2}$$

Problem formulation

- Thus, we consider the following bound constained minimization problem formulation of :

$$\min_{x \in \mathcal{D}} \ f(x). \tag{2.2}$$

- This problem formulation motivates us to adapt an incrementally updated gradient(IUG) method to solve the problem.

Problem formulation

- Thus, we consider the following bound constained minimization problem formulation of :

$$\min_{x\in\mathcal{D}} f(x). \tag{2.2}$$

- This problem formulation motivates us to adapt an incrementally updated gradient(IUG) method to solve the problem.

- To our knowledge, this IUG method is not adopted to deal with the problem of finding the Karcher mean yet.

Incremental gradient(IG) method

- In the case that the number of f_i's consisting of the objective function $f = \sum_{i=1}^{n} f_i$ is large, traditional gradient method would be inefficient since they require evaluating all the gradients of f_i's before updating the iterate.

Incremental gradient(IG) method

- In the case that the number of f_i's consisting of the objective function $f = \sum_{i=1}^{n} f_i$ is large, traditional gradient method would be inefficient since they require evaluating all the gradients of f_i's before updating the iterate.
- Incremental gradient methods, in contrast, update the iterate after evaluation of gradients for only one or a few smooth functions.

Incremental gradient(IG) method

- In the case that the number of f_i's consisting of the objective function $f = \sum_{i=1}^{n} f_i$ is large, traditional gradient method would be inefficient since they require evaluating all the gradients of f_i's before updating the iterate.
- Incremental gradient methods, in contrast, update the iterate after evaluation of gradients for only one or a few smooth functions.
- Blatt et al. proposed a method that computes the gradient of a single component function at each iteration, but instead of updating the iterate using this gradient, it uses the sum of n most recently computed gradients for the unconstrained smooth minimization case.

Incremental gradient(IG) method

- In the case that the number of f_i's consisting of the objective function $f = \sum_{i=1}^{n} f_i$ is large, traditional gradient method would be inefficient since they require evaluating all the gradients of f_i's before updating the iterate.
- Incremental gradient methods, in contrast, update the iterate after evaluation of gradients for only one or a few smooth functions.
- Blatt et al. proposed a method that computes the gradient of a single component function at each iteration, but instead of updating the iterate using this gradient, it uses the sum of n most recently computed gradients for the unconstrained smooth minimization case.

3. Incrementally updated gradient(IUG) method

- Assuming the uniform boundedness and Lipschitz continuity of all the gradients of f_i's as well as the uniqueness of a stationary point and positive definiteness of Hessian of f at the stationary point, the global convergence of this method with a sufficiently small stepsize is shown.

3. Incrementally updated gradient(IUG) method

- Assuming the uniform boundedness and Lipschitz continuity of all the gradients of f_i's as well as the uniqueness of a stationary point and positive definiteness of Hessian of f at the stationary point, the global convergence of this method with a sufficiently small stepsize is shown.

- Blatt's method may be viewed as belonging to a general class of gradient methods that update the gradients for only one or a few f_i's at a time, which we call *incrementally updated gradient(IUG) method*.

3. Incrementally updated gradient(IUG) method

IUG method

- Recently, Tseng and Yun proposed two IUG methods to solve the nonsmooth minimization problem whose objective is the sum of n smooth functions and a convex function. They showed the global convergence for the IUG method using a constant step size, assuming only the Lipschitz continuity of each gradient of n smooth functions.

IUG method

- Recently, Tseng and Yun proposed two IUG methods to solve the nonsmooth minimization problem whose objective is the sum of n smooth functions and a convex function. They showed the global convergence for the IUG method using a constant step size, assuming only the Lipschitz continuity of each gradient of n smooth functions.
- Compared to Blatt's one, IUG method is a more general one for solving a more general problem, and the global convergence is shown under much weaker assumptions.

IUG method

- Recently, Tseng and Yun proposed two IUG methods to solve the nonsmooth minimization problem whose objective is the sum of n smooth functions and a convex function. They showed the global convergence for the IUG method using a constant step size, assuming only the Lipschitz continuity of each gradient of n smooth functions.
- Compared to Blatt's one, IUG method is a more general one for solving a more general problem, and the global convergence is shown under much weaker assumptions.
- The second IUG method uses adaptive stepsizes and hence more practical, and it has a similar global convergence property as the first IUG method.

IUG method

- Recently, Tseng and Yun proposed two IUG methods to solve the nonsmooth minimization problem whose objective is the sum of n smooth functions and a convex function. They showed the global convergence for the IUG method using a constant step size, assuming only the Lipschitz continuity of each gradient of n smooth functions.
- Compared to Blatt's one, IUG method is a more general one for solving a more general problem, and the global convergence is shown under much weaker assumptions.
- The second IUG method uses adaptive stepsizes and hence more practical, and it has a similar global convergence property as the first IUG method.

3. Incrementally updated gradient(IUG) method

- They generalized the previous IG method to handle constraints and nonsmooth regularization and proved the global convergence under much weaker assumptions.
- For this IUG method, in the present paper, we work under the standard framework of Euclidean spaces rather than Riemannian circumstances from a theoretical viewpoint.
- This is mainly due to the fact that an addition of vectors in different tangent spaces of Riemannian manifolds is not possible. Even if it were possible using a parallel transport, it may have no practical meaning in computation.
- At present, it seems to be difficult to consider an effective IG method in a fully Riemannian sense on a symmetric cone.

IUG method for Karcher mean

IUG method for Karcher mean

- The IUG method due to Tseng and Yun exactly fits to deal with the Karcher mean approximation (2.1) and (2.2) where the numbers of the smooth $f_i = \frac{1}{2}\delta(x, a_i)^2$ is large.

IUG method for Karcher mean

- The IUG method due to Tseng and Yun exactly fits to deal with the Karcher mean approximation (2.1) and (2.2) where the numbers of the smooth $f_i = \frac{1}{2}\delta(x, a_i)^2$ is large.

- The following fact plays a key role in the present work:

IUG method for Karcher mean

- The IUG method due to Tseng and Yun exactly fits to deal with the Karcher mean approximation (2.1) and (2.2) where the numbers of the smooth $f_i = \frac{1}{2}\delta(x, a_i)^2$ is large.

- The following fact plays a key role in the present work:

Proposition 1.

$$\|\nabla f_i(y) - \nabla f_i(z)\| \leq L_i \|y - z\| \quad \forall y, z \in \mathcal{D}, \tag{4.1}$$

for some $L_i \geq 0$, $i = 1, \ldots, m$. *Let* $L = \sum_{i=1}^{m} L_i$.

IUG method for Karcher mean

- The IUG method due to Tseng and Yun exactly fits to deal with the Karcher mean approximation (2.1) and (2.2) where the numbers of the smooth $f_i = \frac{1}{2}\delta(x, a_i)^2$ is large.

- The following fact plays a key role in the present work:

Proposition 1.

$$\|\nabla f_i(y) - \nabla f_i(z)\| \leq L_i \|y - z\| \quad \forall y, z \in \mathcal{D}, \tag{4.1}$$

for some $L_i \geq 0$, $i = 1, \ldots, m$. *Let* $L = \sum_{i=1}^{m} L_i$.

4. IUG method for Karcher mean

- Moreover, we give an assumption as follows:

4. IUG method for Karcher mean

- Moreover, we give an assumption as follows:

Assumption.

$\tau_i^k \geq k - K$ *for all* i *and* k, *where* $K \geq 0$ *is an integer.*

4. IUG method for Karcher mean

- Moreover, we give an assumption as follows:

Assumption.

$\tau_i^k \geq k - K$ *for all* i *and* k*, where* $K \geq 0$ *is an integer.*

- Assumption ensures that the gradient of f_i is updated at least once for every $K + 1$ consecutive iterations.

Algorithms

Algorithms

Algorithm 1.

Choose $x^0, x^{-1}, \cdots \in \mathcal{D}$ *and* $t \in]0,1]$. *Initialize* $k = 0$. *Update* $x^{(k+1)}$ *from* x^k *by the following template:*

Algorithms

Algorithm 1.

Choose $x^0, x^{-1}, \cdots \in \mathcal{D}$ *and* $t \in]0,1]$. *Initialize* $k = 0$. *Update* $x^{(k+1)}$ *from* x^k *by the following template:*

Step 1. *Choose* $0 \leq \tau_i^k \leq k$ *for* $i = 1, \ldots, m$,

Step 2. *Update* g^k *by*

$$g^k = \sum_{i=1}^{m} \nabla f_i(x^{\tau_i^k}). \tag{4.2}$$

Step 3. *Find* d^k *by using*

$$d^k = \arg\min_{d \in V, x^k + d \in \mathcal{D}} \left\{ \langle g^k, d \rangle + \frac{1}{2} \|d\|^2 \right\}. \tag{4.3}$$

Step 4. $x^{k+1} = x^k + t d^k$.

- The above framework is quite flexible and allows partially asynchronous updating of the component gradients.

4. IUG method for Karcher mean

- The above framework is quite flexible and allows partially asynchronous updating of the component gradients.
- In the following lemma, we give a descent property of the minimization subproblem (4.3) for finding a search direction.

- The above framework is quite flexible and allows partially asynchronous updating of the component gradients.
- In the following lemma, we give a descent property of the minimization subproblem (4.3) for finding a search direction.

Lemma 1.

For any $x \in \mathcal{D}$, and $g \in V$, let d_g be the solution of the problem

$$\min_{d \in V, x+d \in \mathcal{D}} \left\{ \langle g, d \rangle + \frac{1}{2}\|d\|^2 \right\}.$$

Then

$$\langle g, d \rangle + \frac{1}{2}\|d\|^2 \leq -\frac{1}{2}\|d_g\|^2 \quad \text{or} \quad \langle g, d_g \rangle \leq -\|d_g\|^2. \tag{4.4}$$

4. IUG method for Karcher mean

- An $x \in V$ is *a stationary point* of f if $x \in \mathcal{D}$ and $f'(x; d) \geq 0$ for all $d \in V$.
- The following result characterizes stationarity in terms of $d_{\nabla f(x)}$.

4. IUG method for Karcher mean

- An $x \in V$ is *a stationary point* of f if $x \in \mathcal{D}$ and $f'(x; d) \geq 0$ for all $d \in V$.
- The following result characterizes stationarity in terms of $d_{\nabla f(x)}$.

Lemma 2.

An $x \in \mathcal{D}$ is a stationary point of f if and only if $d_{\nabla f(x)} = 0$.

4. IUG method for Karcher mean

- Now, we have the following global convergence result for the method with a sufficiently small constant stepsize.

- Now, we have the following global convergence result for the method with a sufficiently small constant stepsize.

Theorem 1. (Constant Stepsize Case)

Let $\{x^k\}$ and $\{d^k\}$ be sequences generated by Algorithm 1 under Assumption, and with $t < \frac{2}{L(2K+1)}$. Then $\{d^k\} \to 0$ and every cluster point of $\{x^k\}$ is a stationary point.

4. IUG method for Karcher mean

- We describe the second IG method with adaptive stepsize below.

- We describe the second IG method with adaptive stepsize below.

Algorithm 2.

Choose $x^0, x^{-1}, \ldots \in \mathcal{D}$, $\underline{t} \in]0,1]$, $\beta \in]0,1[$, *and* $\sigma > \frac{1}{2}$. *Initialize* $k = 0$. *Update* $x^{(k+1)}$ *from* x^k *by the following template:*

Step 1. *Choose* $0 \leq \tau_i^k \leq k$ *for* $i = 1, \ldots, m$,

Step 2. *Update* g^k *by (4.2)*

Step 3. *Find* d^k *by using (4.3)*

Step 4. *Choose* $t_k^{\mathrm{init}} \in [\underline{t}, 1]$ *and let* t_k *be the largest element of* $\{t_k^{\mathrm{init}} \beta^j\}_{j=0,1,\ldots}$ *satisfying*

$$f(x^k + t_k d^k) - f(x^k) \leq -\sigma K L \|t_k d^k\|^2 + \frac{L}{2} \sum_{j=(k-K)_+}^{k-1} \|t_j d^j\|^2 \tag{4.5}$$

Step 5. $x^{k+1} = x^k + t d^k$.

- The stepsize t_k in the first IUG method is adaptively selected by decreasing t_k whenever the condition (4.5) is violated.
- In practice, the Lipschitz constant L is not given a priori but we are able to estimate L by increasing L by a certain positive factor whenever the condition (4.5) is not satisfied with starting at an arbitrary estimate of L.
- When t_k is below $\bar{t}$ defined in Theorem 2 below, the condition (4.5) is satisfied with some constant L. Whether L is defined by Proposition 1 is irrelevant.

Theorem 2. (Adaptive Stepsize Case)

Let $\{x^k\}$, $\{d^k\}$, $\{t_k\}$ *be sequences generated by Algorithm 2 under Assumption 1. Then the following results hold.*

(a) *For each* $k \geq 0$, *(4.5) holds whenever* $t_k \leq \bar{t}$, *where* $\bar{t} = \frac{2}{L(2\sigma K + K + 1)}$.

(b) *We have* $t_k \geq \min\{\underline{t}, \beta\bar{t}\}$ *for all* k.

(c) $\{d^k\} \to 0$ *and every cluster point of* $\{x^k\}$ *is a stationary point.*

Conclusions

Conclusions

- In this paper we consider IUG method for the Karther mean motivated by the observations that implementable algorithms of finding the Karcher mean on general settings beyond matrix case are not as many as expected, and the objective function of the considered minimization problem is the sum of many smooth functions.

Conclusions

- In this paper we consider IUG method for the Karther mean motivated by the observations that implementable algorithms of finding the Karcher mean on general settings beyond matrix case are not as many as expected, and the objective function of the considered minimization problem is the sum of many smooth functions.
- We have shown the global convergence of the proposed methods exploiting the Lipschitz continuity of the gradient of the objective function.

Conclusions

- In this paper we consider IUG method for the Karther mean motivated by the observations that implementable algorithms of finding the Karcher mean on general settings beyond matrix case are not as many as expected, and the objective function of the considered minimization problem is the sum of many smooth functions.
- We have shown the global convergence of the proposed methods exploiting the Lipschitz continuity of the gradient of the objective function.
- Even though our method is faster than SD, we need to further accelerate the proposed method so that it becomes more attractive.

4. IUG method for Karcher mean

- Two directions may be taken into account.

4. IUG method for Karcher mean

- Two directions may be taken into account.
- First, a tight bound for Lipschitz constant or a scheme is necessary for adjusting the better stepsize without evaluating the objective value.

- Two directions may be taken into account.
- First, a tight bound for Lipschitz constant or a scheme is necessary for adjusting the better stepsize without evaluating the objective value.
- Second, a fully Riemannian version of the proposed incremental gradient method can be a better alternative.

References

[1] Karcher, H.: Riemannian center of mass and mollifier smoothing. Comm. Pure Appl. Math. **30**, 509–541 (1977)

[2] Afsari, B., Tron, R., Vidal, R.: On the convergence of gradient descent for finding the Riemannian center of mass. SIAM J. Control Optim. **51**, 2230–2260 (2013)

[3] Tseng, P., Yun, S.: Incrementally updated gradient methods for constrained and regularized optimization. J. Optim. Theory Appl. **160**, 832–853 (2014)

[4] Kum, S., Lee, H., Lim, Y.: No dice theorem on symmetric cones. Taiwanese J. Math. **17**, 1967–1982 (2013)

[5] Holbrook, J.: No dice: a determinstic approach to the Cartan centroid. J. Ramanujan Math. Soc. **27**, 509-521 (2012)