Medians of probability measures in Riemannian manifolds and applications to radar target detection

Le Yang

THÈSE

pour l'obtention du grade de

Docteur de l'Université de Poitiers

spécialité : **Mathématiques et leurs intéractions**

préparée au **Laboratoire de Mathématiques et Applications, UMR 6086**

dans le cadre de l'École Doctorale **Science de l'Ingénieur et l'Information S2I**

présentée et soutenue publiquement le 15 décembre 2011 par

Le YANG

Médianes de mesures de probabilité dans les variétés riemanniennes et applications à la détection de cibles radar

---

Directeur de thèse : **Marc ARNAUDON**
Co-directeur de thèse : **Frédéric BARBARESCO**

JURY

Huiling LE	Professeur, University of Nottingham, UK	Rapporteur
Xavier PENNEC	Directeur de Recherche, INRIA Sophia Antipolis	Rapporteur
Michel ÉMERY	Directeur de Recherche CNRS, IRMA	Examinateur
Jean PICARD	Professeur, Université Blaise Pascal	Examinateur
Marc ARNAUDON	Professeur, Université de Poitiers	Directeur
Frédéric BARBARESCO	Expert Radar, Thales Air Systems	Co-directeur

Médianes de mesures de probabilité dans les variétés riemanniennes et applications à la détection de cibles radar

Résumé

Dans cette thèse, nous étudierons les médianes d'une mesure de probabilité dans une variété riemannienne. Dans un premier temps, l'existence et l'unicité des médianes locales seront montrées. Afin de calculer les médianes aux cas pratiques, nous proposerons aussi un algorithme de sous-gradient et prouverons sa convergence. Ensuite, les médianes de Fréchet seront étudiées. Nous montrerons leur cohérence statistique et donnerons des estimations quantitatives de leur robustesse à l'aide de courbures. De plus, nous montrerons que, dans les variétés riemanniennes compactes, les médianes de Fréchet de données génériques sont toujours uniques. Des algorithmes stochastiques et déterministes seront proposés pour calculer les p-moyennes de Fréchet dans les variétés riemanniennes. Un lien entre les médianes et les problèmes de points fixes sera aussi montré. Finalement, nous appliquerons les médiane et la géométrie riemannienne des matrices de covariance Toeplitz à la détection de cible radar.

Mots-clés: statistiques robustes, moyenne, données sphériques, variétés de Riemann, théorème du point fixe, matrices de Toeplitz, analyse de covariance

---

Medians of probability measures in Riemannian manifolds and applications to radar target detection

Abstract

In this thesis, we study the medians of a probability measure in a Riemannian manifold. Firstly, the existence and uniqueness of local medians are proved. In order to compute medians in practical cases, we also propose a subgradient algorithm and prove its convergence. After that, Fréchet medians are considered. We prove their statistical consistency and give some quantitative estimations of their robustness with the aid of curvatures. Moreover, we show that, in compact Riemannian manifolds, the Fréchet medians of generic data points are always unique. Some stochastic and deterministic algorithms are proposed for computing Riemannian p-means. A connection between medians and fixed point problems are also given. Finally, we apply the medians and the Riemannian geometry of Toeplitz covariance matrices to radar target detection.

Key words: median, barycenter, Riemannian manifold, fixed point theorem, Toeplitz covariance matrices

Remerciements

Tout d'abord, je tiens à exprimer ma profonde reconnaissance à mes directeurs de thèse Marc Arnaudon et Frédéric Barbaresco pour m'avoir proposé ce sujet de thèse, ainsi que pour m'avoir fait bénéficier de leurs expériences et de leurs compétences avec tant de patiences et de rigueurs pendant ces trois dernières années. J'adresse particulièrement mes sincères remerciements à Marc pour ses éveils avec énormément de patience, ses encouragements réconfortant et sa sollicitude constante qui m'ont fortement aidé et stimulé. Je remercie Frédéric vivement pour ses conseils et ses références hebdomadaires qui ont été indispensables à l'avancement de la thèse.

Je suis très reconnaissant envers Huiling Le et Xavier Pennec pour m'avoir fait l'honneur d'accepter d'être les rapporteurs de mon travail.

Je tiens à exprimer toute ma reconnaissance à Michel Émery et Jean Picard pour leur participation à mon jury.

Je souhaite exprimer ma gratitude à Pierre Torasso pour son aide et ses encouragements qui m'ont été précieux pendant mon séjour à Poitiers.

Mes remerciements vont également au directeur du laboratoire Pol Vanhaecke qui m'a accueilli dans des conditions de travail très agréables. Sa gentillesse, sa disponibilité et son humour ont rendu tout problème facile.

Je désirais remercier Hervé Sabourin et Christian Olivier pour m'avoir donné l'occasion de continuer mes études en France, ainsi que pour m'avoir beaucoup aidé depuis mon arrivé à Poitiers.

Je remercie également Clément Dombry et Anthony Phan pour notre collaboration inoubliable. Mes remerciements vont également à Abderrazak Bouaziz pour son encouragement. J'aimerais aussi remercier Alain Miranville, Anne Bertrand, Morgan Pierre, Arnaud Rougirel, Lionel Ducos, Rupert Yu, Frédéric Bosio et tous les autres membres du laboratoire pour l'ambiance très agréable et respectueuse.

Mes remerciements s'adressent également à Jocelyne Attab, Natalie Mongin, Brigitte Brault, Natalie Marlet et Benoît Métrot pour leur gentillesse, leur patience et leur efficacité qui ont rendu mes difficultés administratives, informatiques et de documentations beaucoup plus légères.

Merci à mes amis et camarades doctorants: Frédéric, Anis, Jule, Caroline Pintoux, Caroline Lemarie, Hélène, Guilhem, Houssam, Haydi, Guilnard, Gang, Florent, Brice, Azariel, Khaoula, Kohélé et Armel.

Je remercie amicalement Qinghua, Xiujuan, Xinxin, Laoyu, Liwei, Shanshan, Jian, Huijie, Shaopeng et Zhongxun pour tous les bons moments passés ensemble.

Finalement, je ne pourrais jamais oublier ma famille pour leur soutient constant et leurs encouragements, particulièrement mon épouse Jia. Xmm, cette thèse est aussi la tienne.

Contents

Chapter 1

Introduction

1.1 A short review of previous work

The history of medians can be traced back to 1629 when P. Fermat initiated a challenge (see [39]): given three points in the plan, find a fourth one such that the sum of its distances to the three given points is minimum. The answer to this question, which was found firstly by E. Torricelli (see [84]) in 1647, is that if each angle of the triangle is smaller than $2\pi/3$, then the minimum point is such that the three segments joining it and the vertices of the triangle form three angles equal to $2\pi/3$; and in the opposite case, the minimum point is the vertex whose angle is greater than or equal to $2\pi/3$. This point is called the median or the Fermat-Torricelli point of the triangle. Naturally, the question of finding the triangle median can be generalized to finding a point that minimizes the sum of its distances to N given points in the plan and, if necessary, with weighted distances. This was proposed by J. Steiner in 1838 (see [81]) and was also considered by R. Sturm in 1884 (see [83]). The second author showed that the minimum point must be unique if the data points are not contained in a single line. This generalized Fermat's problem was also used by the economist A. Weber in 1909 (see [89]) as a mathematical model for the optimal location of a facility in order to serve several clients. From then on, the generalized Fermat's problem is also called Steiner's problem or Weber's problem or single facility location problem in economics and in the theory of operation research.

It is often very important to know the value of the median in practical cases. But in contrast to the barycenter, it is much more difficult to find the median of given points in the plan. For example, with the aid of Galois theory, C. Bajaj showed that (see [11]) if the number of generic data points is no less than five, then the generalized Fermat's problem cannot be solved by arithmetic operations and extraction of roots. The first algorithm intended to find medians in Euclidean spaces was proposed by E. Weiszfeld in 1937 (see [90]), whose original idea is to construct a sequence of weighted barycenters of data points which converges to the median. But this algorithm, which is essentially a gradient descent method, has a vital flaw that it does not converge if the median coincides with some data point, which is clearly an event of positive probability. The first completely convergent algorithm for computing medians was proposed by L. M. Ostresh in 1978 (see [73]), whose key ingredient is to eliminate the singular term when the algorithm hits a data point. Since then various algorithms for computing medians in Euclidean spaces were proposed and improved by many authors.

The notion of median also appears in statistics since a long time ago. In 1774, when P. S. Laplace tried to find an appropriate notion of the middle point for a group of observation values, he introduced in [61] "the middle of probability", the point that minimizes the sum of its absolute differences to data points, this is exactly the one dimensional median. In this work he also established that for a probability distribution on the real line with density f, the median m solves the equation

$$\int_{-\infty}^{m} f(x)dx = \frac{1}{2},$$

and the minimization problem

$$\min_{y\in\mathbf{R}} \int_{-\infty}^{+\infty} |x-y|f(x)dx.$$

A sufficiently general notion of median in metric spaces was proposed in 1948 by M. Fréchet in his famous article [48], where he defined a p-mean of a random variable X to be a point which minimizes the expectation of its distance at the power p to X. This flexible definition allows us to define various typical values, among which there are two important cases: $p = 1$ and $p = 2$, corresponding to the notions of median and mean, respectively.

Apparently, the median and mean are two notions of centrality for data points. As a result, one may wonder that which one is more advantageous? In fact, this is a longstanding debate which propelled the development of early statistics. Perhaps the perspective that the mean is more advantageous than the median is due to the choice of error laws in the early days. Laplace perceived the obvious advantage of the mean over the median when he used the normal distribution as the error law. But for other distributions, the normal distribution is not always an appropriate error law, as noted by F. Y. Edgeworth (see [37]) who insisted that one would lose less by taking the median instead of the mean if the error observations are not normally distributed. Another example, which is given by R. A. Fisher in [44], is that the empirical mean for a Cauchy distribution has the same accuracy as one single observation. Consequently, it would be better to use the median instead of the mean in this situation. Probably the most significant advantage of the median over the mean is that the former is robust but the latter is not, that is to say, the median is much less sensitive to outliers than the mean. Roughly speaking (see [66]), in order to move the median of a group of data points to arbitrarily far, at least a half of data points should be moved. On the contrary, in order to move the mean of a group of data points to arbitrarily far, it suffices to move one data point. So that medians are in some sense more prudent than means, as argued by M. Fréchet. The robustness property makes the median an important estimator in situations when there are lots of noise and disturbing factors.

The first application of the mean to Riemannian geometry is an ingenious artifice due to E. Cartan (see [30]), who used the uniqueness of means of measures on complete, simply connected Riemannian manifolds with non-positive curvature to show that every compact group of isometries of such manifolds has a common fixed point, which leads to the only known proof of the conjugacy of maximal compact subgroups of a semisimple Lie group. The first formal definition of means for probability measures on Riemannian manifolds was made by H. Karcher in [53] in order to generalize the mollifier smoothing techniques to the settings of Riemannian manifolds. To introduce Karcher's result concerning means, consider a Riemannian manifold M with Riemannian distance d and

$$B(a,\rho)=\{x\in M: d(x,a)<\rho\}$$

is a geodesic ball in M centered at a with a finite radius ρ. Let Δ be an upper bound of sectional curvatures in $\bar{B}(a,\rho)$ and inj be the injectivity radius of $\bar{B}(a,\rho)$. Under the following condition:

$$\rho<\min\Big\{\,\frac{\pi}{4\sqrt{\Delta}}\,,\,\frac{\mathrm{inj}}{2}\,\Big\},\tag{1.1}$$

where if $\Delta\le 0$, then $\pi/(4\sqrt{\Delta})$ is interpreted as $+\infty$, Karcher showed that, with the aid of estimations of Jacobi fields, the local energy functional

$$F_\mu:\quad \bar{B}(a,\rho)\longrightarrow \mathbf{R}_+,\quad x\longmapsto\int_M d^2(x,p)\mu(dp)\tag{1.2}$$

is strictly convex, thus it has a unique minimizer $b(\mu)$, which he called the Riemannian center of mass of the probability measure μ. Moreover, $b(\mu)$ is also the unique solution of the following equation:

$$\int_M \exp_x^{-1} p\ \mu(dp)=0_x,\quad x\in\bar{B}(a,\rho).\tag{1.3}$$

From then on, local means of probability measures on Riemannian manifolds are also called Karcher means, meanwhile, global means are often called Fréchet means. A rather general result concerning the uniqueness of local means was proved by W. S. Kendall in [55], where the condition (1.1) was replaced by the following much weaker one:

$$\rho<\min\Big\{\,\frac{\pi}{2\sqrt{\Delta}}\,,\,\mathrm{inj}\,(a)\,\Big\},\tag{1.4}$$

where inj (a) stands for the injectivity radius of the point a. But one should be aware that under the assumption (1.4) the ball $B(a,\rho)$ is only weakly convex: every pair of points in $B(a,\rho)$ can be joined by one and only one geodesic lying in $B(a,\rho)$. So that the distance $d(x,p)$ in (1.2) should be understood as the length of the geodesic joining x and p in $B(a,\rho)$, which is not necessarily the distance induced by the Riemannian metric of M in general. As a particular case of Kendall's result, the condition

$$\rho < \frac{1}{2}\min\left\{\frac{\pi}{\sqrt{\Delta}}, \mathrm{inj}\right\} \tag{1.5}$$

is sufficient to ensure the uniqueness of the Kacher means of μ. Furthermore, Kendall's method for proving the uniqueness of Karcher means is also very ingenious and interesting. His basic observation, which he owed to M. Emery, is a relationship between Karcher means and the convex functions on $B(a,\rho)$: for every Karcher mean x of μ, one has

$$\varphi(x) \leq \int_M \varphi(p)\mu(dp) \quad \text{for every convex function } \varphi \text{ on } B(a,\rho). \tag{1.6}$$

In view of this and the evident fact that the product of two Karcher means of μ is a Karcher mean of the product measure $\mu \otimes \mu$, the proof of the uniqueness of Karcher means of μ is reduced to the construction of some convex function on $B(a,\rho) \times B(a,\rho)$ which vanishes exactly on the diagonal. An example of such functions is given by Kendall in [55], detailed discussions on the construction can be found in [56]. Some generalizations of Karcher mean are given by many authors. For instance, M. Emery and G. Mokobodzki defined in [38] the exponential barycenters and convex barycenters for measures on affine manifolds with the aid of (1.3) and (1.6), respectively. They also showed that a point x is a convex barycenter of a probability μ if and only if there exists a continuous martingale starting from x with terminal law μ. The uniqueness of exponential barycenters are generalized by M. Arnaudon and X. M. Li in [9] to probability measures on convex affine manifolds with semilocal convex geometry. Moreover, the behavior of exponential barycenters when measures are pushed by stochastic flows is also considered in [9]. In order to study harmonic maps between Riemannian manifolds with probabilistic methods, J. Picard also gave a generalized notion of barycenters in [75]. As we noted before, Karcher means are only local minimizers of the energy functional f_μ in (1.2), but it is easily seen that f_μ can be defined not only on the closed ball $\bar{B}(a,\rho)$ but also on the whole manifold M as long as the second moment of μ is finite. This leads to the global minimizers of the second moment function of μ, which is just the original definition of means made by Fréchet. Global minimizers are more useful in statistics than local ones, so that it is necessary to know whether or under which conditions the Karcher mean of μ is in fact the Fréchet mean. For the case when μ is a discrete measure supported by finitely many points in the closed upper hemisphere, S. R. Buss and J. P. Fillmore showed in [29] that if the support of μ is not totally contained in the equator then μ has a unique Karcher mean which lies in the open hemisphere and equals to the Fréchet mean. Inspired by the methods of Buss and Fillmore, B. Afsari showed in [1] that if the upper curvature bound Δ and the injectivity radius inj in (1.5) is replaced by the ones of the larger ball $B(a, 2\rho)$, then all the Fréchet p-means of μ lie inside $B(a, \rho)$. Particularly, the Karcher means coincide with Fréchet means. The key point of Afsari's proof is an ingenious comparison argument in which not only the geodesic triangles in the manifold are compared with the ones in model spaces but also the whole configurations. This very beautiful and creative idea also leads to a new geometric proof of Kendall's result on the uniqueness of Karcher means. Gradient descent methods for computing Karcher means of probability measures in Riemannian manifolds are proposed by many authors. For example, by H. Le in [63] and by X. Pennec in [74]. New stochastic and deterministic algorithms with explicit stepsizes and error estimates for computing the Fréchet p-means of probability measures are given in Chapter 4 of this dissertation.

Medians of discrete sample points on the sphere are studied by economists and operational research experts in the 1970s and 1980s, but they used the name "location problems on a sphere". For data points lying in a spherical disc of radius smaller than $\pi/4$, Drezner and Wesolowsky showed in [35] that the cost function is unimodal in that disc and the Fréchet median is unique if the data points are not contained in a single great circle. A. A. Aly and his coauthors showed that (see [4]) the Fréchet medians are contained in the spherically convex hull of the sample points. Particularly, if all the sample points are contained in the open upper hemisphere, then all the Fréchet medians also lie in it. In order to show this particular case, they proposed in [4] an ingenious method using symmetry. More precisely, they observed that for every point outside the upper hemisphere the value of the cost function will be diminished if the point is replaced by its reflection with respect to the equator. This is just the method used by Buss and Fillmore in [29] and by Afsari in [1] to show the insideness of spherical means and Fréchet p-means. It is also shown by Z. Drezner in [36] that if all the sample points are contained in a great circle, then one of the sample points will be a Fréchet median. By the way, spherical medians also appears in statistical literature since the paper [43] of N. I. Fisher who proposed spherical analogues of spatial medians for unimodal, bipolar axial and girdle distributions on the sphere in order to carry out robust estimations for directional data. Perhaps the first work about Fréchet medians on Riemannian manifolds is the paper named "Generalized Fermat's problem" by R. Noda and his coauthors (see [72]). They proved the uniqueness, characterizations and position estimations of Fréchet medians for discrete sample points lying in a Cartan-Hadamard manifold. By the way, it seems that, probably due to its name which is very similar to that of the famous Fermat's last theorem, the article [72] is not known at all for people who work in the domain of nonlinear statistics, including me before 2011. In order to do robust statistics for data living in a Riemannian manifold P. T. Fletcher and his coauthors defined in [47] the local medians for discrete sample points. They showed the existence and uniqueness of local medians and gave a Riemannian analogue of Weiszfeld algorithm to compute it. But their algorithm is only proven to be convergent when the manifold is nonnegatively curved. A complete generalization of their results to arbitrary probability measures on Riemannian manifolds without curvature constraint is given in Chapter 2 of this dissertation, where results concerning the existence, uniqueness, characterization and position estimation of medians are proved under condition (1.1). Moreover, a subgradient algorithm that always converges to medians without curvature condition is also proposed. It is shown in [1] that if the upper curvature bound Δ and the injectivity radius inj in (1.5) is replaced by the ones of the larger ball $B(a, 2\rho)$, then all the Fréchet medians of μ are contained in $B(a, \rho)$. A generalization of this result is given in Chapter 3, where positions of Fréchet medians are estimated under the assumption that μ has more than a half mass lying in $B(a, \rho)$ and verifies some concentration condition. Moreover, we also show that if one takes data points randomly in a compact Riemannian manifold, then the Fréchet medians are almost surely unique.

1.2 The work of this dissertation

1.2.1 Motivation: radar target detection

Suggested by J. C. Maxwell's seminal work on electromagnetism, H. Hertz carried out an experiment in 1886 which validated that radio waves could be reflected by metallic objects. This provided C. Hüelsmeyer the theoretical foundation of his famous patent on "telemobiloscope" in 1904. He showed publicly in Germany and Netherlands that his device was able to detect remote metallic objects such as ships, even in dense fog or darkness, so that collisions could be avoided. Hüelsmeyer's "telemobiloscope" is recognized as the primogenitor of modern radar even though it could only detect the direction of an object, neither its distance nor its speed. This is because the basic idea of radar was already born: send radio waves in a predetermined direction and then receive the possible echoes reflected by a target. In order to know the distance and the radial speed of the target, it suffices to send successively two radio waves. In fact, it is easily seen that the distance d of the target can be computed by the formula

$$d = \frac{c\Delta t}{2},$$

where c is the speed of light and Δt is the time interval between every emission and reception in the direction under test. Moreover, the radial speed v of the target can be deduced by the Doppler effect which states that the frequency of a wave is changed for an observer moving relatively to the source of the wave. More precisely,

$$v = \frac{\lambda\Delta\varphi}{4\pi\Delta t},$$

where λ and $\Delta\varphi$ are the wavelength and the skewing of the two emitted radio waves, respectively. As a result, the direction, the distance and the speed of the target can all be determined.

For simplicity, from now on we only consider a fixed direction in which a radar sends radio waves. Since the range of emitted waves are finite, we can divide this direction into some intervals each of which represents a radar cell under test. The radar sends each time a rafale of radio waves in this direction and then receive the returning echoes. For each echo we measure its amplitude r and phase φ, so that it can be represented by a complex number $z = re^{i\varphi}$. As a result, the observation value of each radar cell is a complex vector $Z = (z_1, \dots, z_N)$, where N is the number of waves emitted in each rafale.

The aim of target detection is to know whether there is a target at the location of some radar cell in this direction. Intuitively speaking, a target is an object whose behavior on reflectivity or on speed is very different from its environment. The classical methods for target detection is to compute the difference between the discrete Fourier transforms of the radar observation values of the cell under test and that of its ambient cells. The bigger this difference is, the more likely a target appears at the location of the cell under test. However, the performance of these classical methods based on Doppler filtering using discrete Fourier transforms together with the Constant False Alarm Rate (CFAR) is not very satisfactory due to their low resolutions issues in perturbed radar environment or with smaller bunch of pulses.

In order to overcome these drawbacks, a lot of mathematical models for spectra estimation were introduced, among which the method based on autoregressive models proposed by F. Barbaresco in [13] is proved to be very preferable. We shall introduce this method in Chapter 6 of this dissertation. The main difference between this new method and the classical ones is that, instead of using directly the radar observation value Z of each cell, we regard it as a realization of a centered stationary Gaussian process and identify it to its covariance matrix $R = \mathbf{E}[ZZ^*]$. Thus the new observation value for each radar cell is a covariance matrix which is also Toeplitz due to the stationarity of the process. As a result, the principle for target detection becomes to find the cells where the covariance matrix differs greatly from the average matrix of its neighborhood. Once such cells are determined we can conclude that there are targets in these locations. In order to carry out this new method, there are two important things which should be considered seriously. One is to define a good distance between two Toeplitz covariance matrices. The other is to give a reasonable definition of the average of covariance matrices, which should be robust to outliers so as to be adapted to perturbed radar environment, and develop an efficient method to compute it in practical cases. These works will be done in the following by studying the Riemannian geometry of Toeplitz covariance matrices and the medians of probability measures on Riemannian manifolds.

1.2.2 Main results

In Chapter 2 of this dissertation we define local medians of a probability measure on a Riemannian manifold, give their characterization and a natural condition to ensure their uniqueness. In order to compute medians in practical cases, we also propose a subgradient algorithm and prove its convergence.

In more detail, let M be a complete Riemannian manifold with Riemannian metric $\langle \cdot , \cdot \rangle$ and Riemannian distance d. We fix an open geodesic ball

$$B(a,\rho) = \{x \in M : d(x,a) < \rho\}$$

in M centered at a with a finite radius ρ. Let δ and Δ denote respectively a lower and an upper bound of sectional curvatures K in $\bar{B}(a,\rho)$. The injectivity radius of $\bar{B}(a,\rho)$ is denoted by $\mathrm{inj}\,(\bar{B}(a,\rho))$. Furthermore, we assume that the radius of the ball verifies

$$\rho < \min\Big\{ \frac{\pi}{4\sqrt{\Delta}} , \frac{\mathrm{inj}\,(\bar{B}(a,\rho))}{2} \Big\}, \tag{2.7}$$

where if $\Delta \leq 0$, then $\pi/(4\sqrt{\Delta})$ is interpreted as $+\infty$.

We consider a probability measure μ on M whose support is contained in the open ball $B(a,\rho)$ and define a function

$$f : \qquad \bar{B}(a,\rho) \longrightarrow \mathbf{R}_+ , \qquad x \longmapsto \int_M d(x,p)\mu(dp).$$

This function is 1-Lipschitz, hence continuous on the compact set $\bar{B}(a,\rho)$. The convexity of the distance function on $\bar{B}(a,\rho)$ yields that f is also convex. Hence we don't need to distinguish its local minima from its global ones. Now we can give the following definition:

Definition 1.1. A minimum point of f is called a median of μ. The set of all the medians of μ will be denoted by $\mathfrak{M}_\mu$. The minimal value of f will be denoted by f_*.

It is easily seen that $\mathfrak{M}_\mu$ is compact and convex. Moreover, by computing the right derivative of f we can prove the following characterization of $\mathfrak{M}_\mu$, which is proved in [58] for a finite number of points in an Euclidean space.

Theorem 1.2. *The set $\mathfrak{M}_\mu$ is characterized by*

$$\mathfrak{M}_\mu = \big\{x \in \bar{B}(a,\rho) : |H(x)| \leq \mu\{x\}\big\},$$

where for $x \in \bar{B}(a,\rho)$,

$$H(x) := \int_{M\setminus\{x\}} \frac{-\exp_x^{-1} p}{d(x,p)} \mu(dp),$$

is a tangent vector at x satisfying $|H(x)| \leq 1$.

Observing that every geodesic triangle in $\bar{B}(a,\rho)$ has at most one obtuse angle, we can prove the following result which gives a position estimation for the medians of μ.

Proposition 1.3. *$\mathfrak{M}_\mu$ is contained in the smallest closed convex subset of $B(a,\rho)$ containing the support of μ.*

In Euclidean case, it is well known that if the sample points are not collinear, then their medians are unique. Hence we get a natural condition of μ to ensure the uniqueness for medians in Riemannian case:

$*$ *The support of μ is not totally contained in any geodesic. This means that for every geodesic γ: $[0,1] \to \bar{B}(a,\rho)$, we have $\mu(\gamma[0,1]) < 1$.*

This condition implies that f is strictly convex along every geodesic in $\bar{B}(a,\rho)$, so that it has one and only one minimizer, as stated by the theorem below.

Theorem 1.4. *If condition $*$ holds, then μ has a unique median.*

With further analysis, we can show a stronger quantitative version of Theorem 1.4, which is crucial in the error estimations of the subgradient algorithm as well as in the convergence proof of the stochastic algorithm for computing medians in Chapter 4.

Theorem 1.5. *If condition $*$ holds, then there exits a constant $\tau > 0$ such that for every $x \in \bar{B}(a,\rho)$ one has*

$$f(x) \geq f_* + \tau d^2(x,m),$$

where m is the unique median of μ.

The main results of approximating medians of μ by subgradient method is summarized in the following theorem. The idea stems from the basic observation that $H(x)$ is a subgradient of f at x for every $x \in \bar{B}(a,\rho)$.

Theorem 1.6. *Let $(t_k)_k$ be a sequence of real numbers such that*

$$t_k > 0, \quad \lim_{k\to\infty} t_k = 0 \quad \text{and} \quad \sum_{k=0}^{\infty} t_k = +\infty.$$

Define a sequence $(x_k)_k$ by $x_0 \in \bar{B}(a,\rho)$ and for $k \geq 0$,

$$x_{k+1} = \begin{cases} x_k, & \text{if } H(x_k) = 0; \\ \exp_{x_k}\Big(-t_k \dfrac{H(x_k)}{|H(x_k)|}\Big), & \text{if } H(x_k) \neq 0. \end{cases}$$

Then there exists some constant $T > 0$ such that if we choose $t_k \leq T$ for every $k \geq 0$, then the sequence $(x_k)_k$ is contained in $\bar{B}(a,\rho)$ and verifies

$$\lim_{k\to\infty} d(x_k, \mathfrak{M}_\mu) = 0 \quad \text{and} \quad \lim_{k\to\infty} f(x_k) = f_*.$$

Moreover, if the sequence $(t_k)_k$ *also verifies*

$$\sum_{k=0}^{\infty} t_k^2 < +\infty,$$

then there exists some $m \in \mathfrak{M}_\mu$ *such that* $x_k \longrightarrow m$.

Remark 1.7. We can choose the constant T in Theorem 1.6 to be

$$T = \frac{\rho - \sigma}{C(\rho, \delta)F(\rho, \Delta) + 1},$$

where $\sigma = \sup\{d(p, a) : p \in \operatorname{supp} \mu\}$,

$$F(\rho, \Delta) = \begin{cases} 1, & \text{if } \Delta \geq 0; \\ \cosh(2\rho\sqrt{-\Delta}), & \text{if } \Delta < 0, \end{cases}$$

and

$$C(\rho, \delta) = \begin{cases} 1, & \text{if } \delta \geq 0; \\ 2\rho\sqrt{-\delta}\coth(2\rho\sqrt{-\delta}), & \text{if } \delta < 0. \end{cases}$$

The proposition below gives the error estimation of the algorithm in Theorem 1.6.

Proposition 1.8. *Let condition* $*$ *hold and the stepsizes* $(t_k)_k$ *in Theorem 1.6 satisfy*

$$\lim_{k\to\infty} t_k = 0 \quad \textit{and} \quad \sum_{k=0}^{\infty} t_k = +\infty.$$

Then there exists $N \in \mathbf{N}$, *such that for every* $k \geq N$,

$$d^2(x_k, m) \leq b_k,$$

where m *is the unique median of* μ *and the sequence* $(b_k)_{k\geq N}$ *is defined by*

$$b_N = (\rho + \sigma)^2 \quad \textit{and} \quad b_{k+1} = (1 - 2\tau t_k)b_k + C(\rho, \delta)t_k^2\,, \quad k \geq N,$$

which converges to 0 *when* $k \to \infty$. *More explicitly, for every* $k \geq N$,

$$b_{k+1} = (\rho + \sigma)^2 \prod_{i=N}^{k}(1 - 2\tau t_i) + C(\rho, \delta)\Big(\sum_{j=N+1}^{k} t_{j-1}^2 \prod_{i=j}^{k}(1 - 2\tau t_i) + t_k^2\;\Big).$$

Chapter 3 is devoted to some basic results about Fréchet medians, or equivalently, global medians. We prove the consistency of Fréchet medians in proper metric spaces, give a quantitative estimation for the robustness of Fréchet medians in Riemannian manifolds and show the almost sure uniqueness of Fréchet sample medians in compact Riemannian manifolds.

In section 3.1, we work in a proper metric space (M, d) (recall that a metric space is proper if and only if every bounded and closed subset is compact). Let $P_1(M)$ denote the set of all the probability measures μ on M verifying

$$\int_M d(x_0, p)\mu(dp) < \infty, \text{ for some } x_0 \in M.$$

For every $\mu \in P_1(M)$ we can define a function

$$f_\mu : \qquad M \longrightarrow \mathbf{R}_+ , \qquad x \longmapsto \int_M d(x,p)\mu(dp).$$

This function is 1-Lipschitz hence continuous on M. Since M is proper, f_μ attains its minimum (see [76, p. 42]), so we can give the following definition:

Definition 1.9. Let μ be a probability measure in $P_1(M)$, then a *global* minimum point of f_μ is called a Fréchet median of μ. The set of all the Fréchet medians of μ is denoted by Q_μ. Let f_μ^* denote the global minimum of f_μ.

By the Kantorovich-Rubinstein duality of L^1-Wasserstein distance (see [88, p. 107]), we can show that Fréchet medians are characterized by 1-Lipschitz functions. A corresponding result that Riemannian barycenters are characterized by convex functions can be found in [55, Lemma 7.2].

Proposition 1.10. *Let $\mu \in P_1(M)$ and M be also separable, then*

$$Q_\mu = \left\{ x \in M : \varphi(x) \le f_\mu^* + \int_M \varphi(p)\mu(dp), \text{ for every } \varphi \in \mathrm{Lip}_1(M) \right\},$$

where $\mathrm{Lip}_1(M)$ denotes the set of all the 1-Lipschitz functions on M.

The following theorem states that the uniform convergence of first moment functions yields the convergence of Fréchet medians.

Theorem 1.11. *Let $(\mu_n)_{n\in\mathbf{N}}$ be a sequence in $P_1(M)$ and μ be another probability measure in $P_1(M)$. If $(f_{\mu_n})_n$ converges uniformly on M to f_μ, then for every $\varepsilon > 0$, there exists $N \in \mathbf{N}$, such that for every $n \ge N$ we have*

$$Q_{\mu_n} \subset B(Q_\mu, \varepsilon) := \{x \in M : d(x, Q_\mu) < \varepsilon\}.$$

As a corollary to Theorem 1.11, Fréchet medians are strongly consistent estimators. The consistency of Fréchet means is proved in [25].

Corollary 1.12. *Let $(X_n)_{n\in\mathbf{N}}$ be a sequence of i.i.d random variables of law $\mu \in P_1(M)$ and $(m_n)_{n\in\mathbf{N}}$ be a sequence of random variables such that $m_n \in Q_{\mu_n}$ with $\mu_n = \frac{1}{n}\sum_{k=1}^n \delta_{X_k}$. If μ has a unique Fréchet median m, then $m_n \longrightarrow m$ a.s.*

The framework of sections 3.2 and 3.3 is a complete Riemannian manifold (M, d) whose dimension is no less than 2. We fix a closed geodesic ball

$$\bar{B}(a,\rho) = \{x \in M : d(x,a) \leq \rho\}$$

in M centered at a with a finite radius $\rho > 0$ and a probability measure $\mu \in P_1(M)$ such that

$$\mu(\bar{B}(a,\rho)) = \alpha > \frac{1}{2}.$$

The aim of Section 3.2 is to estimate the positions of the Fréchet medians of μ, which gives a quantitative estimation for robustness. To this end, the following type of functions are of fundamental importance for our methods. Let $x, z \in M$, define

$$h_{x,z}: \quad \bar{B}(a,\rho) \longrightarrow \mathbf{R}, \quad p \longmapsto d(x,p) - d(z,p).$$

Obviously, $h_{x,z}$ is continuous and attains its minimum.

By a simple estimation on the minimum of $h_{x,a}$ we get the following basic result.

Theorem 1.13. *The set Q_μ of all the Fréchet medians of μ verifies*

$$Q_\mu \subset \bar{B}\Big(a, \frac{2\alpha\rho}{2\alpha - 1}\Big) := B_*.$$

Remark 1.14. It is easily seen that the conclusions of Proposition 3.6 and Theorem 1.13 also hold if M is only a proper metric space.

Remark 1.15. As a direct corollary to Theorem 1.13, if μ is a probability measure in $P_1(M)$ such that for some point $m \in M$ one has $\mu\{m\} > 1/2$, then m is the unique Fréchet median of μ.

In view of Theorem 1.13, let Δ be an upper bound of sectional curvatures in B_* and inj be the injectivity radius of B_*. By computing the minima of some typical functions $h_{x,z}$ in model spaces $\mathbb{S}^2$, $\mathbb{E}^2$ and $\mathbb{H}^2$, and then comparing with the ones in M, we get the following main result of section 3.2.

Theorem 1.16. *Assume that*

$$\frac{2\alpha\rho}{2\alpha - 1} < r_* := \min\{\frac{\pi}{\sqrt{\Delta}}, \mathrm{inj}\}, \tag{2.8}$$

where if $\Delta \leq 0$, then $\pi/\sqrt{\Delta}$ is interpreted as $+\infty$.

i) If $\Delta > 0$ and $Q_\mu \subset \bar{B}(a, r_*/2)$, then

$$Q_\mu \subset \bar{B}\Big(a, \frac{1}{\sqrt{\Delta}} \arcsin\big(\frac{\alpha \sin(\sqrt{\Delta}\rho)}{\sqrt{2\alpha - 1}}\big)\Big).$$

Moreover, any of the two conditions below implies $Q_\mu \subset \bar{B}(a, r_*/2)$:

$$a)\quad \frac{2\alpha\rho}{2\alpha-1}\le\frac{r_*}{2};\qquad b)\quad \frac{2\alpha\rho}{2\alpha-1}>\frac{r_*}{2}\quad \textit{and}\quad F_{\alpha,\rho,\Delta}(\frac{r_*}{2}-\rho)\le 0,$$

where $F_{\alpha,\rho,\Delta}(t)=\cot(\sqrt{\Delta}(2\alpha-1)t)-\cot(\sqrt{\Delta}t)-2\cot(\sqrt{\Delta}\rho),\ t\in(0,\dfrac{\rho}{2\alpha-1}]$.

ii) If $\Delta=0$, *then*

$$Q_\mu\subset\bar{B}\Big(a,\frac{\alpha\rho}{\sqrt{2\alpha-1}}\Big).$$

iii) If $\Delta<0$, *then*

$$Q_\mu\subset\bar{B}\Big(a,\frac{1}{\sqrt{-\Delta}}\operatorname{arcsinh}\big(\frac{\alpha\sinh(\sqrt{-\Delta}\rho)}{\sqrt{2\alpha-1}}\big)\Big).$$

Finally any of the above three closed balls is contained in the open ball $B(a,r_*/2)$.

Remark 1.17. Although we have chosen the framework of this section to be a Riemannian manifold, the essential tool that has been used is the hinge version of the triangle comparison theorem. Consequently, Theorem 1.16 remains true if M is a CAT(Δ) space (see [26, Chapter 2]) and r_* is replaced by $\pi/\sqrt{\Delta}$.

Remark 1.18. For the case when $\alpha=1$, the assumption (2.8) becomes

$$\rho<\frac{1}{2}\min\{\frac{\pi}{\sqrt{\Delta}}\ ,\mathrm{inj}\,\}.$$

Observe that in this case, when $\Delta>0$, the condition $F_{1,\rho,\Delta}(r_*/2-\rho)\le 0$ is trivially true in case of need. Hence Theorem 1.16 yields that $Q_\mu\subset\bar{B}(a,\rho)$, which is exactly what the Theorem 2.1 in [1] says for medians.

Before introduce the results of section 3.3 we give some notations. For each point $x\in M$, S_x denotes the unit sphere in T_xM. Moreover, for a tangent vector $v\in S_x$, the distance between x and its cut point along the geodesic starting from x with velocity v is denoted by $\tau(v)$. Certainly, if there is no cut point along this geodesic, then we define $\tau(v)=+\infty$. For every point $(x_1,\dots,x_N)\in M^N$, where $N\ge 3$ is a fixed natural number, we write

$$\mu(x_1,\dots,x_N)=\frac{1}{N}\sum_{k=1}^{N}\delta_{x_k}.$$

The set of all the Fréchet medians of $\mu(x_1,\dots,x_N)$, is denoted by $Q(x_1,\dots,x_N)$.

The following theorem states that in order to get the uniqueness of Fréchet medians, it suffices to move two data points towards a common median along some minimizing geodesics for a little distance.

Theorem 1.19. *Let $(x_1,\dots,x_N)\in M^N$ and $m\in Q(x_1,\dots,x_N)$. Fix two normal geodesics $\gamma_1,\gamma_2:[0,+\infty)\to M$ such that $\gamma_1(0)=x_1$, $\gamma_1(d(x_1,m))=m$, $\gamma_2(0)=x_2$ and $\gamma_2(d(x_2,m))=m$. Assume that*

$$x_2\notin\begin{cases}\gamma_1[0,\tau(\dot\gamma_1(0))], & \text{if } \tau(\dot\gamma_1(0))<+\infty;\\ \gamma_1[0,+\infty), & \text{if } \tau(\dot\gamma_1(0))=+\infty.\end{cases}$$

Then for every $t\in(0,d(x_1,m)]$ and $s\in(0,d(x_2,m)]$ we have

$$Q(\gamma_1(t),\gamma_2(s),x_3,\dots,x_N)=\{m\}.$$

Generally speaking, the non uniqueness of Fréchet medians is due to some symmetric properties of data points. As a result, generic data points should have a unique Fréchet median. In mathematical language, this means that the set of all the particular positions of data points is of Lebesgue measure zero. After eliminate all these particular cases we obtain the following main result of section 3.3:

Theorem 1.20. *Assume that M is compact. Then $\mu(x_1,\dots,x_N)$ has a unique Fréchet median for almost every $(x_1,\dots,x_N)\in M^N$.*

Remark 1.21. In probability language, Theorem 1.20 is equivalent to say that if $(X_1,\dots,X_N)$ is an M^N-valued random variable with density, then $\mu(X_1,\dots,X_N)$ has a unique Fréchet median almost surely. Clearly, the same statement is also true if $X_1,\dots,X_N$ are independent and M-valued random variables with density.

The Chapter 4 of this dissertation is a collaborative work of Marc Arnaudon, Clément Dombry, Anthony Phan and me. We consider a probability measure μ supported by a regular geodesic ball in a manifold and, for any $p\geq 1$, define a stochastic algorithm which converges almost surely to the p-mean e_p of μ. Assuming furthermore that the functional to minimize is regular around e_p, we prove that a natural renormalization of the inhomogeneous Markov chain converges in law into an inhomogeneous diffusion process. We give the explicit expression of this process, as well as its local characteristic. Finally, we show that the p-mean of μ can also be computed by the method of gradient descent. The questions concerning the choice of stepsizes and error estimates of this deterministic method are also considered.

In more detail, let M be a Riemannian manifold whose sectional curvatures $K(\sigma)$ verify $-\beta^2\leq K(\sigma)\leq\alpha^2$, where α,β are positive numbers. Denote by ρ the Riemannian distance on M. Let $B(a,r)$ be a geodesic ball in M and μ be a probability measure with support included in a compact convex subset K_μ of $B(a,r)$. Fix $p\in[1,\infty)$. We will always make the following assumptions on (r,p,μ):

Assumption 1.22. The support of μ is not reduced to one point. Either $p>1$ or the support of μ is not contained in a line, and the radius r satisfies

$$r<r_{\alpha,p}\quad\text{with}\begin{cases} r_{\alpha,p} &= \frac{1}{2}\min\left\{\text{inj}(M),\frac{\pi}{2\alpha}\right\}, \quad \text{if } p\in[1,2);\\ r_{\alpha,p} &= \frac{1}{2}\min\left\{\text{inj}(M),\frac{\pi}{\alpha}\right\}, \quad \text{if } p\in[2,\infty).\end{cases}$$

Under Assumption 1.22, it has been proved in [1, Theorem 2.1] that the function

$$\begin{aligned} H_p : M &\longrightarrow \mathbb{R}_+ \\ x &\longmapsto \int_M \rho^p(x,y)\mu(dy) \end{aligned}$$

has a unique minimizer e_p in M, the p-mean of μ, and moreover $e_p \in B(a,r)$. If $p=1$, e_1 is the median of μ.

In the following theorem, we define a stochastic gradient algorithm $(X_k)_{k\geq 0}$ to approximate the p-mean e_p and prove its convergence. In the sequel, let

$$K = \bar{B}(a, r-\varepsilon) \quad \text{with} \quad \varepsilon = \frac{\rho(K_\mu, B(a,r)^c)}{2}.$$

Theorem 1.23. *Let $(P_k)_{k\geq 1}$ be a sequence of independent $B(a,r)$-valued random variables, with law μ. Let $(t_k)_{k\geq 1}$ be a sequence of positive numbers satisfying*

$$\forall k \geq 1, \quad t_k \leq \min\left(\frac{1}{C_{p,\mu,K}}, \frac{\rho(K_\mu, B(a,r)^c)}{2p(2r)^{p-1}}\right),$$

$$\sum_{k=1}^{\infty} t_k = +\infty \quad \textit{and} \quad \sum_{k=1}^{\infty} t_k^2 < \infty,$$

where $C_{p,\mu,K} > 0$ is a constant.

Letting $x_0 \in K$, define inductively the random walk $(X_k)_{k\geq 0}$ by

$$X_0 = x_0 \quad \textit{and for } k \geq 0 \quad X_{k+1} = \exp_{X_k}\left(-t_{k+1} \operatorname{grad}_{X_k} F_p(\cdot, P_{k+1})\right)$$

where $F_p(x,y) = \rho^p(x,y)$, with the convention $\operatorname{grad}_x F_p(\cdot, x) = 0$.

The random walk $(X_k)_{k\geq 1}$ converges in L^2 and almost surely to e_p.

The fluctuation of the random walk $(X_k)_k$ defined in Theorem 1.23 is summarized in the following theorem.

Theorem 1.24. *Assume that in Theorem 1.23*

$$t_k = \min\left(\frac{\delta}{k}, \min\left(\frac{1}{C_{p,\mu,K}}, \frac{\rho(K_\mu, B(a,r)^c)}{2p(2r)^{p-1}}\right)\right), \quad k \geq 1,$$

for some $\delta > 0$. We define for $n \geq 1$ the Markov chain $(Y_k^n)_{k\geq 0}$ in $T_{e_p}M$ by

$$Y_k^n = \frac{k}{\sqrt{n}} \exp_{e_p}^{-1} X_k.$$

Assume that H_p is C^2 in a neighborhood of e_p and $\delta > C_{p,\mu,K}^{-1}$. Then the sequence of processes $\left(Y_{[nt]}^n\right)_{t\geq 0}$ converges weakly in $\mathbb{D}((0,\infty), T_{e_p}M)$ to a diffusion process given by

$$y_\delta(t) = \sum_{i=1}^{d} t^{1-\delta\lambda_i} \int_0^t s^{\delta\lambda_i - 1} \langle \delta\sigma\, dB_s, e_i \rangle e_i, \quad t \geq 0,$$

where B_t is the standard Brownian motion in $T_{e_p}M$ and $\sigma \in \mathrm{End}(T_{e_p}M)$ satisfying

$$\sigma\sigma^* = \mathbb{E}\left[\operatorname{grad}_{e_p} F_p(\cdot, P_1) \otimes \operatorname{grad}_{e_p} F_p(\cdot, P_1)\right],$$

$(e_i)_{1\leq i\leq d}$ is an orthonormal basis diagonalizing the symmetric bilinear form $\nabla dH_p(e_p)$ and $(\lambda_i)_{1\leq i\leq d}$ are the associated eigenvalues.

Gradient descent algorithms for computing e_p are given in the following theorem. In view of Theorem 1.6, it suffices to consider the case when $p > 1$.

Theorem 1.25. *Assume that $p > 1$. Let $x_0 \in \bar{B}(a, r)$ and for $k \geq 0$ define*

$$x_{k+1} = \exp_{x_k}(-t_k \operatorname{grad}_{x_k} H_p),$$

where $(t_k)_k$ is a sequence of real numbers such that

$$0 < t_k \leq \frac{p\varepsilon^{p+1}}{\pi p^2 (2r)^{2p-1} \beta \coth(2\beta r) + p\varepsilon^p}, \quad \lim_{k\to\infty} t_k = 0 \quad \text{and} \quad \sum_{k=0}^{\infty} t_k = +\infty.$$

Then the sequence $(x_k)_k$ is contained in $\bar{B}(a, \rho)$ and converges to e_p.

The following proposition gives the error estimations of the gradient descent algorithms in Theorem 1.25.

Proposition 1.26. *Assume that $t_k < C_{p,\mu,K}^{-1}$ for every k in Theorem 1.25, then the following error estimations hold:*
i) if $1 < p < 2$, then for $k \geq 1$,

$$\rho^2(x_k, e_p) \leq 4r^2 \prod_{i=0}^{k-1} (1 - C_{p,\mu,K} t_i) + C(\beta, r, p)\bigg(\sum_{j=1}^{k-1} t_{j-1}^2 \prod_{i=j}^{k-1} (1 - C_{p,\mu,K} t_i) + t_{k-1}^2\bigg) := b_k;$$

ii) if $p \geq 2$, then for $k \geq 1$,

$$H_p(x_k) - H_p(e_p) \leq (2r)^p \prod_{i=0}^{k-1} (1 - C_{p,\mu,K} t_i) + C(\beta, r, p)\bigg(\sum_{j=1}^{k-1} t_{j-1}^2 \prod_{i=j}^{k-1} (1 - C_{p,\mu,K} t_i) + t_{k-1}^2\bigg) := c_k,$$

where the constant

$$C(\beta, r, p) = \begin{cases} p^2(2r)^{2p-1}\beta\coth(2\beta r), & \textit{if } 1 < p < 2; \\ p^3(2r)^{3p-4}\left(2\beta r\coth(2\beta r) + p - 2\right), & \textit{if } p \geq 2. \end{cases}$$

Moreover, the sequences $(b_k)_k$ *and* $(c_k)_k$ *both tend to zero.*

In Chapter 5 of this dissertation we show that, under some conditions, local medians can be interpreted as solutions to fixed point problems. It is also shown that the associated iterated sequences converge to the medians. As a result, this gives rise to another way for computing Riemannian sample medians. The idea to add the penalty term $\mu\{x\}d(y, x)$ in the definition of h_x is to ensure that the fixed point mapping T diminishes the value of f. The main results of Chapter 5 generalize those of [87], in which all the results are proved only in Euclidean spaces.

The framework of Chapter 5 is the same to that of Chapter 2, so I omit it here. Besides, we shall make some assumption on the probability measure μ:

1) μ is not a Dirac measure,

2) μ has a unique median m,

3) for every x in the support of μ,

$$\int_{M\setminus\{x\}} \frac{1}{d(x,p)}\mu(dp) < \infty,$$

4) for every convergent sequence $(y_n)_n$ in $B(a, \rho)$,

$$\lim_{\mu A\to 0}\limsup_{n\to\infty}\int_{A\setminus\{y_n\}} \frac{1}{d(y_n,p)}\mu(dp) = 0,$$

5) the atoms of μ are isolated.

Remark 1.27. It is easily seen that if $N \geq 3$, $\sum_{k=1}^N \omega_k = 1$, $\omega_k > 0$ and $p_1, \dots, p_N$ are distinct points in $B(a, \rho)$ which are not contained in a single geodesic, then the probability measure

$$\mu = \sum_{k=1}^{N} \omega_k \delta_{p_k}, \tag{2.9}$$

satisfies all the above conditions.

The following type of functions are important in the sequel. For every x in the open ball $B(a, \rho)$, we define

$$h_x : \bar{B}(a,\rho) \longrightarrow \mathbf{R}_+, \qquad y \longmapsto \frac{1}{2}\int_{M\setminus\{x\}} \frac{d^2(y,p)}{d(x,p)}\mu(dp) + \mu\{x\}d(y,x).$$

Observe that h_x is continuous, strictly convex. Hence it has a unique minimum point which is denoted by $T(x)$.

The following theorem says that the medians of μ coincide with the fixed points of the mapping T.

Theorem 1.28. *The medians of μ are characterized by*

$$\mathfrak{M}_\mu = \{x \in B(a,\rho) : \ T(x) = x\}.$$

The main result of Chapter 5 is the following theorem:

Theorem 1.29. *Let $x_0 \in B(a,\rho)$, define a sequence $(x_n)_n$ by*

$$x_{n+1} = T(x_n), \quad n \geq 0.$$

Then $x_n \longrightarrow m$.

Example 1.30. Assume that M is a Euclidean space. Let $N \geq 3$, $\sum_{k=1}^{N} \omega_k = 1$, $\omega_k > 0$ and $p_1, \dots, p_N$ are distinct points in M which are not contained in a single line, then Theorem 1.29 holds for the probability measure

$$\mu = \sum_{k=1}^{N} \omega_k \delta_{p_k}.$$

If the median of μ dose not coincide with any data point p_k, then by choosing an appropriate starting point, we may assume that $\mu\{x_n\} = 0$ for every $n \geq 0$. As a result,

$$h_{x_n}(y) = \frac{1}{2} \sum_{k=1}^{N} \omega_k \frac{\|y - p_k\|^2}{\|x_n - p_k\|}.$$

It follows that

$$x_{n+1} = \sum_{k=1}^{N} \frac{\omega_k p_k}{\|x_n - p_k\|} \Big/ \sum_{k=1}^{N} \frac{\omega_k}{\|x_n - p_k\|},$$

which is exactly the Weiszfeld algorithm.

This dissertation is ended by its Chapter 6, in which we study the Riemannian geometry of the manifold of Toeplitz covariance matrices of order n. The explicit expression of the reflection coefficients reparametrization and its inverse are obtained. With the Riemannian metric given by the Hessian of a Kähler potential, we show that the manifold is in fact a Cartan-Hadamard manifold with lower sectional curvature bound -4. The geodesics in this manifold are also computed. Finally, we apply the subgradient algorithm introduced in Chapter 2 and the Riemannian geometry of Toeplitz covariance matrices to radar target detection.

In more detail, let $\mathcal{T}_n$ be the set of Toeplitz Hermitian positive definite matrices of order n. It is an open submanifold of $\mathbf{R}^{2n-1}$. Each element $R_n \in \mathcal{T}_n$ can be written as

$$R_n = \begin{bmatrix} r_0 & \overline{r}_1 & \dots & \overline{r}_{n-1} \\ r_1 & r_0 & \dots & \overline{r}_{n-2} \\ \vdots & \ddots & \ddots & \vdots \\ r_{n-1} & \dots & r_1 & r_0 \end{bmatrix}.$$

For every $1 \leq k \leq n-1$, the upper left $(k+1)$-by-$(k+1)$ corner of R_n is denoted by R_k. It is associated to a k-th order autoregressive model whose Yule-Walker equation is

$$\begin{bmatrix} r_0 & \overline{r}_1 & \dots & \overline{r}_k \\ r_1 & r_0 & \dots & \overline{r}_{k-1} \\ \vdots & \ddots & \ddots & \vdots \\ r_k & \dots & r_1 & r_0 \end{bmatrix} \begin{bmatrix} 1 \\ a_1^{(k)} \\ \vdots \\ a_k^{(k)} \end{bmatrix} = \begin{bmatrix} P_k \\ 0 \\ \vdots \\ 0 \end{bmatrix},$$

where $a_1^{(k)}, \dots, a_k^{(k)}$ are the optimal prediction coefficients and $P_k = \det R_{k+1} / \det R_k$ is the mean squared error.

The last optimal prediction coefficient $a_k^{(k)}$ is called the k-th reflection coefficient and is denoted by μ_k. It is easily seen that $\mu_1, \dots, \mu_{n-1}$ are uniquely determined by the matrix R_n. Moreover, the classical Levinson's recursion (see e.g. [86]) gives that $|\mu_k| < 1$. Hence, by letting $P_0 = r_0$, we obtain a map between two submanifolds of $\mathbf{R}^{2n-1}$:

$$\varphi : \quad \mathcal{T}_n \longrightarrow \mathbf{R}_+^* \times \mathbf{D}^{n-1}, \quad R_n \longmapsto (P_0, \mu_1, \dots, \mu_{n-1}),$$

where $\mathbf{D} = \{z \in \mathbf{C} : |z| < 1\}$ is the unit disc of the complex plane.

Using the Cramer's rule and the method of Schur complement (see e.g. [95]) we get the following proposition.

Proposition 1.31. *φ is a diffeomorphism, whose explicit expression is*

$$\mu_k = (-1)^k \frac{\det S_k}{\det R_k}, \quad \textit{where} \quad S_k = R_{k+1} \begin{pmatrix} 2, \dots, k+1 \\ 1, \dots, k \end{pmatrix}$$

is the submatrix of R_{k+1} obtained by deleting the first row and the last column. On the other hand, if $(P_0, \mu_1, \dots, \mu_{n-1}) \in \mathbf{R}_+^* \times \mathbf{D}^{n-1}$, then its inverse image R_n under φ can be calculated by the following algorithm:

$$r_0 = P_0, \quad r_1 = -P_0 \mu_1,$$

$$r_k = -\mu_k P_{k-1} + \alpha_{k-1}^T J_{k-1} R_{k-1}^{-1} \alpha_{k-1}, \quad 2 \leq k \leq n-1,$$

where

$$\alpha_{k-1} = \begin{bmatrix} r_1 \\ \vdots \\ r_{k-1} \end{bmatrix}, \quad J_{k-1} = \begin{bmatrix} 0 & \dots & 0 & 1 \\ 0 & \dots & 1 & 0 \\ & \dots & & \\ 1 & \dots & 0 & 0 \end{bmatrix} \quad \textit{and} \quad P_{k-1} = P_0 \prod_{i=1}^{k-1} (1 - |\mu_i|^2).$$

From now on, we regard $\mathcal{T}_n$ as a Riemannian manifold whose metric, which is introduced in [14] by the Hessian of the Kähler potential (see e.g. [12] for the definition of a Kähler potential)

$$\Phi(R_n) = -\ln(\det R_n) - n \ln(\pi e),$$

is given by

$$ds^2 = n\frac{dP_0^2}{P_0^2} + \sum_{k=1}^{n-1}(n-k)\frac{|d\mu_k|^2}{(1-|\mu_k|^2)^2}, \tag{2.10}$$

where $(P_0, \mu_1, \dots, \mu_{n-1}) = \varphi(R_n)$.

The metric (2.10) is a Bergman type metric and it will be shown in the appendix of Chapter 6 that this metric is not equal to the Fisher information metric of $\mathcal{T}_n$. But J. Burbea and C. R. Rao have proved in [28, Theorem 2] that the Bergman metric and the Fisher information metric do coincide for some probability density functions of particular forms. A similar potential function was used by S. Amari in [5] to derive the Riemannian metric of multi-variate Gaussian distributions by means of divergence functions. We refer to [78] for more account on the geometry of Hessian structures.

With the metric given by (2.10) the space $\mathbf{R}_+^* \times \mathbf{D}^{n-1}$ is just the product of the Riemannian manifolds $(\mathbf{R}_+^*, ds_0^2)$ and $(\mathbf{D}, ds_k^2)_{1 \le k \le n-1}$, where

$$ds_0^2 = n\frac{dP_0^2}{P_0^2} \quad \text{and} \quad ds_k^2 = (n-k)\frac{|d\mu_k|^2}{(1-|\mu_k|^2)^2}.$$

The latter is just $n-k$ times the classical Poincaré metric of $\mathbf{D}$. Hence $(\mathbf{R}_+^* \times \mathbf{D}^{n-1}, ds^2)$ is a Cartan-Hadamard manifold whose sectional curvatures K verify $-4 \le K \le 0$. The Riemannian distance between two different points x and y in $\mathbf{R}_+^* \times \mathbf{D}^{n-1}$ is given by

$$d(x,y) = \left(n\sigma(P,Q)^2 + \sum_{k=1}^{n-1}(n-k)\tau(\mu_k,\nu_k)^2 \right)^{1/2},$$

where $x = (P, \mu_1, \dots, \mu_{n-1})$, $y = (Q, \nu_1, \dots, \nu_{n-1})$,

$$\sigma(P,Q) = |\ln(\frac{Q}{P})| \quad \text{and} \quad \tau(\mu_k,\nu_k) = \frac{1}{2}\ln\frac{1+|\frac{\nu_k-\mu_k}{1-\bar{\mu}_k\nu_k}|}{1-|\frac{\nu_k-\mu_k}{1-\bar{\mu}_k\nu_k}|}.$$

The geodesic from x to y in $\mathcal{T}_n$ parameterized by arc length is given by

$$\gamma(s,x,y)=(\gamma_0(\frac{\sigma(P,Q)}{d(x,y)}s),\gamma_1(\frac{\tau(\mu_1,\nu_1)}{d(x,y)}s),\ldots,\gamma_1(\frac{\tau(\mu_{n-1},\nu_{n-1})}{d(x,y)}s)),$$

where γ_0 is the geodesic in $(\mathbf{R}_+^*, ds_0^2)$ from P to Q parameterized by arc length and for $1\le k\le n-1$, γ_k is the geodesic in $(\mathbf{D}, ds_k^2)$ from μ_k to ν_k parameterized by arc length. More precisely,

$$\gamma_0(t)=Pe^{t\,\mathrm{sign}(Q-P)},$$

and for $1\le k\le n-1$,

$$\gamma_k(t)=\frac{(\mu_k+e^{i\theta_k})e^{2t}+(\mu_k-e^{i\theta_k})}{(1+\bar{\mu}_k e^{i\theta_k})e^{2t}+(1-\bar{\mu}_k e^{i\theta_k})},\quad \text{where} \quad \theta_k=\arg\frac{\nu_k-\mu_k}{1-\bar{\mu}_k\nu_k}.$$

Particularly,

$$\gamma'(0,x,y)=(\gamma_0'(0)\frac{\sigma(P,Q)}{d(x,y)},\gamma_1'(0)\frac{\tau(\mu_1,\nu_1)}{d(x,y)},\ldots,\gamma_{n-1}'(0)\frac{\tau(\mu_{n-1},\nu_{n-1})}{d(x,y)}).$$

Let $v=(v_0,v_1,\ldots,v_{n-1})$ be a tangent vector in $T_x(\mathbf{R}_+^*\times\mathbf{D}^{n-1})$, then the geodesic starting from x with velocity v is given by

$$\zeta(t,x,v)=(\zeta_0(t),\zeta_1(t),\ldots,\zeta_{n-1}(t)),$$

where ζ_0 is the geodesic in $(\mathbf{R}_+^*, ds_0^2)$ starting from P with velocity v_0 and for $1\le k\le n-1$, ζ_k is the geodesic in $(\mathbf{D}, ds_k^2)$ starting from μ_k with velocity v_k. More precisely,

$$\zeta_0(t)=Pe^{\frac{v_0}{P}t},$$

and for $1\le k\le n-1$,

$$\zeta_k(t)=\frac{(\mu_k+e^{i\theta_k})e^{\frac{2|v_k|t}{1-|\mu_k|^2}}+(\mu_k-e^{i\theta_k})}{(1+\bar{\mu}_k e^{i\theta_k})e^{\frac{2|v_k|t}{1-|\mu_k|^2}}+(1-\bar{\mu}_k e^{i\theta_k})},\quad \text{where} \quad \theta_k=\arg v_k.$$

Numerical and radar simulations of the subgradient algorithme and the new detection method which are introduced respectively in Chapter 2 and in the subsection 1.2.1 can be found in section 6.4.

Chapter 2

Riemannian median and its estimation

Abstract

In this chapter, we define the geometric medians of a probability measure on a Riemannian manifold, give their characterization and a natural condition to ensure their uniqueness. In order to compute geometric medians in practical cases, we also propose a subgradient algorithm and prove its convergence as well as estimating the error of approximation and the rate of convergence. The convergence property of this subgradient algorithm, which is a generalization of the classical Weiszfeld algorithm in Euclidean spaces to the context of Riemannian manifolds, also improves a recent result of P. T. Fletcher *et al.* [*NeuroImage* 45 (2009) S143-S152].

2.1 Introduction

Throughout this chapter, M is a complete Riemannian manifold with Riemannian metric $\langle \cdot, \cdot \rangle$ and Riemannian distance d. The gradient operator and the hessian operator on M are denoted by grad and Hess respectively. Moreover, for every point p in M, let d_p denote the distance function to p defined by $d_p(x) = d(x,p)$, $x \in M$.

We fix an open geodesic ball

$$B(a,\rho) = \{x \in M : d(x,a) < \rho\}$$

in M centered at a with a finite radius ρ. Let δ and Δ denote respectively a lower and an upper bound of sectional curvatures K in $\bar{B}(a,\rho)$. The injectivity radius of $\bar{B}(a,\rho)$ is denoted by $\operatorname{inj}(\bar{B}(a,\rho))$. Furthermore, we assume that the radius of the ball verifies

$$\rho < \min\Big\{ \frac{\pi}{4\sqrt{\Delta}}, \frac{\operatorname{inj}(\bar{B}(a,\rho))}{2} \Big\}, \tag{1.1}$$

where if $\Delta \le 0$, then $\pi/(4\sqrt{\Delta})$ is interpreted as $+\infty$. Hence $\bar{B}(a,\rho)$ is convex (see [33, Theorem 5.14]), that is, for every two points x and y in $\bar{B}(a,\rho)$, there is a unique shortest geodesic from x to y in M, and this geodesic lies in $\bar{B}(a,\rho)$. Moreover, the geodesics in $\bar{B}(a,\rho)$ vary continuously with their endpoints. As a consequence, the angle comparison theorem of Alexandrov (see for example [2], [3, p.3] and [26, Proposition II.4.9]) can be applied in $\bar{B}(a,\rho)$. Similarly, it is easy to check that the angle comparison theorem of Toponogov (see [33, Theorem 2.2]) can also be applied in $\bar{B}(a,\rho)$. Thus we have to introduce some notations of model spaces that provide us many geometric information.

Notation 2.1. Let κ be a real number, the model space M^2_κ is defined as follows:

1) if $\kappa > 0$, then M^2_κ is obtained from the sphere $\mathbb{S}^2$ by multiplying the distance function by $1/\sqrt{\kappa}$;

2) if $\kappa = 0$, then M^2_0 is the Euclidean space $\mathbb{E}^2$;

3) if $\kappa < 0$, then M^2_κ is obtained from the hyperbolic space $\mathbb{H}^2$ by multiplying the distance function by $1/\sqrt{-\kappa}$.

The diameter of M^2_κ will be denoted by D_κ. More precisely,

$$D_\kappa = \begin{cases} \pi/\sqrt{\kappa}, & \text{if } \kappa > 0; \\ +\infty, & \text{if } \kappa \le 0. \end{cases}$$

The distance between two points $\bar{A}$ and $\bar{B}$ in M^2_κ will be written as $\bar{d}(\bar{A},\bar{B})$.

Moreover, we write for $t \in \mathbf{R}$,

$$S_\kappa(t) = \begin{cases} \sin(\sqrt{\kappa}\,t)/\sqrt{\kappa}, & \text{if } \kappa > 0; \\ t, & \text{if } \kappa = 0; \\ \sinh(\sqrt{-\kappa}\,t)/\sqrt{-\kappa}, & \text{if } \kappa < 0. \end{cases}$$

For the necessity of later comparison arguments, we recall some terminology about triangles (see [26, p.158 and Lemma I.2.14]). A *geodesic triangle* $\triangle ABC$ in M is a figure consisting of three distinct points A, B, C of M called the vertices and a choice of three shortest geodesics AB, BC, CA joining them called the sides. A *comparison triangle* in M^2_κ of $\triangle ABC$ is a geodesic triangle $\triangle \bar{A}\bar{B}\bar{C}$ in M^2_κ such that $\bar{d}(\bar{A},\bar{B}) = d(A,B)$, $\bar{d}(\bar{B},\bar{C}) = d(B,C)$, $\bar{d}(\bar{C},\bar{A}) = d(C,A)$. Note that if the perimeter of $\triangle ABC$ is less than $2D_\kappa$, that is, if $d(A,B)+d(B,C)+d(C,A) < 2D_\kappa$, then its comparison triangle in M^2_κ exits, and it is unique up to an isometry. Hence every geodesic triangle in $\bar{B}(a,\rho)$ has its comparison triangles in M^2_δ and M^2_Δ, respectively. Note that, by convexity of $\bar{B}(a,\rho)$, every geodesic triangle in it is uniquely determined by its three vertices.

The two estimations below are very useful in the following, which are direct corollaries to the classical Hessian comparison theorems.

Lemma 2.2. *Let $p \in \bar{B}(a,\rho)$ and $\gamma : [0,b] \to \bar{B}(a,\rho)$ be a geodesic, then*

i)

$$\mathrm{Hess}\, d_p(\dot\gamma(t),\dot\gamma(t)) \geq D(\rho,\Delta) |\, \dot\gamma_p^{\mathrm{nor}}(t)\,|^2$$

for every $t \in [0,b]$ such that $\gamma(t) \neq p$, where $D(\rho,\Delta) = S'_\Delta(2\rho)/S_\Delta(2\rho) > 0$ and $\dot\gamma_p^{\mathrm{nor}}(t)$ is the normal component of $\dot\gamma(t)$ with respect to the geodesic from p to $\gamma(t)$ in $\bar{B}(a,\rho)$.

ii)

$$\mathrm{Hess}\, \frac{1}{2} d_p^2(\dot\gamma(t),\dot\gamma(t)) \leq C(\rho,\delta)|\dot\gamma|^2$$

for every $t \in [0,b]$, where the constant $C(\rho,\delta) \geq 1$ is defined by

$$C(\rho,\delta) = \begin{cases} 1, & \text{if } \delta \geq 0; \\ 2\rho\sqrt{-\delta}\coth(2\rho\sqrt{-\delta}), & \text{if } \delta < 0. \end{cases}$$

Proof. Since in $\bar{B}(a,\rho)$ we have $\delta \leq K \leq \Delta$, by the classical Hessian comparison theorem (see [77, Lemma IV.2.9]) we get, for $\gamma(t) \neq p$,

$$\frac{S'_\Delta(d(\gamma(t),p))}{S_\Delta(d(\gamma(t),p))} |\, \dot\gamma(t)^{\mathrm{nor}}\,|^2 \leq \mathrm{Hess}\, d_p(\dot\gamma(t),\dot\gamma(t)) \leq \frac{S'_\delta(d(\gamma(t),p))}{S_\delta(d(\gamma(t),p))} |\, \dot\gamma(t)^{\mathrm{nor}}\,|^2.$$

Since $\theta \longmapsto S'_\Delta(\theta)/S_\Delta(\theta)$ is nonincreasing for $\theta > 0$ if $\Delta \leq 0$ and for $\theta \in (0,\pi)$ if $\Delta > 0$, so the left inequality together with $d(\gamma(t),p) \leq 2\rho$ proves the first assertion. To show the second one, let $\dot\gamma(t)^{\tan}$ be the tangential component of $\dot\gamma(t)$ with respect to the geodesic from p to $\gamma(t)$ then

$$\begin{aligned} \mathrm{Hess}\, \frac{1}{2} d_p^2(\dot\gamma(t),\dot\gamma(t)) &= d(\gamma(t),p)\, \mathrm{Hess}\, d_p(\dot\gamma(t),\dot\gamma(t)) + |\, \dot\gamma(t)^{\tan}\,|^2 \\ &\leq d(\gamma(t),p) \frac{S'_\delta(d(\gamma(t),p))}{S_\delta(d(\gamma(t),p))} |\, \dot\gamma(t)^{\mathrm{nor}}\,|^2 + |\, \dot\gamma(t)^{\tan}\,|^2 \\ &\leq \max\left\{ d(\gamma(t),p) \frac{S'_\delta(d(\gamma(t),p))}{S_\delta(d(\gamma(t),p))},\ 1 \right\} |\dot\gamma|^2. \end{aligned}$$

Observe that $\theta S'_\delta(\theta)/S_\delta(\theta) \leq 1$ for $\theta > 0$ if $\delta = 0$ and for $\theta \in [0,\pi)$ if $\delta > 0$, thus for the case when $\delta \geq 0$ we have

$$d(\gamma(t),p) \frac{S'_\delta(d(\gamma(t),p))}{S_\delta(d(\gamma(t),p))} \leq 1.$$

If $\delta < 0$, then $\theta \longmapsto \theta S'_\delta(\theta)/S_\delta(\theta) \geq 1$ and is nondecreasing for $\theta \geq 0$, thus we get

$$d(\gamma(t), p)\frac{S'_\delta(d(\gamma(t), p))}{S_\delta(d(\gamma(t), p))} \leq 2\rho \frac{S'_\delta(2\rho)}{S_\delta(2\rho)} = 2\rho\sqrt{-\delta}\coth(2\rho\sqrt{-\delta}),$$

hence the estimation holds for every $\delta \in \mathbf{R}$. Finally, the case when $\gamma(t) = p$ is trivial and the proof is complete. □

2.2 Definition of Riemannian medians

As in [53], we consider a probability measure μ on M whose support is contained in the open ball $B(a, \rho)$ and define a function

$$f: \qquad \bar{B}(a, \rho) \longrightarrow \mathbf{R}_+, \qquad x \longmapsto \int_M d(x, p)\mu(dp).$$

This function is 1-Lipschitz, hence continuous on the compact set $\bar{B}(a, \rho)$. Moreover, by the first estimation in Lemma 2.2, it is also convex. The convexity of f yields that its local minima coincide with its global ones, so that we don't need to distinguish the two. Now we give the following definition.

Definition 2.3. A minimum point of f is called a median of μ. The set of all the medians of μ will be denoted by $\mathfrak{M}_\mu$. The minimal value of f will be denoted by f_*.

It is easily seen that $\mathfrak{M}_\mu$ is compact and convex. In order to give a characterization of $\mathfrak{M}_\mu$, we need the following proposition, which implies that f is not differentiable in general.

Proposition 2.4. *Let $\gamma : [0, b] \to \bar{B}(a, \rho)$ be a geodesic, then*

$$\frac{d}{dt} f(\gamma(t))\big|_{t=t_0+} = \langle\, \dot{\gamma}(t_0),\, H(\gamma(t_0))\,\rangle + \mu\{\gamma(t_0)\}|\dot{\gamma}|, \quad t_0 \in [0, b\,),$$

$$\frac{d}{dt} f(\gamma(t))\big|_{t=t_0-} = \langle\, \dot{\gamma}(t_0),\, H(\gamma(t_0))\,\rangle - \mu\{\gamma(t_0)\}|\dot{\gamma}|, \quad t_0 \in (0, b\,],$$

where for $x \in \bar{B}(a, \rho)$,

$$H(x) = \int_{M\setminus\{x\}} \frac{-\exp_x^{-1} p}{d(x, p)}\mu(dp),$$

is a tangent vector at x satisfying $|H(x)| \leq 1$. Particularly, if $\mu\{x\} = 0$, then $\operatorname{grad} f(x) = H(x)$. *Moreover, H is continuous on $\bar{B}(a, \rho) \setminus \operatorname{supp}(\mu)$.*

Proof. We only prove the first identity since the proof of the second one is similar. Let $t_0 \in [\,0, b\,)$ and $\varepsilon > 0$ be sufficiently small, then

$$\begin{aligned}\frac{f(\gamma(t_0+\varepsilon)) - f(\gamma(t_0))}{\varepsilon} &= \int_M \frac{d(\gamma(t_0+\varepsilon), p) - d(\gamma(t_0), p)}{\varepsilon}\mu(dp) \\ &= \int_{M\setminus\{\gamma(t_0)\}} \frac{d(\gamma(t_0+\varepsilon), p) - d(\gamma(t_0), p)}{\varepsilon}\mu(dp) + \mu\{\gamma(t_0)\}|\dot\gamma|.\end{aligned}$$

Letting $\varepsilon \to 0+$ and using bounded convergence we get

$$\begin{aligned}\frac{d}{dt}f(\gamma(t))\big|_{t=t_0+} &= \int_{M\setminus\{\gamma(t_0)\}} \frac{d}{dt}d(\gamma(t), p)\big|_{t=t_0}\mu(dp) + \mu\{\gamma(t_0)\}|\dot\gamma| \\ &= \int_{M\setminus\{\gamma(t_0)\}} \langle\, \dot\gamma(t_0),\, \operatorname{grad} d_p(\gamma(t_0))\,\rangle \mu(dp) + \mu\{\gamma(t_0)\}|\dot\gamma| \\ &= \langle\, \dot\gamma(t_0),\, H(\gamma(t_0))\,\rangle + \mu\{\gamma(t_0)\}|\dot\gamma|.\end{aligned}$$

□

Now we give the characterization of $\mathfrak{M}_\mu$ which is proved in [58] for a finite number of points in an Euclidean space.

Theorem 2.5. $\mathfrak{M}_\mu = \big\{x \in \bar B(a,\rho) : |H(x)| \le \mu\{x\}\big\}$.

Proof. $(\subset)$ Let $x \in \mathfrak{M}_\mu$. If $H(x) = 0$, then there is nothing to prove. So we assume that $H(x) \neq 0$. Consider the geodesic in $\bar B(a,\rho)$:

$$\gamma(t) = \exp_x(-t\frac{H(x)}{|H(x)|}), \quad t \in [\,0,\, b\,].$$

By Proposition 2.4 and the definition of $\mathfrak{M}_\mu$ we get

$$|H(x)| = \mu\{x\} - \frac{d}{dt}f(\gamma(t))\big|_{t=0+} \le \mu\{x\}.$$

$(\supset)$ Let $x \in \bar B(a,\rho)$ such that $|H(x)| \le \mu\{x\}$. For every geodesic $\gamma : [0,1] \to \bar B(a,\rho)$ with $\gamma(0) = x$ and $\gamma(1) = y$, by the convexity of f, Proposition 2.4 and Cauchy-Schwartz inequality, we get

$$f(y) - f(x) \ge \frac{d}{dt}f(\gamma(t))\big|_{t=0+} \ge |\dot\gamma|(-|H(x)| + \mu\{x\}) \ge 0,$$

so that $x \in \mathfrak{M}_\mu$.

□

In order to describe the location of $\mathfrak{M}_\mu$, we need the following geometric lemma which is also useful in the next section.

Lemma 2.6. *Let $\triangle ABC$ be a geodesic triangle in $\bar{B}(a,\rho)$ such that $\angle A \geq \pi/2$, then $\angle B < \pi/2$ and $\angle C < \pi/2$.*

Proof. We prove this for the case when $\Delta > 0$. The proof for the cases when $\Delta \leq 0$ is similar. It suffices to show that $\angle B < \pi/2$. Let $d(B,C) = a_1$, $d(C,A) = b_1$ and $d(A,B) = c_1$. Consider a comparison triangle $\triangle \bar{A}\bar{B}\bar{C}$ of $\triangle ABC$ in M^2_Δ. Since $K \leq \Delta$ in $\bar{B}(a,\rho)$, Alexandrov's theorem yields that $\angle A \leq \angle \bar{A}$, $\angle B \leq \angle \bar{B}$. By the following identity in M^2_Δ,

$$\sin(\sqrt{\Delta}a_1)\cos\angle\bar{B} = \cos(\sqrt{\Delta}b_1)\sin(\sqrt{\Delta}c_1) - \sin(\sqrt{\Delta}b_1)\cos(\sqrt{\Delta}c_1)\cos\angle\bar{A},$$

we get $\angle \bar{B} < \pi/2$ and this completes the proof. □

Proposition 2.7. *$\mathfrak{M}_\mu$ is contained in the smallest closed convex subset of $B(a,\rho)$ containing the support of μ.*

Proof. Let V be this set. By Theorem 2.5 it suffices to show that if $x \in \bar{B}(a,\rho)\setminus V$ then $H(x) \neq 0$. In fact, let y be a point in V such that $d(x,y) = \inf\{d(x,p) : p \in V\}$, then the convexity of V yields $\angle xyp \geq \pi/2$ for every $p \in V$. Hence by Lemma 2.6 we get $\angle pxy < \pi/2$ and this gives that

$$\begin{aligned}\langle\, H(x), \exp_x^{-1} y\,\rangle &= \int_V \frac{\langle -\exp_x^{-1} p,\, \exp_x^{-1} y\,\rangle}{d(x,p)}\mu(dp)\\ &= -d(x,y)\int_V \cos\angle pxy\,\mu(dp) < 0.\end{aligned}$$

The proof is completed by observing that $\exp_x^{-1} y \neq 0$. □

2.3 Uniqueness of Riemannian medians

In Euclidean case, it is well known that if the sample points are not colinear, then the geometric median is unique. Hence we get a natural condition of μ to ensure the uniqueness for medians in Riemannian case:

$*$ *The support of μ is not totally contained in any geodesic. This means that for every geodesic γ: $[0,1] \to \bar{B}(a,\rho)$, we have $\mu(\gamma[0,1]) < 1$.*

Before giving the uniqueness theorem for medians, we introduce a procedure of extending geodesics in $\bar{B}(a,\rho)$. Let γ: $[0,1] \to \bar{B}(a,\rho)$ be a geodesic such that $\gamma(0) = x$ and $\gamma(1) = y$. By the completeness of M and the fact that the diameter of $\bar{B}(a,\rho)$ equals $2\rho < \mathrm{inj}\,(\bar{B}(a,\rho))$, we can extend γ from its endpoint y, along the direction of $\dot{\gamma}(1)$, to the point $\hat{y}$ where the extended geodesic firstly hits the boundary of $\bar{B}(a,\rho)$. Similarly, we can also apply this procedure in the opposite direction: extend γ from its starting point x, along the direction of $-\dot{\gamma}(0)$, to the point $\hat{x}$ where the extended geodesic firstly hits the boundary of $\bar{B}(a,\rho)$. Then we write $\hat{\gamma}$: $[0,1] \to \bar{B}(a,\rho)$ the geodesic such that $\hat{\gamma}(0) = \hat{x}$ and $\hat{\gamma}(1) = \hat{y}$. Obviously, $\gamma[0,1] \subset \hat{\gamma}[0,1]$. Furthermore, the strong convexity of $B(a,\rho)$ (see [33, Theorem 5.14]) yields $\dot{\gamma}_p^{\mathrm{nor}}(t) \neq 0$ for every $p \in \bar{B}(a,\rho) \setminus \hat{\gamma}[0,1]$ and $t \in [0,1]$.

Theorem 2.8. *If condition $*$ holds, then the median of μ is unique.*

Proof. We will prove this by showing that f is strictly convex, that is, for every geodesic γ: $[0,1] \to \bar{B}(a,\rho)$, the function $f \circ \gamma$ is strictly convex. By the first estimation in Lemma 2.2, for every $p \in \bar{B}(a,\rho) \setminus \hat{\gamma}[0,1]$ the function $t \mapsto d(\gamma(t),p)$ is strictly convex, and for $p \in \hat{\gamma}[0,1]$ it is trivially convex. Since the condition $*$ yields that $\mu(\bar{B}(a,\rho) \setminus \hat{\gamma}[0,1]) > 0$, by integration we obtain the strict convexity of f and the proof is completed. □

Remark 2.9. For the Riemannian barycenters of μ in $B(a,\rho)$ to be unique, Kendall showed that (see [55, Theorem 7.3] and [56]) the assumption

$$\rho < \min\Big\{ \frac{\pi}{2\sqrt{\Delta}}\, , \operatorname{inj}(a) \Big\} \tag{3.2}$$

suffices (certainly, without condition $*$). Naturally, one may wonder that, under condition $*$, whether the median of μ remains unique if condition (1.1) is replaced by the weaker one (3.2). Unfortunately, this is not true. A counterexample may be found in [54], which shows that if three points on the upper hemisphere $\mathbb{S}^2_+$ symmetrically located in a circle which is parallel and close to the equator, then there are at least three medians.

In the proof of Theorem 2.8, we have seen that f is strictly convex if condition $*$ holds. However, under the same condition, we can show that f is in fact strongly convex (see [85, Definition 6.1.1]): there exits a constant $\tau > 0$ such that for every geodesic γ: $[0,1] \to \bar{B}(a,\rho)$, the following inequalities hold,

$$f(\gamma(t)) \leq (1-t)f(\gamma(0)) + tf(\gamma(1)) - \tau|\dot{\gamma}|^2(1-t)t, \quad t \in [0,1]. \tag{3.3}$$

This is equivalent to say that for every geodesic γ: $[0,1] \to \bar{B}(a,\rho)$, the function $t \mapsto f(\gamma(t)) - \tau|\dot{\gamma}|^2 t^2$ is convex on $[0,1]$. To see this, we begin with an equivalent formulation of condition $*$.

Lemma 2.10. *Condition $*$ holds if and only if there exist two constants $\varepsilon_\mu > 0$ and $\eta_\mu > 0$, such that for every geodesic γ: $[0,1] \to \bar{B}(a,\rho)$, we have*

$$\mu(B(\gamma,\varepsilon_\mu)) \leq 1 - \eta_\mu,$$

where for $\varepsilon > 0$, $B(\gamma,\varepsilon) = \{x \in \bar{B}(a,\rho) : d(x,\gamma[0,1]) < \varepsilon\}$.

Proof. We only have to show the necessity since the sufficiency is trivial. Assume that this is not true, then for every $\varepsilon > 0$ and $\eta > 0$, there exists a geodesic γ: $[0,1] \to \bar{B}(a,\rho)$ such that $\mu(B(\gamma,\varepsilon)) > 1-\eta$. Then we obtain a sequence of geodesics $(\gamma_n)_{n\geq 1}$: $[0,1] \to \bar{B}(a,\rho)$ verifying $\mu(B(\gamma_n, 1/n)) > 1 - 1/n$. Since the sequence $(\gamma_n(0), \dot{\gamma}_n(0))_n$ is contained in the compact set $E = \{(x,v) \in TM : x \in \bar{B}(a,\rho), |v| \leq 2\rho\}$, there is a subsequence $(\gamma_{n_k}(0), \dot{\gamma}_{n_k}(0))_k$ and a point $(x,v) \in E$, such that $(\gamma_{n_k}(0), \dot{\gamma}_{n_k}(0)) \to (x,v)$. Let γ be the geodesic starting from x with velocity v, by the classical theory of ordinary differential equations we know that $\gamma_{n_k} \to \gamma$ uniformly on $[0,1]$. Particularly, $\gamma[0,1] \subset \bar{B}(a,\rho)$. Then for every $j \geq 1$, $B(\gamma_{n_k}, 1/n_k) \subset B(\gamma, 1/j)$ for sufficiently large k, hence $\mu(B(\gamma, 1/j)) \geq 1 - 1/n_k$. By letting $k \to \infty$ and then letting $j \to \infty$, we get $\mu(\gamma[0,1]) = 1$. This contradicts condition $*$. □

The lemma below gives a basic angle estimation, which is also useful in the next section. In the following, we write

$$\sigma = \sup\{d(p,a) : p \in \operatorname{supp}\mu\}.$$

Note that $\sigma < \rho$, since the support of μ is contained in the open ball $B(a,\rho)$.

Lemma 2.11. *Let $\triangle ABC$ be a geodesic triangle in $\bar{B}(a,\rho)$ such that $A = a$, $B \in \bar{B}(a,\sigma)$ and $C \in \bar{B}(a,\rho) \setminus \bar{B}(a,\sigma)$, then*

$$\cos\angle C \geq \frac{S_\Delta(d(C,A)-\sigma)}{S_\Delta(d(C,A)+\sigma)}.$$

Proof. We prove this lemma for the case when $\Delta > 0$. The proof for the cases when $\Delta \leq 0$ is similar. Let $d(B,C) = a_1$, $d(C,A) = b_1$ and $d(A,B) = c_1$. Consider a comparison triangle $\triangle\bar{A}\bar{B}\bar{C}$ of $\triangle ABC$ in M^2_Δ. Since $K \leq \Delta$ in $\bar{B}(a,\rho)$, Alexandrov's Theorem yields that $\angle C \leq \angle\bar{C}$. Hence, observing $c_1 \leq \sigma$ and $b_1 - \sigma \leq a_1 \leq b_1 + \sigma$, we obtain

$$\begin{aligned}
\cos\angle C \geq \cos\angle\bar{C} &= \frac{\cos(\sqrt{\Delta}c_1) - \cos(\sqrt{\Delta}a_1)\cos(\sqrt{\Delta}b_1)}{\sin(\sqrt{\Delta}a_1)\sin(\sqrt{\Delta}b_1)} \\
&\geq \frac{\cos(\sqrt{\Delta}\sigma) - \cos(\sqrt{\Delta}(b_1-\sigma))\cos(\sqrt{\Delta}b_1)}{\sin(\sqrt{\Delta}(b_1+\sigma))\sin(\sqrt{\Delta}b_1)} \\
&= \frac{\sin(\sqrt{\Delta}(b_1-\sigma))}{\sin(\sqrt{\Delta}(b_1+\sigma))} = \frac{S_\Delta(d(C,A)-\sigma)}{S_\Delta(d(C,A)+\sigma)}.
\end{aligned}$$

□

Lemma 2.12. *Let $\triangle ABC$ be a geodesic triangle in $\bar{B}(a,\rho)$ such that $\angle A \geq \pi/2$. Consider a geodesic triangle $\triangle A_1B_1C_1$ in M^2_δ such that $\bar{d}(A_1,C_1) = d(A,C)$, $\bar{d}(A_1,B_1) = d(A,B)$ and $\angle A_1 = \angle A$. Then $\sin\angle C \geq \sin\angle C_1$.*

Proof. We prove this for the case when $\delta > 0$. The proof for the cases when $\delta \le 0$ is similar. Firstly, observe that by Lemma 2.6 we have $\angle B < \pi/2$ and $\angle C < \pi/2$. Let $\triangle A_2B_2C_2$ be a comparison triangle of $\triangle ABC$ in M_δ^2. Since $K \ge \delta$ in $\bar{B}(a,\rho)$, Toponogov's theorem yields that $\angle A_2 \le \angle A = \angle A_1$ and $\angle B_2 \le \angle B$. Assume that $d(B,C) = a_2$, $d(A,C) = b_2$, $d(A,B) = c_2$ and $\bar{d}(B_1,C_1) = a_1$. Since $\angle A_2 \le \angle A_1$, we have $a_2 \le a_1$. By the law of cosines in M_δ^2 we obtain

$$\begin{aligned}
\frac{d}{da_1}\cos\angle C_1 &= \frac{d}{da_1}\frac{\cos(\sqrt{\delta}c_2) - \cos(\sqrt{\delta}a_1)\cos(\sqrt{\delta}b_2)}{\sin(\sqrt{\delta}b_2)\sin(\sqrt{\delta}a_1)} \\
&= \frac{\sqrt{\delta}(\cos(\sqrt{\delta}b_2) - \cos(\sqrt{\delta}a_1)\cos(\sqrt{\delta}c_2))}{\sin(\sqrt{\delta}b_2)\sin^2(\sqrt{\delta}a_1)} \\
&\ge \frac{\sqrt{\delta}(\cos(\sqrt{\delta}b_2) - \cos(\sqrt{\delta}a_2)\cos(\sqrt{\delta}c_2))}{\sin(\sqrt{\delta}b_2)\sin^2(\sqrt{\delta}a_1)} \\
&= \frac{\sqrt{\delta}\sin(\sqrt{\delta}a_2)\sin(\sqrt{\delta}c_2)}{\sin(\sqrt{\delta}b_2)\sin^2(\sqrt{\delta}a_1)}\cos\angle B_2 > 0.
\end{aligned}$$

So that $\cos\angle C_1$ is nondecreasing with respect to a_1 when $a_1 \ge a_2$. Hence $\cos\angle C_1 \ge \cos\angle C_2$, i.e. $\angle C_1 \le \angle C_2$. Then observing $\angle C_2 \le \angle C < \pi/2$, we obtain $\sin\angle C \ge \sin\angle C_1$. The proof is complete. □

The following lemma is important for giving a lower bound of the Hessian of the distance function. Though the two estimations below are very simple, they are sufficient for our purposes.

Lemma 2.13. *Let $\triangle ABC$ be a geodesic triangle in $\bar{B}(a,\rho)$.*

i) If $\angle A = \pi/2$, then

$$\sin\angle C \ge L_1(\rho,\delta)d(A,B),$$

where the constant

$$L_1(\rho,\delta) = \begin{cases} 2\sqrt{\delta}/\pi, & \text{if } \delta > 0; \\ 1/4\rho, & \text{if } \delta = 0; \\ \sqrt{-\delta}/\sinh(4\rho\sqrt{-\delta}), & \text{if } \delta < 0. \end{cases}$$

ii) If $A \in \partial\bar{B}(a,\rho)$, $B \in \bar{B}(a,\sigma)$ and $\angle A \ge \pi/2$, then

$$\sin\angle C \ge L(\sigma,\rho,\delta,\Delta)d(A,B),$$

where the constant

$$L(\sigma,\rho,\delta,\Delta) = L_1(\rho,\delta)\frac{S_\Delta(\rho-\sigma)}{S_\Delta(\rho+\sigma)}.$$

Proof. We prove this lemma for the case when $\delta > 0$. The proof for the cases when $\delta \leq 0$ is similar. Let $\triangle A_1B_1C_1$ be as in Lemma 2.12, then $\sin\angle C \geq \sin\angle C_1$. Let $\bar{d}(B_1, C_1) = a_1$ and $d(A, B) = c_1$, then the law of sines in M_δ^2 yields

$$\sin\angle C_1 = \frac{\sin(\sqrt{\delta}c_1)}{\sin(\sqrt{\delta}a_1)}\sin\angle A \geq \frac{2\sqrt{\delta}c_1}{\pi}\sin\angle A,$$

since $\sin\theta \geq 2\theta/\pi$ for $\theta \in [0, \pi/2]$. Hence the first estimation holds if $\angle A = \pi/2$. To show the second one, by the convexity of $\bar{B}(a, \rho)$ we get

$$\angle A \leq \angle CAa + \angle aAB < \frac{\pi}{2} + \angle aAB.$$

Since $\angle A \geq \pi/2$, by Lemma 2.11,

$$\sin\angle A \geq \cos\angle aAB \geq \frac{S_\Delta(\rho - \sigma)}{S_\Delta(\rho + \sigma)}.$$

The second estimation follows immediately and the proof is completed. $\square$

We know that f is convex, thus along every geodesic it has a second derivative in the sense of distribution, the following proposition gives its specific form as well as the Taylor's formulae.

Proposition 2.14. *Let γ: $[\,0, b\,] \to \bar{B}(a, \rho)$ be a geodesic, then*

for $t_0 \in [\,0, b\,)$ and $t \in [\,t_0, b\,]$,

$$f(\gamma(t)) = f(\gamma(t_0)) + \frac{d}{ds}f(\gamma(s))\big|_{s=t_0+}(t - t_0) + \int_{(t_0,t)}(t - s)\nu(ds); \tag{3.4}$$

for $t_0 \in (\,0, b\,]$ and $t \in [\,0, t_0\,]$,

$$f(\gamma(t)) = f(\gamma(t_0)) + \frac{d}{ds}f(\gamma(s))\big|_{s=t_0-}(t - t_0) + \int_{(t,t_0)}(s - t)\nu(ds),$$

with ν being the second derivative of $f \circ \gamma$ on $(0, b)$ in the sense of distribution, which is a bounded positive measure given by

$$\nu = \left(\int_{M\setminus\gamma[0,b]}\operatorname{Hess} d_p(\dot{\gamma}, \dot{\gamma})\mu(dp)\right)\cdot\lambda|_{(0,b)} + 2|\dot{\gamma}|\cdot(\mu\circ\gamma)|_{(0,b)},$$

where $\lambda|_{(0,b)}$ and $(\mu\circ\gamma)|_{(0,b)}$ denote the restrictions of Lebesgue measure and the measure $\mu\circ\gamma$ on $(0, b)$, respectively.

Proof. We shall only prove the first identity, since the proof of the second one is similar. Observe that since γ is a homeomorphism of $(0,b)$ onto its image, $\mu\circ\gamma$ is a well defined measure on $(0,b)$. By Taylor's formula,

$$
\begin{aligned}
&\int_{M\setminus\gamma[0,b]} d(\gamma(t),p)\mu(dp)\\
&=\int_{M\setminus\gamma[0,b]}\Big(d(\gamma(t_0),p)+\frac{d}{ds}d(\gamma(s),p)\big|_{s=t_0}(t-t_0)+\int_{t_0}^{t}(t-s)\frac{d^2}{ds^2}d(\gamma(s),p)ds\Big)\mu(dp)\\
&=\int_{M\setminus\gamma[0,b]} d(\gamma(t_0),p)\mu(dp)+\langle\,\dot\gamma(t_0),\int_{M\setminus\gamma[0,b]}\frac{-\exp^{-1}_{\gamma(t_0)}p}{d(\gamma(t_0),p)}\mu(dp)\,\rangle(t-t_0)\\
&\quad+\int_{t_0}^{t}(t-s)ds\int_{M\setminus\gamma[0,b]}\operatorname{Hess}d_p(\dot\gamma(s),\dot\gamma(s))\mu(dp).
\end{aligned}
$$

It is easily seen that

$$
\int_{M\setminus\gamma[0,b]} d(\gamma(t_0),p)=f(\gamma(t_0))-\int_{\gamma[0,t_0)}d(\gamma(t_0),p)\mu(dp)-\int_{\gamma(t_0,b]}d(\gamma(t_0),p)\mu(dp),
$$

and, by Proposition 2.4, that

$$
\begin{aligned}
&\Big\langle\,\dot\gamma(t_0),\int_{M\setminus\gamma[0,b]}\frac{-\exp^{-1}_{\gamma(t_0)}p}{d(\gamma(t_0),p)}\mu(dp)\,\Big\rangle\\
&=\langle\,\dot\gamma(t_0),H(\gamma(t_0))\,\rangle-\Big\langle\,\dot\gamma(t_0),\int_{\gamma[0,t_0)\cup\gamma(t_0,b]}\frac{-\exp^{-1}_{\gamma(t_0)}p}{d(\gamma(t_0),p)}\mu(dp)\,\Big\rangle\\
&=\langle\,\dot\gamma(t_0),H(\gamma(t_0))\,\rangle-|\dot\gamma|\mu(\gamma[\,0,t_0\,))+|\dot\gamma|\mu(\gamma(\,t_0,b\,])\\
&=\frac{d}{ds}f(\gamma(s))\big|_{s=t_0+}-|\dot\gamma|\mu(\gamma[\,0,t_0\,])+|\dot\gamma|\mu(\gamma(\,t_0,b\,]).
\end{aligned}
$$

Since

$$
f(\gamma(t))=\int_{M\setminus\gamma[0,b]}d(\gamma(t),p)\mu(dp)+\int_{\gamma[0,b]}d(\gamma(t),p)\mu(dp)),
$$

we obtain

$$
\begin{aligned}
&f(\gamma(t))-f(\gamma(t_0))-\frac{d}{ds}f(\gamma(s))\big|_{s=t_0+}(t-t_0)\\
&\quad-\int_{t_0}^{t}(t-s)ds\int_{M\setminus\gamma[0,b]}\operatorname{Hess}d_p(\dot\gamma(s),\dot\gamma(s))\mu(dp)\\
&=-\int_{\gamma[0,t_0)}d(\gamma(t_0),p)\mu(dp)-\int_{\gamma(t_0,b]}d(\gamma(t_0),p)\mu(dp)-|\dot\gamma|\mu(\gamma[\,0,t_0\,])(t-t_0)\\
&\quad+|\dot\gamma|\mu(\gamma(\,t_0,b\,])(t-t_0)+\int_{\gamma[0,b]}d(\gamma(t),p)\mu(dp)
\end{aligned}
$$

$$
\begin{aligned}
&= \Big(-\int_{\gamma[0,t_0)} d(\gamma(t_0),p)\mu(dp) - |\dot\gamma|\mu(\gamma[0,t_0))(t-t_0) + \int_{\gamma[0,t_0)} d(\gamma(t),p)\mu(dp) \Big) \\
&\quad + \Big(-\int_{\gamma(t_0,b]} d(\gamma(t_0),p)\mu(dp) + |\dot\gamma|\mu(\gamma(\,t_0,b\,])(t-t_0) + \int_{\gamma(t_0,b]} d(\gamma(t),p)\mu(dp) \Big) \\
&= \Big(-\int_{\gamma[0,t_0)} d(\gamma(t_0),p)\mu(dp) - \int_{\gamma[0,t_0)} d(\gamma(t_0),\gamma(t))\mu(dp) + \int_{\gamma[0,t_0)} d(\gamma(t),p)\mu(dp) \Big) \\
&\quad + \Big(-\int_{\gamma(t_0,t)} d(\gamma(t_0),p)\mu(dp) + \int_{\gamma(t_0,t)} d(\gamma(t_0),\gamma(t))\mu(dp) + \int_{\gamma(t_0,t)} d(\gamma(t),p)\mu(dp) \Big) \\
&\quad + \Big(-\int_{\gamma[t,b]} d(\gamma(t_0),p)\mu(dp) + \int_{\gamma[t,b]} d(\gamma(t_0),\gamma(t))\mu(dp) + \int_{\gamma[t,b]} d(\gamma(t),p)\mu(dp) \Big) \\
&= 2\int_{\gamma(t_0,t)} d(\gamma(t),p)\mu(dp) = 2|\dot\gamma| \int_{(t_0,t)} (t-s)(\mu\circ\gamma)(ds).
\end{aligned}
$$

Hence (3.4) holds. To show that ν is the second derivative of $f\circ\gamma$ on $(0,b)$ in the sense of distribution, let $\varphi \in C_c^\infty(\,0,b)$ and choose $t_0 = 0$ in (3.4), then Fubini's theorem and integration by parts yield

$$
\begin{aligned}
&\int_{(0,b)} f(\gamma(t))\varphi''(t)dt \\
&= f(\gamma(0)) \int_{(0,b)} \varphi''(t)dt + \frac{d}{ds} f(\gamma(s))\big|_{s=0+} \int_{(0,b)} t\varphi''(t)dt + \int_{(0,b)} \varphi''(t)dt \int_{(0,t)} (t-s)\nu(ds) \\
&= \int_{(0,b)} \nu(ds) \int_{(s,b)} (t-s)\varphi''(t)dt = \int_{(0,b)} \varphi(s)\nu(ds).
\end{aligned}
$$

The proof is complete. ∎

Now we are ready to show that condition $*$ yields the strong convexity of f. Certainly, this also gives a proof of the uniqueness of the median.

Theorem 2.15. *If condition $*$ holds, then f is strongly convex. More precisely, (3.3) holds for $\tau = (1/2)\,\varepsilon_\mu^2\,\eta_\mu\, D(\rho,\Delta)\, L(\sigma,\rho,\delta,\Delta)^2 > 0$. Moreover, with this choice of τ, for every $x \in \bar B(a,\rho)$,*

$$f(x) \geq f_* + \tau d^2(x,m),$$

where m is the unique median of μ.

Proof. Let $\gamma\colon [\,0,1\,] \to \bar B(a,\rho)$ be a geodesic, then by the first estimation in Lemma 2.2 we obtain that for every $s \in [\,0,1\,]$,

$$
\begin{aligned}
\int_{M\setminus\gamma[0,1]} \mathrm{Hess}\, d_p(\dot\gamma(s),\dot\gamma(s))\mu(dp) &\geq \int_{\bar B(a,\sigma)\setminus B(\hat\gamma,\varepsilon_\mu)} D(\rho,\Delta)|\dot\gamma_p^{\mathrm{nor}}(s)|^2 \mu(dp) \\
&= D(\rho,\Delta)|\dot\gamma|^2 \int_{\bar B(a,\sigma)\setminus B(\hat\gamma,\varepsilon_\mu)} \sin^2 \angle(\dot\gamma(s), \exp^{-1}_{\gamma(s)} p)\mu(dp).
\end{aligned}
$$

Then for every $p \in \bar{B}(a,\sigma) \setminus B(\hat{\gamma},\varepsilon_\mu)$, let $q = q(p)$ be the metric projection of p onto $\hat{\gamma}[0,1]$. If $q \in \hat{\gamma}(0,1)$ and $\gamma(s) \neq q$, then the geodesic triangle $\triangle pq\gamma(s)$ is a right triangle with $\angle pq\gamma(s) = \pi/2$. Hence the first estimation in Lemma 2.13 yields that

$$\sin\angle(\dot{\gamma}(s), \exp^{-1}_{\gamma(s)} p) \geq L_1(\rho,\delta) d(p,q) \geq L_1(\rho,\delta)\varepsilon_\mu.$$

If $q \in \{\hat{\gamma}(0), \hat{\gamma}(1)\}$ and $\gamma(s) \neq q$, then $\angle pq\gamma(s) \geq \pi/2$. Hence by the second estimation in Lemma 2.13,

$$\sin\angle(\dot{\gamma}(s), \exp^{-1}_{\gamma(s)} p) \geq L(\sigma,\rho,\delta,\Delta) d(p,q) \geq L(\sigma,\rho,\delta,\Delta)\varepsilon_\mu.$$

Since $L(\sigma,\rho,\delta,\Delta) < L_1(\rho,\delta)$, we always have

$$\sin\angle(\dot{\gamma}(s), \exp^{-1}_{\gamma(s)} p) \geq L(\sigma,\rho,\delta,\Delta)\varepsilon_\mu.$$

Thus by Lemma 2.10 we obtain

$$\int_{M\setminus\gamma[0,1]} \operatorname{Hess} d_p(\dot{\gamma}(s),\dot{\gamma}(s))\mu(dp) \geq \varepsilon_\mu^2\, \eta_\mu\, D(\rho,\Delta)\, L(\sigma,\rho,\delta,\Delta)^2 |\dot{\gamma}|^2 = 2\tau|\dot{\gamma}|^2.$$

Then for every $t \in [0,1)$ by Proposition 2.14,

$$\begin{aligned} f(\gamma(1)) &= f(\gamma(t)) + \frac{d}{ds} f(\gamma(s))\big|_{s=t+}(1-t) + \int_{(t,1)} (1-s)\nu(ds) \\ &\geq f(\gamma(t)) + \frac{d}{ds} f(\gamma(s))\big|_{s=t+}(1-t) + 2\,\tau|\dot{\gamma}|^2 \int_{(t,1)} (1-s)ds \\ &= f(\gamma(t)) + \frac{d}{ds} f(\gamma(s))\big|_{s=t+}(1-t) + \tau|\dot{\gamma}|^2(1-t)^2. \end{aligned} \tag{3.5}$$

Similarly, for every $t \in (0,1]$,

$$f(\gamma(0)) \geq f(\gamma(t)) + \frac{d}{ds} f(\gamma(s))\big|_{s=t-}(-t) + \tau|\dot{\gamma}|^2 t^2. \tag{3.6}$$

It follows by (3.5), (3.6) and Proposition 2.4 that for every $t \in (0,1)$,

$$\begin{aligned} f(\gamma(t)) \leq &(1-t) f(\gamma(0)) + t f(\gamma(1)) - \tau|\dot{\gamma}|^2(1-t)t \\ &- \Big(\frac{d}{ds} f(\gamma(s))\big|_{s=t+} - \frac{d}{ds} f(\gamma(s))\big|_{s=t-}\Big)(1-t)t \\ \leq &(1-t) f(\gamma(0)) + t f(\gamma(1)) - \tau|\dot{\gamma}|^2(1-t)t. \end{aligned}$$

The strong convexity of f is proved. To show the last inequality, let $\gamma\colon [0,1] \to \bar{B}(a,\rho)$ be the geodesic such that $\gamma(0) = m$ and $\gamma(1) = x$. Then (3.5) yields

$$f(x) \geq f(m) + \frac{d}{ds} f(\gamma(s))\big|_{s=0+} + \tau|\dot{\gamma}|^2 \geq f(m) + \tau|\dot{\gamma}|^2 = f_* + \tau d^2(x,m).$$

The proof is completed. □

2.4 A subgradient algorithm

To begin with, we recall the definition of the subgradient of a convex function on a Riemannian manifold. For our purpose, it suffices to consider this notion in a convex subset of the manifold.

Definition 2.16. Let U be a convex subset of M and h be a convex function defined on U. For every $x \in U$, a vector $v \in T_xM$ is called a subgradient of h at x if for every geodesic $\gamma\colon [0,b] \to U$ with $\gamma(0) = x$, we have

$$h(\gamma(t)) \geq h(x) + \langle \dot\gamma(0), v \rangle t, \quad t \in [0,b].$$

Our idea to approximate the median of μ by subgradient method stems from the following simple observation.

Lemma 2.17. *For every $x \in \bar{B}(a,\rho)$, $H(x)$ is a subgradient of f at x.*

Proof. Let $\gamma\colon [0,b] \to \bar{B}(a,\rho)$ be a geodesic such that $\gamma(0) = x$, then by Proposition 2.4, together with the convexity of f, we get for every $t \in [0,b]$,

$$\begin{aligned} f(\gamma(t)) &\geq f(\gamma(0)) + \frac{d}{ds} f(\gamma(s))\Big|_{s=0+} t \\ &= f(x) + (\langle \dot\gamma(0), H(x) \rangle + \mu\{x\}|\dot\gamma|)\, t \\ &\geq f(x) + \langle \dot\gamma(0), H(x) \rangle t. \end{aligned}$$

□

We proceed to give some notations which are necessary for introducing the subgradient algorithm.

Notation 2.18. If $x \in \bar{B}(a,\rho)$ and $H(x) \neq 0$, then we write

$$\gamma_x(t) = \exp_x\Big(-t\frac{H(x)}{|H(x)|}\Big), \quad t \geq 0.$$
$$r_x = \sup\{t \in [0, 2\rho] : \gamma_x(t) \in \bar{B}(a,\rho)\}.$$

Note that for $x \in \bar{B}(a,\rho)$ such that $H(x) \neq 0$, by the convexity of $\bar{B}(a,\rho)$ and the fact that $2\rho < \operatorname{inj}(\bar{B}(a,\rho))$, we have $\gamma_x[0, r_x] \subset \bar{B}(a,\rho)$.

The following simple lemma shows that every r_x is strictly positive. More importantly, it ensures theoretically the possibility of the choice of stepsizes in the convergence theorem of the subgradient algorithm.

Lemma 2.19. $\inf\{r_x : x \in \bar{B}(a,\rho),\ H(x) \neq 0\} > 0$.

Proof. Since the support of μ is contained in $B(a,\rho)$, for every $x \in \partial\bar{B}(a,\rho)$, $H(x)$ is transverse to $\partial\bar{B}(a,\rho)$, and hence $r_x > 0$ for every $x \in \bar{B}(a,\rho)$ such that $H(x) \neq 0$. Moreover, there exists $\varepsilon > 0$ such that $\operatorname{supp}(\mu) \subset B(a, \rho - \varepsilon)$. Then for $x \in$

$B(a, \rho - \varepsilon)$ such that $H(x) \neq 0$ we have $r_x \geq \rho - d(x, a) > \varepsilon$. On the other hand, since H is continuous on $\bar{B}(a, \rho) \setminus B(a, \rho - \varepsilon)$, r_x vary continuously with x on this compact set. Thus there exists a point $x_0 \in \bar{B}(a, \rho) \setminus B(a, \rho - \varepsilon)$ such that $\inf\{r_x : x \in \bar{B}(a, \rho) \setminus B(a, \rho - \varepsilon)\} = r_{x_0}$. Hence we get $\inf\{r_x : x \in \bar{B}(a, \rho),\ H(x) \neq 0\} \geq \min\{\varepsilon, r_{x_0}\} > 0$. □

Now we introduce the subgradient algorithm to approximate the medians of the probability measure μ.

Algorithm 2.20. Subgradient algorithm for Riemannian medians:
Step 1:
Choose a starting point $x_0 \in \bar{B}(a, \rho)$ and let $k = 0$.
Step 2:
If $H(x_k) = 0$, then $x_k \in \mathfrak{M}_\mu$ and stop. If not, then go to step 3.
Step 3:
Choose a stepsize $t_k \in (\,0,\, r_{x_k}]$ and let $x_{k+1} = \gamma_{x_k}(t_k)$, then come back to step 2 with $k = k + 1$.

Remark 2.21. It should be noted that, in the above algorithm, we have already restricted every stepsize t_k to be in the interval $(\,0,\, r_{x_k}]$. From now on, we shall always make this restriction implicitly to ensure that the sequence $(x_k)_k$ will never get out of the ball $\bar{B}(a, \rho)$.

Now we turn to the convergence proof of the above algorithm under some further conditions of the stepsizes. In Euclidean spaces, it is well known that the following type of inequalities are of fundamental importance to conclude the convergence of subgradient algorithms (see for example [34, 71]):

$$||x_{k+1} - y||^2 \leq ||x_k - y||^2 + \alpha t_k^2 + \beta \frac{2t_k}{||v_k||}(f(y) - f(x_k)).$$

For a nonnegatively curved Riemannian manifold, Ferreira and Oliveira obtained a generalization of the above inequality in [40] by using Toponogov's comparison theorem. But their method is not applicable in our case since $\bar{B}(a, \rho)$ is not assumed to be nonnegatively curved. However, we can still obtain a similar result using a different method.

Lemma 2.22. *Let $(x_k)_k$ be the sequence generated by Algorithm 6.17. If $H(x_k) \neq 0$, then for every point $y \in \bar{B}(a, \rho)$,*

$$d^2(x_{k+1}, y) \leq d^2(x_k, y) + C(\rho, \delta) t_k^2 + \frac{2t_k}{|H(x_k)|}(f(y) - f(x_k)).$$

Particularly,

$$d^2(x_{k+1}, \mathfrak{M}_\mu) \leq d^2(x_k, \mathfrak{M}_\mu) + C(\rho, \delta) t_k^2 + 2t_k(f_* - f(x_k)). \tag{4.7}$$

Proof. By Taylor's formula and the second estimation in Lemma 2.2, there exists $\xi \in (0, t_k)$ such that

$$\begin{aligned}
\frac{1}{2}d^2(x_{k+1}, y) =& \frac{1}{2}d^2(\gamma_{x_k}(t_k), y) \\
=& \frac{1}{2}d^2(\gamma_{x_k}(0), y) + \frac{d}{dt}\left[\frac{1}{2}d^2(\gamma_{x_k}(t), y)\right]_{t=0} t_k + \frac{1}{2}\frac{d^2}{dt^2}\left[\frac{1}{2}d^2(\gamma_{x_k}(t), y)\right]_{t=\xi} t_k^2 \\
=& \frac{1}{2}d^2(x_k, y) + \langle\, \dot\gamma_{x_k}(0), \operatorname{grad}\frac{1}{2}d_y^2(x_k)\,\rangle\, t_k + \frac{1}{2}\operatorname{Hess}\frac{1}{2}d_y^2(\dot\gamma_{x_k}(\xi), \dot\gamma_{x_k}(\xi))\, t_k^2 \\
\leq& \frac{1}{2}d^2(x_k, y) + \frac{\langle\, H(x_k), \exp_{x_k}^{-1} y\,\rangle}{|H(x_k)|}\, t_k + \frac{C(\rho, \delta)}{2}\, t_k^2.
\end{aligned}$$

By Lemma 2.17, $H(x_k)$ is a subgradient of f at point x_k and hence

$$\langle\, H(x_k), \exp_{x_k}^{-1} y\,\rangle \leq f(y) - f(x_k).$$

Consequently,

$$\frac{1}{2}d^2(x_{k+1}, y) \leq \frac{1}{2}d^2(x_k, y) + \frac{t_k}{|H(x_k)|}(f(y) - f(x_k)) + \frac{C(\rho, \delta)}{2}t_k^2,$$

the first inequality holds. The second one follows from $f_* \leq f(x_k)$ and $|H(x_k)| \leq 1$. ∎

As in the Euclidean case, once the fundamental inequality is established, the convergence of the subgradient algorithm is soon achieved and the proof is elementary. Since the fundamental inequality (4.7) in Lemma 2.22 is very similar to the Euclidean one, the proof of the following convergence theorem is also very similar to the one in Euclidean case.

Theorem 2.23. *If the stepsizes $(t_k)_k$ verify*

$$\lim_{k\to\infty} t_k = 0 \quad \textit{and} \quad \sum_{k=0}^{\infty} t_k = +\infty,$$

then the sequence $(x_k)_k$ generated by Algorithm 6.17 satisfies

$$\lim_{k\to\infty} d(x_k, \mathfrak{M}_\mu) = 0 \quad \textit{and} \quad \lim_{k\to\infty} f(x_k) = f_*.$$

Moreover, if the stepsizes $(t_k)_k$ also verify

$$\sum_{k=0}^{\infty} t_k^2 < +\infty,$$

then there exists some $m \in \mathfrak{M}_\mu$ such that $x_k \to m$.

Proof. Without loss of generality, we may assume that $H(x_k) \neq 0$ for every $k \geq 0$. Our first step is to show that

$$\liminf_{k\to\infty} f(x_k) = f_*$$

If this is not true, then there exist $N_1 \in \mathbf{N}$ and $\eta > 0$ such that for every $k \geq N_1$ we have $f_* - f(x_k) \leq -\eta$. By Lemma 2.22, we get

$$d^2(x_{k+1}, \mathfrak{M}_\mu) \leq d^2(x_k, \mathfrak{M}_\mu) + t_k(C(\rho,\delta)t_k - 2\eta)$$

Since $\lim_{k\to\infty} t_k = 0$, we can suppose that $C(\rho,\delta)t_k < \eta$ for every $k \geq N_1$ and hence

$$d^2(x_{k+1}, \mathfrak{M}_\mu) \leq d^2(x_k, \mathfrak{M}_\mu) - \eta t_k$$

by summing the above inequalities we get

$$\eta \sum_{i=N_1}^{k} t_i \leq d^2(x_{N_1}, \mathfrak{M}_\mu) - d^2(x_{k+1}, \mathfrak{M}_\mu) \leq d^2(x_{N_1}, \mathfrak{M}_\mu)$$

which contradicts with $\sum_{k=0}^{\infty} t_k = +\infty$, this proves the assertion.

Now for fixed $\varepsilon > 0$, there exists $N_2 \in \mathbf{N}$ such that $C(\rho,\delta)t_k < 2\varepsilon$ for every $k \geq N_2$. We consider the following two cases:
If $f(x_k) > f_* + \varepsilon$, then by Lemma 2.22 we obtain that

$$\begin{aligned} d^2(x_{k+1}, \mathfrak{M}_\mu) &\leq d^2(x_k, \mathfrak{M}_\mu) + C(\rho,\delta)t_k^2 + 2t_k(f_* - f(x_k)) \\ &< d^2(x_k, \mathfrak{M}_\mu) + (C(\rho,\delta)t_k - 2\varepsilon)t_k < d^2(x_k, \mathfrak{M}_\mu) \end{aligned}$$

If $f(x_k) \leq f_* + \varepsilon$ then $x_k \in L_\varepsilon = \{x \in \bar{B}(a,\rho) : f(x) \leq f_* + \varepsilon\}$ and if we write $l_\varepsilon = \sup\{d(y, \mathfrak{M}_\mu) : y \in L_\varepsilon\}$, hence in this case we have

$$d(x_{k+1}, \mathfrak{M}_\mu) \leq d(x_{k+1}, x_k) + d(x_k, \mathfrak{M}_\mu) \leq t_k + l_\varepsilon$$

In conclusion, we always have that for $k \geq N_2$,

$$d(x_{k+1}, \mathfrak{M}_\mu) \leq \max\{d(x_k, \mathfrak{M}_\mu), t_k + l_\varepsilon\}$$

By induction we get for every $n \geq k$,

$$\begin{aligned} d(x_{n+1}, \mathfrak{M}_\mu) &\leq \max\{d(x_k, \mathfrak{M}_\mu), \max\{t_k, t_{k+1}, \ldots, t_n\} + l_\varepsilon\} \\ &\leq \max\{d(x_k, \mathfrak{M}_\mu), \sup\{t_i : i \geq k\} + l_\varepsilon\} \end{aligned}$$

thus we get

$$\limsup_{n\to\infty} d(x_n, \mathfrak{M}_\mu) \leq \max\{d(x_k, \mathfrak{M}_\mu), \sup\{t_i : i \geq k\} + l_\varepsilon\}$$

$\liminf_{k\to\infty} f(x_k) = f_*$ yields that $\liminf_{k\to\infty} d(x_k, \mathfrak{M}_\mu) = 0$, by taking the inferior limit on the right hand side we obtain

$$\limsup_{n\to\infty} d(x_n, \mathfrak{M}_\mu) \leq l_\varepsilon$$

Now we show that $l_\varepsilon \to 0$ when $\varepsilon \to 0$. By monotonicity of l_ε, it suffices to show this along some sequence. In fact, observe that L_ε is compact, thus for every $\varepsilon > 0$, there exists $y_\varepsilon \in L_\varepsilon$ such that $l_\varepsilon = d(y_\varepsilon, \mathfrak{M}_\mu)$. Since $\bar{B}(a,\rho)$ is compact, there exist a sequence $\varepsilon_k \to 0$ and $y \in \bar{B}(a,\rho)$ such that $y_{\varepsilon_k} \to y$. Since $f(y_{\varepsilon_k}) \leq f_* + \varepsilon_k$, we have $f(y) \leq f_*$ and hence $y \in \mathfrak{M}_\mu$. Consequently $l_{\varepsilon_k} \to 0$. Thus we get $d(x_k, \mathfrak{M}_\mu) \to 0$ and this yields $f(x_k) \to f_*$.

If $\sum_{k=0}^{\infty} t_k^2 < +\infty$, the compactness of $\bar{B}(a,\rho)$ and $f(x_k) \to f_*$ imply that the sequence $(x_k)_k$ has some cluster point $m \in \mathfrak{M}_\mu$, hence Lemma 2.22 yields

$$d^2(x_{k+1}, m) \leq d^2(x_k, m) + C(\rho,\delta) t_k^2$$

Then for every $n \geq k$, by summing the above inequalities we get

$$d^2(x_{n+1}, m) \leq d^2(x_k, m) + C(\rho,\delta) \sum_{i=k}^{n} t_i^2$$

Let $n \to \infty$ and we get

$$\limsup_{n\to\infty} d^2(x_n, m) \leq d^2(x_k, m) + C(\rho,\delta) \sum_{i=k}^{\infty} t_i^2$$

the proof will be completed by observing that the right hand side of the above inequality posseses a subsequence that converges to 0.

$\square$

Now we consider the problem of the choice of stepsizes. By Lemma 2.19 we can choose $(t_k)_k$ that verifies the conditions of the preceding theorem and hence yields the desired convergence of our algorithm. For example, we may take $t_k = r_{x_k}/(k+1)$ for every $k \geq 0$. But the drawback is that we do not know much about r_{x_k}. However, with further analysis we can obtain an explicit lower bound for it.

Lemma 2.24. *For every $x \in \bar{B}(a,\rho) \setminus \bar{B}(a,\sigma)$,*

$$r_x \geq \frac{2d(x,a) S_\Delta(d(x,a) - \sigma)}{C(\rho,\delta) S_\Delta(d(x,a) + \sigma)}.$$

Proof. Since $x \in \bar{B}(a,\rho) \setminus \bar{B}(a,\sigma)$, we have $H(x) \neq 0$ and hence r_x is well defined. Moreover, the diameter of $\bar{B}(a,\rho)$ is $2\rho < \operatorname{inj}(\bar{B}(a,\rho))$, thus the definition of r_x

yields that $\gamma_x(r_x) \in \partial \bar{B}(a,\rho)$. By Taylor's formula and the second estimation in Lemma 2.2, there exists $\xi \in (0, r_x)$ such that

$$\begin{aligned}
\frac{1}{2}\rho^2 &= \frac{1}{2}d^2(\gamma_x(r_x), a) \\
&= \frac{1}{2}d^2(x,a) + \frac{d}{dt}\left[\frac{1}{2}d^2(\gamma_x(t),a)\right]_{t=0} r_x + \frac{1}{2}\frac{d^2}{dt^2}\left[\frac{1}{2}d^2(\gamma_x(t),a)\right]_{t=\xi} r_x^2 \\
&= \frac{1}{2}d^2(x,a) + \langle\, \dot\gamma_x(0),\, \mathrm{grad}\,\frac{1}{2}d_a^2(x)\,\rangle\, r_x + \frac{1}{2}\,\mathrm{Hess}\,\frac{1}{2}d_a^2(\dot\gamma_x(\xi), \dot\gamma_x(\xi))\, r_x^2 \\
&\le \frac{1}{2}d^2(x,a) + \frac{\langle\, H(x),\, \exp_x^{-1} a\,\rangle}{|H(x)|}\, r_x + \frac{C(\rho,\delta)}{2}\, r_x^2.
\end{aligned}$$

Gauss Lemma yields that $\langle\, \exp_x^{-1} p,\, \exp_x^{-1} a\,\rangle > 0$ for $p \in \mathrm{supp}\,\mu$, hence

$$\langle\, H(x),\, \exp_x^{-1} a\,\rangle = -\int_{\mathrm{supp}\,\mu} \frac{\langle\, \exp_x^{-1} p,\, \exp_x^{-1} a\,\rangle}{d(x,p)}\mu(dp) < 0.$$

Combine this with $d(x,a) \le \rho$, $C(\rho,\delta) > 0$ and $r_x > 0$, we obtain that

$$\begin{aligned}
r_x &\ge \frac{1}{C(\rho,\delta)}\left\{ -\frac{\langle\, H(x),\, \exp_x^{-1} a\,\rangle}{|H(x)|} + \sqrt{\frac{\langle\, H(x),\, \exp_x^{-1} a\,\rangle^2}{|H(x)|^2} + C(\rho,\delta)(\rho^2 - d^2(x,a))}\right\} \\
&\ge \frac{1}{C(\rho,\delta)}\left\{ -\frac{\langle\, H(x),\, \exp_x^{-1} a\,\rangle}{|H(x)|} + \frac{|\langle\, H(x),\, \exp_x^{-1} a\,\rangle|}{|H(x)|}\right\} \\
&= \frac{-2}{C(\rho,\delta)}\frac{\langle\, H(x),\, \exp_x^{-1} a\,\rangle}{|H(x)|} \ge \frac{-2}{C(\rho,\delta)}\langle\, H(x),\, \exp_x^{-1} a\,\rangle \\
&= \frac{2}{C(\rho,\delta)}\int_{\mathrm{supp}\,\mu} \frac{\langle\, \exp_x^{-1} p,\, \exp_x^{-1} a\,\rangle}{d(x,p)}\mu(dp) \\
&= \frac{2d(x,a)}{C(\rho,\delta)}\int_{\mathrm{supp}\,\mu} \cos\angle\, pxa\ \mu(dp).
\end{aligned}$$

Now it suffices to use Lemma 2.11 to obtain that for every $p \in \mathrm{supp}(\mu)$,

$$\cos\angle\, pxa \ge \frac{S_\Delta(d(x,a)-\sigma)}{S_\Delta(d(x,a)+\sigma)}.$$

□

We are ready to give the desired lower bound.

Lemma 2.25. *For every $x \in \bar{B}(a,\rho)$ such that $H(x) \ne 0$ we have*

$$r_x \ge \frac{\rho - \sigma}{C(\rho,\delta)F(\rho,\Delta) + 1},$$

where the constant $F(\rho,\Delta) \ge 1$ is given by

$$F(\rho,\Delta) = \begin{cases} 1, & \text{if } \Delta \ge 0; \\ \cosh(2\rho\sqrt{-\Delta}), & \text{if } \Delta < 0. \end{cases}$$

Proof. We prove this for the case when $\Delta > 0$. The proof for the cases when $\Delta \leq 0$ is similar. For every $x \in \bar{B}(a,\rho) \setminus \bar{B}(a,\sigma)$, by the preceding lemma we have

$$r_x \geq \frac{2d(x,a)}{C(\rho,\delta)} \frac{\sin(\sqrt{\Delta}(d(x,a)-\sigma))}{\sin(\sqrt{\Delta}(d(x,a)+\sigma))}.$$

Note that $0 < \sqrt{\Delta}(d(x,a) \pm \sigma) < 2\rho\sqrt{\Delta} < \pi/2$ and that $(\sin u / \sin v) \geq (u/v)$ for $0 < u \leq v \leq \pi/2$, then we obtain

$$r_x \geq \frac{2d(x,a)}{C(\rho,\delta)} \frac{d(x,a)-\sigma}{d(x,a)+\sigma} \geq \frac{2d(x,a)}{C(\rho,\delta)} \frac{d(x,a)-\sigma}{2d(x,a)} = \frac{d(x,a)-\sigma}{C(\rho,\delta)}.$$

On the other hand, we always have $r_x \geq \rho - d(x,a)$ and hence

$$r_x \geq \max\{\rho - d(x,a), \frac{d(x,a)-\sigma}{C(\rho,\delta)}\}.$$

Observe that

$$\min\left\{ \max\{\rho - d(x,a), \frac{d(x,a)-\sigma}{C(\rho,\delta)}\} : \sigma < d(x,a) \leq \rho \right\} = \frac{\rho-\sigma}{C(\rho,\delta)+1},$$

then we obtain

$$r_x \geq \frac{\rho-\sigma}{C(\rho,\delta)+1}.$$

Moreover, for every $x \in \bar{B}(a,\sigma)$ such that $H(x) \neq 0$,

$$r_x \geq \rho - \sigma > \frac{\rho-\sigma}{C(\rho,\delta)+1}.$$

The proof is complete. □

Thanks to the above estimation, we get a practically useful version of Theorem 2.23.

Theorem 2.26. *Let $(a_k)_k$ be a sequence in $(0,1]$ such that*

$$\lim_{k\to\infty} a_k = 0 \quad \textit{and} \quad \sum_{k=0}^{\infty} a_k = +\infty.$$

Then we can choose

$$t_k = \frac{(\rho-\sigma)a_k}{C(\rho,\delta)F(\rho,\Delta)+1}$$

in Algorithm 6.17 and, with this choice of stepsizes, the generated sequence $(x_k)_k$ satisfies

$$\lim_{k\to\infty} d(x_k, \mathfrak{M}_\mu) = 0 \quad \textit{and} \quad \lim_{k\to\infty} f(x_k) = f_*.$$

Moreover, if $(a_k)_k$ *also verifies that*

$$\sum_{k=0}^{\infty} a_k^2 < +\infty,$$

then there exists some $m \in \mathfrak{M}_\mu$ *such that* $x_k \to m$.

Proof. This is a simple corollary to Lemma 2.25 and Theorem 2.23. □

Now we turn to the questions of error estimates and the rate of convergence of the subgradient algorithm under condition $*$.

Proposition 2.27. *Let condition* $*$ *hold and the stepsizes* $(t_k)_k$ *satisfy*

$$\lim_{k\to\infty} t_k = 0 \quad \textit{and} \quad \sum_{k=0}^{\infty} t_k = +\infty.$$

Then there exists $N \in \mathbf{N}$, *such that for every* $k \geq N$,

$$d^2(x_k, m) \leq b_k,$$

where m *is the unique median of* μ *and the sequence* $(b_k)_{k\geq N}$ *is defined by*

$$b_N = (\rho+\sigma)^2 \quad \textit{and} \quad b_{k+1} = (1-2\tau t_k)b_k + C(\rho,\delta)t_k^2\,, \quad k \geq N,$$

which converges to 0 *when* $k \to \infty$. *More explicitly, for every* $k \geq N$,

$$b_{k+1} = (\rho+\sigma)^2 \prod_{i=N}^{k}(1-2\tau t_i) + C(\rho,\delta)\Big(\sum_{j=N+1}^{k} t_{j-1}^2 \prod_{i=j}^{k}(1-2\tau t_i) + t_k^2\ \Big).$$

Proof. Since $t_k \to 0$, there exists $N \in \mathbf{N}$ such that for every $k \geq N$ we have $2\tau t_k < 1$. By Theorem 2.15,

$$f(x_k) - f_* \geq \tau d^2(x_k, m).$$

Combining this and Lemma 2.22 we obtain

$$d^2(x_{k+1}, m) \leq (1-2\tau t_k)d^2(x_k, m) + C(\rho,\delta)t_k^2.$$

By Proposition 2.7, $d^2(x_N, m) \leq (\rho+\sigma)^2 = b_N$. Then by induction it is easily seen that $d^2(x_k, m) \leq b_k$ for every $k \geq N$. To prove $b_k \to 0$, we first show

$$\liminf_{k\to\infty} b_k = 0.$$

If this is not true, then there exist $N_1 \geq N$ and $\eta > 0$ such that for every $k \geq N_1$ we have $b_k > \eta$ and $C(\rho,\delta)t_k < \tau\eta$. Thus

$$b_{k+1} = b_k + t_k(C(\rho,\delta)t_k - 2\tau b_k) \leq b_k - \tau\eta t_k.$$

By summing the above inequalities we get

$$\tau\eta \sum_{i=N_1}^{k} t_i \leq b_{N_1} - b_{k+1} \leq b_{N_1},$$

which contradicts $\sum_{k=0}^{\infty} t_k = +\infty$ and the assertion is proved.

For every $k \geq N$, we consider the following two cases:
If $b_k > C(\rho,\delta)t_k/(2\tau)$, then

$$b_{k+1} < b_k - 2\tau t_k(C(\rho,\delta)t_k/(2\tau)) + C(\rho,\delta)t_k^2 = b_k.$$

If $b_k \leq C(\rho,\delta)t_k/(2\tau)$, then

$$b_{k+1} \leq (1 - 2\tau t_k)C(\rho,\delta)t_k/(2\tau) + C(\rho,\delta)t_k^2 = C(\rho,\delta)t_k/(2\tau).$$

Hence we always have

$$b_{k+1} \leq \max\{b_k, C(\rho,\delta)t_k/(2\tau)\},$$

which yields by induction that for every $n \geq k$,

$$b_{n+1} \leq \max\{b_k, (C(\rho,\delta)/(2\tau)) \max\{t_k, t_{k+1}, \dots, t_n\}\}.$$

Then by taking the superior limit on the left hand side and then the inferior limit on the right hand side we conclude that $b_k \to 0$.

Finally, the explicit expressions of $(b_k)_k$ are obtained by induction. □

We proceed to show that if $(t_k)_k$ is chosen to be the harmonic series, then the rate of convergence of our algorithm is sublinear. To do this, we use the following lemma in [70].

Lemma 2.28. *Let $(u_k)_{k\geq 0}$ be a sequence of nonnegative real numbers such that*

$$u_{k+1} \leq \Big(1 - \frac{\alpha}{k+1}\Big)u_k + \frac{\zeta}{(k+1)^2},$$

where α and ζ are positive constants. Then

$$u_{k+1} \leq \begin{cases} \dfrac{1}{(k+2)^{\alpha}}\Big(u_0 + \dfrac{2^{\alpha}\zeta(2-\alpha)}{1-\alpha}\Big), & \text{if} \quad 0 < \alpha < 1; \\[2ex] \dfrac{\zeta(1 + \ln(k+1))}{k+1}, & \text{if} \quad \alpha = 1; \\[2ex] \dfrac{1}{(\alpha-1)(k+2)}\Big(\zeta + \dfrac{(\alpha-1)u_0 - \zeta}{(k+2)^{\alpha-1}}\Big), & \text{if} \quad \alpha > 1. \end{cases}$$

Proposition 2.29. *Let condition $*$ hold and we choose $t_k = r/(k+1)$ for every $k \geq 0$ with some constant $r > 0$, then*

$$d^2(x_{k+1}, m) \leq \begin{cases} \dfrac{1}{(k+2)^\alpha}\left((\rho+\sigma)^2 + \dfrac{2^\alpha r^2 C(\rho,\delta)(2-\alpha)}{1-\alpha}\right), & \text{if} \quad 0 < \alpha < 1; \\ \dfrac{r^2 C(\rho,\delta)}{k+1}(1 + \ln(k+1)), & \text{if} \quad \alpha = 1; \\ \dfrac{1}{(\alpha-1)(k+2)}\left(r^2 C(\rho,\delta) + \dfrac{(\alpha-1)(\rho+\sigma)^2 - r^2 C(\rho,\delta)}{(k+2)^{\alpha-1}}\right), & \text{if} \quad \alpha > 1, \end{cases}$$

where m is the unique median of μ and $\alpha = 2\tau r$.

Proof. As in the proof of Proposition 2.27, we have for every $k \geq 0$,

$$d^2(x_{k+1}, m) \leq (1 - \frac{2\tau r}{k+1}) d^2(x_k, m) + \frac{r^2 C(\rho,\delta)}{(k+1)^2}.$$

Then it suffices to use the preceding lemma with $\alpha = 2\tau r$ and $\zeta = r^2 C(\rho,\delta)$ by observing that $d(x_0, m) \leq \rho + \sigma$. □

Chapter 3

Some properties of Fréchet medians in Riemannian manifolds

Abstract

The consistency of Fréchet medians is proved for probability measures in proper metric spaces. In the context of Riemannian manifolds, assuming that the probability measure has more than a half mass concentrated in a convex ball, the positions of its Fréchet medians are estimated in terms of the upper bound of sectional curvatures and the concentrated mass. It is also shown that, in compact Riemannian manifolds, the Fréchet sample medians of generic data points are always unique.

3.1 Consistency of Fréchet medians in metric spaces

Let (M, d) be a proper metric space (recall that a metric space is proper if and only if every bounded and closed subset is compact) and $P_1(M)$ denote the set of all the probability measures μ on M verifying

$$\int_M d(x_0, p)\mu(dp) < \infty, \text{ for some } x_0 \in M.$$

For every $\mu \in P_1(M)$ we can define a function

$$f_\mu : \qquad M \longrightarrow \mathbf{R}_+ , \qquad x \longmapsto \int_M d(x, p)\mu(dp).$$

This function is 1-Lipschitz hence continuous on M. Since M is proper, f_μ attains its minimum (see [76, p. 42]), so we can give the following definition:

Definition 3.1. Let μ be a probability measure in $P_1(M)$, then a *global* minimum point of f_μ is called a Fréchet median of μ. The set of all the Fréchet medians of μ is denoted by Q_μ. Let f^*_μ denote the global minimum of f_μ.

Observe that Q_μ is compact, since the triangle inequality implies that $d(x,y) \leq 2f_\mu^*$ for every $x, y \in Q_\mu$.

To introduce the next proposition, let us recall that the L^1-Wasserstein distance between two elements μ and ν in $P_1(M)$ is defined by

$$W_1(\mu,\nu) = \inf_{\pi \in \Pi(\mu,\nu)} \int_{M\times M} d(x,y) d\pi(x,y),$$

where $\Pi(\mu,\nu)$ is the set of all the probability measures on $M \times M$ with margins μ and ν. As a useful case for us, observe that $f_\mu(x) = W_1(\delta_x, \mu)$ for every $x \in M$. The set of all the 1-Lipschitz functions on M is denoted by $\mathrm{Lip}_1(M)$.

As is well known that Riemannian barycenters are characterized by convex functions (see [55, Lemma 7.2]), the following proposition shows that Fréchet medians can be characterized by Lipschitz functions.

Proposition 3.2. *Let $\mu \in P_1(M)$ and M be also separable, then*

$$Q_\mu = \left\{ x \in M : \varphi(x) \leq f_\mu^* + \int_M \varphi(p)\mu(dp), \ \textit{for every } \varphi \in \mathrm{Lip}_1(M) \right\}.$$

Proof. The separability of M ensures that the duality formula of Kantorovich-Rubinstein (see [88, p. 107]) can be applied, so that for every $x \in M$,

$$\begin{aligned} x \in Q_\mu &\iff f_\mu(x) \leq f_\mu^* \\ &\iff W_1(\delta_x, \mu) \leq f_\mu^* \\ &\iff \sup_{\varphi \in \mathrm{Lip}_1(M)} \left| \varphi(x) - \int_M \varphi(p)\mu(dp) \right| \leq f_\mu^* \\ &\iff \varphi(x) \leq f_\mu^* + \int_M \varphi(p)\mu(dp), \ \text{for every } \varphi \in \mathrm{Lip}_1(M), \end{aligned}$$

as desired. $\square$

We proceed to show the main result of this section.

Theorem 3.3. *Let $(\mu_n)_{n\in\mathbf{N}}$ be a sequence in $P_1(M)$ and μ be another probability measure in $P_1(M)$. If $(f_{\mu_n})_n$ converges uniformly on M to f_μ, then for every $\varepsilon > 0$, there exists $N \in \mathbf{N}$, such that for every $n \geq N$ we have*

$$Q_{\mu_n} \subset B(Q_\mu, \varepsilon) := \{x \in M : d(x, Q_\mu) < \varepsilon\}.$$

Proof. We prove this by contradiction. Suppose that the assertion is not true, then without loss of generality, we can assume that there exist some $\varepsilon > 0$ and a sequence $(x_n)_{n\in\mathbf{N}}$ such that $x_n \in Q_{\mu_n}$ and $x_n \notin B(Q_\mu, \varepsilon)$ for every n. If this sequence is bounded, then by choosing a subsequence we can assume that $(x_n)_n$ converges to a point $x_* \notin B(Q_\mu, \varepsilon)$ because $M \setminus B(Q_\mu, \varepsilon)$ is closed. However, observe that the uniform convergence of $(f_{\mu_n})_n$ to f_μ implies $f^*_{\mu_n} \longrightarrow f^*_\mu$, hence one gets

$$\begin{aligned}
|f_\mu(x_*) - f^*_\mu| &\leq |f_\mu(x_*) - f_{\mu_n}(x_*)| + |f_{\mu_n}(x_*) - f_{\mu_n}(x_n)| + |f_{\mu_n}(x_n) - f^*_\mu| \\
&\leq \sup_{x \in M} |f_{\mu_n}(x) - f_\mu(x)| + d(x_*, x_n) + |f^*_{\mu_n} - f^*_\mu| \longrightarrow 0.
\end{aligned}$$

So that $f_\mu(x_*) = f^*_\mu$, that is to say $x_* \in Q_\mu$. This is impossible, hence $(x_n)_n$ is not bounded. Now we fix a point $\bar{x} \in Q_\mu$, always by choosing a subsequence we can assume that $d(x_n, \bar{x}) \longrightarrow +\infty$, then

$$\begin{aligned}
f_\mu(x_n) &= \int_M d(x_n, p)\mu(dp) \geq \int_M (d(x_n, \bar{x}) - d(\bar{x}, p))\mu(dp) \\
&= d(x_n, \bar{x}) - f^*_\mu \longrightarrow +\infty. \qquad (1.1)
\end{aligned}$$

On the other hand,

$$\begin{aligned}
|f_\mu(x_n) - f^*_\mu| &\leq |f_\mu(x_n) - f_{\mu_n}(x_n)| + |f_{\mu_n}(x_n) - f^*_\mu| \\
&\leq \sup_{x \in M} |f_{\mu_n}(x) - f_\mu(x)| + |f^*_{\mu_n} - f^*_\mu| \longrightarrow 0.
\end{aligned}$$

This contradicts (1.1), the proof is complete. □

Remark 3.4. A sufficient condition to ensure the uniform convergence of $(f_{\mu_n})_n$ on M to f_μ is that $W_1(\mu_n, \mu) \longrightarrow 0$, since

$$\sup_{x \in M} |f_{\mu_n}(x) - f_\mu(x)| = \sup_{x \in M} |W_1(\delta_x, \mu_n) - W_1(\delta_x, \mu)| \leq W_1(\mu_n, \mu).$$

The consistency of Fréchet means is proved in [25, Theorem 2.3]. The consistency of Fréchet medians given below is a corollary to Theorem 3.3. A similar result can be found in [76, p. 44].

Corollary 3.5. *Let $(X_n)_{n \in \mathbf{N}}$ be a sequence of i.i.d random variables of law $\mu \in P_1(M)$ and $(m_n)_{n \in \mathbf{N}}$ be a sequence of random variables such that $m_n \in Q_{\mu_n}$ with $\mu_n = \frac{1}{n}\sum_{k=1}^n \delta_{X_k}$. If μ has a unique Fréchet median m, then $m_n \longrightarrow m$ a.s.*

Proof. By Theorem 3.3 and Remark 3.4, it suffices to show that $\mu_n \xrightarrow{W_1} \mu$ a.s. This is equivalent to show that (see [88, p. 108]) for every $f \in C_b(M)$,

$$\frac{1}{n}\sum_{k=1}^n f(X_k) \longrightarrow \int_M f(p)\mu(dp) \quad \text{a.s.}$$

and for every $x \in M$,

$$\frac{1}{n}\sum_{k=1}^n d(x, X_k) \longrightarrow \int_M d(x,p)\mu(dp) \quad \text{a.s.}$$

These two assertions are trivial corollaries to the strong law of large numbers, hence the result holds. □

3.2 Robustness of Fréchet medians in Riemannian manifolds

Throughout this section, we assume that M is a complete Riemannian manifold with dimension no less than 2, whose Riemannian distance is denoted by d. We fix a closed geodesic ball

$$\bar{B}(a,\rho) = \{x \in M : d(x,a) \leq \rho\}$$

in M centered at a with a finite radius $\rho > 0$ and a probability measure $\mu \in P_1(M)$ such that

$$\mu(\bar{B}(a,\rho)) = \alpha > \frac{1}{2}.$$

The aim of this section is to estimate the positions of the Fréchet medians of μ, which gives a quantitative estimation for robustness. To this end, the following type of functions are of fundamental importance for our methods. Let $x, z \in M$, define

$$h_{x,z} : \quad \bar{B}(a,\rho) \longrightarrow \mathbf{R}, \quad p \longmapsto d(x,p) - d(z,p).$$

Obviously, $h_{x,z}$ is continuous and attains its minimum.

Our method of estimating the position of Q_μ is essentially based on the following simple observation.

Proposition 3.6. *Let $x \in \bar{B}(a,\rho)^c$ and assume that there exists $z \in M$ such that*

$$\min_{p \in \bar{B}(a,\rho)} h_{x,z}(p) > \frac{1-\alpha}{\alpha}\, d(x,z),$$

then $x \notin Q_\mu$.

Proof. Clearly one has

$$\begin{aligned} f_\mu(x) - f_\mu(z) &= \int_{\bar{B}(a,\rho)} (d(x,p) - d(z,p))\mu(dp) + \int_{M \setminus \bar{B}(a,\rho)} (d(x,p) - d(z,p))\mu(dp) \\ &\geq \alpha \min_{p \in \bar{B}(a,\rho)} h_{x,z}(p) - (1-\alpha) d(x,z) > 0. \end{aligned}$$

The proof is complete. □

By choosing the dominating point $z = a$ in Proposition 3.6 we get the following basic estimation.

Theorem 3.7. *The set Q_μ of all the Fréchet medians of μ verifies*

$$Q_\mu \subset \bar{B}\left(a, \frac{2\alpha\rho}{2\alpha - 1}\right).$$

Proof. Observe that for every $p \in \bar{B}(a, \rho)$,

$$h_{x,a}(p) = d(x, p) - d(a, p) \geq d(x, a) - 2d(a, p) \geq d(x, a) - 2\rho.$$

Hence Proposition 3.6 yields

$$\begin{aligned} Q_\mu \cap \bar{B}(a, \rho)^c &\subset \Big\{x \in M : \min_{p \in \bar{B}(a,\rho)} h_{x,a}(p) \leq \frac{1-\alpha}{\alpha} d(x, a)\Big\} \\ &\subset \Big\{x \in M : d(x, a) - 2\rho \leq \frac{1-\alpha}{\alpha} d(x, a)\Big\} \\ &= \Big\{x \in M : d(x, a) \leq \frac{2\alpha\rho}{2\alpha - 1}\Big\}. \end{aligned}$$

The proof is complete. □

Remark 3.8. It is easily seen that the conclusions of Proposition 3.6 and Theorem 3.7 also hold if M is only a proper metric space.

Remark 3.9. As a direct corollary to Theorem 3.7, if μ is a probability measure in $P_1(M)$ such that for some point $m \in M$ one has $\mu\{m\} > 1/2$, then m is the unique Fréchet median of μ.

Thanks to Theorem 3.7, from now on we only have to work in the closed geodesic ball

$$B_* = \bar{B}\left(a, \frac{2\alpha\rho}{2\alpha - 1}\right).$$

Thus let Δ be an upper bound of sectional curvatures in B_* and inj be the injectivity radius of B_*. Moreover, we shall always assume that the following concentration condition is fulfilled throughout the rest part of this section:

Assumption 3.10.

$$\frac{2\alpha\rho}{2\alpha - 1} < r_* := \min\Big\{\frac{\pi}{\sqrt{\Delta}}, \mathrm{inj}\Big\},$$

where if $\Delta \leq 0$, then $\pi/\sqrt{\Delta}$ is interpreted as $+\infty$.

In view of Proposition 3.6 and Theorem 3.7, estimating the position of Q_μ can be achieved by estimating the minimum of the functions $h_{x,z}$ for some $x, z \in B_*$. The following lemma enables us to use the comparison argument proposed in [1] to compare the configurations in B_* with the ones in model spaces in order to obtain lower bounds of the functions $h_{x,z}$.

Lemma 3.11. *Let $x \in B_* \setminus \bar{B}(a, \rho)$ and y be the intersection point of the boundary of $\bar{B}(a, \rho)$ and the minimal geodesic joining x and a. Let $z \neq x$ be another point on the minimal geodesic joining x and a. Assume that $d(a, x) + d(a, z) < r_*$, then*

$$\operatorname{argmin} h_{x,z} \subset \{p \in \bar{B}(a, \rho) :\ d(x, p) + d(p, z) + d(z, x) < 2r_*\}.$$

Proof. Let $p \in \bar{B}(a, \rho)$ such that $d(x,p) + d(p,z) + d(z,x) \geq 2r_*$, then

$$\begin{aligned} h_{x,z}(p) &\geq 2r_* - d(x,z) - 2d(z,p) \\ &> 2(d(a,x) + d(a,z)) - d(x,z) - 2(d(a,z) + \rho) \\ &= d(x,y) - d(a,y) + d(a,z). \end{aligned} \tag{2.2}$$

If $d(a,y) > d(a,z)$, then (2.2) yields $h_{x,z}(p) > h_{x,z}(y)$, thus p cannot be a minimum point of $h_{x,z}$. On the other hand, if $d(a,y) \leq d(a,z)$, then (2.2) gives that $h_{x,z}(p) > d(x,y) + d(y,z) \geq d(x,z)$, which is impossible. Hence in either case, every minimum point p of $h_{x,z}$ must verify $d(x,p) + d(p,z) + d(z,x) < 2r_*$. □

As a preparation for the comparison arguments in the following, let us recall the definition of model spaces. For a real number κ, the model space $\mathbb{M}^2_\kappa$ is defined as follows:
1) if $\kappa > 0$, then $\mathbb{M}^2_\kappa$ is obtained from the sphere $\mathbb{S}^2$ by multiplying the distance function by $1/\sqrt{\kappa}$;
2) if $\kappa = 0$, then $\mathbb{M}^2_\kappa$ is the Euclidean space $\mathbb{E}^2$;
3) if $\kappa < 0$, then $\mathbb{M}^2_\kappa$ is obtained from the hyperbolic space $\mathbb{H}^2$ by multiplying the distance function by $1/\sqrt{-\kappa}$.
Moreover, the distance between two points $\bar{x}$ and $\bar{y}$ in $\mathbb{M}^2_\kappa$ will be denoted by $\bar{d}(\bar{x}, \bar{y})$.

The following proposition says that for the positions of Fréchet medians, if comparisons can be done, then the model space $\mathbb{M}^2_\Delta$ is the worst case.

Proposition 3.12. *Consider in $\mathbb{M}^2_\Delta$ the same configuration as that in Lemma 3.11: a closed geodesic ball $\bar{B}(\bar{a}, \rho)$ and a point $\bar{x}$ such that $\bar{d}(\bar{x}, \bar{a}) = d(x, a)$. We denote $\bar{y}$ the intersection point of the boundary of $\bar{B}(\bar{a}, \rho)$ and the minimal geodesic joining $\bar{x}$ and $\bar{a}$. Let $\bar{z}$ be a point in the minimal geodesic joining $\bar{x}$ and $\bar{a}$ such that $\bar{d}(\bar{a}, \bar{z}) = d(a, z)$. Assume that $d(a,x) + d(a,z) < r_*$, then*

$$\min_{p \in \bar{B}(a,\rho)} h_{x,z}(p) \geq \min_{\bar{p} \in \bar{B}(\bar{a},\rho)} \bar{h}_{\bar{x},\bar{z}}(\bar{p}),$$

where $\bar{h}_{\bar{x},\bar{z}}(\bar{p}) := \bar{d}(\bar{x}, \bar{p}) - \bar{d}(\bar{z}, \bar{p})$.

Proof. Let $p \in \operatorname{argmin} h_{x,z}$. Consider a comparison point $\bar{p} \in \mathbb{M}^2_\Delta$ such that $\bar{d}(\bar{z}, \bar{p}) = d(z, p)$ and $\angle\, \bar{a}\bar{z}\bar{p} = \angle\, azp$. Then the assumption $d(a,x) + d(a,z) < r_*$ and the hinge version of Alexandrov-Toponogov comparison theorem (see [32, Exercise IX.1, p. 420]) yield that $\bar{d}(\bar{a}, \bar{p}) \leq d(a,p) = \rho$, i.e. $\bar{p} \in \bar{B}(\bar{a}, \rho)$. Now by hinge comparison again and Lemma 3.11, we get $\bar{d}(\bar{p}, \bar{x}) \leq d(p, x)$, which implies that

$$h_{x,z}(p) \geq \bar{h}_{\bar{x},\bar{z}}(\bar{p}) \geq \min_{\bar{p} \in \bar{B}(\bar{a},\rho)} \bar{h}_{\bar{x},\bar{z}}(\bar{p}).$$

The proof is complete. □

According to Proposition 3.12, it suffices to find the minima of the functions $h_{x,z}$ when M equals $\mathbb{S}^2$, $\mathbb{E}^2$ and $\mathbb{H}^2$, which are of constant curvatures 1, 0 and -1, respectively.

Proposition 3.13. *Let $t, u \geq 0$ such that $u < \rho + t \leq 2\alpha\rho/(2\alpha - 1)$.*

i) If $M = \mathbb{S}^2$, let $x = (\sin(\rho+t), 0, \cos(\rho+t))$ and $z = (\sin u, 0, \cos u)$. Assume that $\rho + t + u < \pi$, then

$$\min_{\bar{B}(a,\rho)} h_{x,z} = \begin{cases} t - \rho + u, & \text{if} \quad \cot u \geq 2\cot\rho - \cot(\rho+t); \\ \\ \arccos\left(\cos(\rho+t-u) + \dfrac{\sin^2\rho \sin^2(\rho+t-u)}{2\sin u \sin(\rho+t)}\right), & \text{if} \quad \text{not.} \end{cases}$$

ii) If $M = \mathbb{E}^2$, let $a = (0,0)$, $x = (\rho + t, 0)$, $z = (u, 0)$, then

$$\min_{\bar{B}(a,\rho)} h_{x,z} = \begin{cases} t - \rho + u, & \text{if} \quad u \leq \dfrac{(\rho+t)\rho}{\rho+2t}; \\ \\ (\rho+t-u)\sqrt{1 - \dfrac{\rho^2}{u(\rho+t)}}, & \text{if} \quad \text{not.} \end{cases}$$

iii) If $M = \mathbb{H}^2$, let $a = (0,0,1)$, $x = (\sinh(\rho+t), 0, \cosh(\rho+t))$ and $z = (\sinh u, 0, \cosh u)$, then

$$\min_{\bar{B}(a,\rho)} h_{x,z} = \begin{cases} t - \rho + u, & \text{if} \quad \coth u \geq 2\coth\rho - \coth(\rho+t); \\ \\ \operatorname{arccosh}\left(\cosh(\rho+t-u) - \dfrac{\sinh^2\rho \sinh^2(\rho+t-u)}{2\sinh u \sinh(\rho+t)}\right), & \text{if} \quad \text{not.} \end{cases}$$

We shall only prove the result for the case when $M = \mathbb{S}^2$, since the proofs for $M = \mathbb{E}^2$ and $M = \mathbb{H}^2$ are similar and easier. The proof consists of some lemmas, the first one below says that $h_{x,z}$ is smooth at its minimum points which can only appear on the boundary of the ball $\bar{B}(a, \rho)$.

Lemma 3.14. *Let x' and z' be the antipodes of x and z. Then $z' \notin \bar{B}(a, \rho)$ and all the local minimum points of $h_{x,z}$ are contained in $\partial\bar{B}(a,\rho) \setminus \{x'\}$.*

Proof. It is easily seen that $d(z', a) = \pi - u > \rho$, so that $z' \notin \bar{B}(a, \rho)$. Observe that x' is a global maximum point of $h_{x,z}$ which is not locally constant, so that x' cannot be a local minimum. Now let $p \in B(a, \rho)$ be a local minimum of $h_{x,z}$, then $h_{x,z}$ is smooth at p. It follows that $\operatorname{grad} h_{x,z}(p) = 0$, which yields that $h_{x,z}(p) = d(x, z)$, this is a contradiction. The proof is complete. $\square$

The following lemma characterizes the global minimum points of $h_{x,z}$.

Lemma 3.15. *The set of global minimum points of $h_{x,z}$ verifies*

$$\operatorname{argmin} h_{x,z} = \begin{cases} \{y\}, & \text{if } \cot u \geq 2\cot\rho - \cot(\rho+t); \\ \{\, p \in \partial\bar{B}(a,\rho) : \dfrac{\sin(\rho+t)}{\sin d(x,p)} = \dfrac{\sin u}{\sin d(z,p)}\}, & \text{if not,} \end{cases}$$

where y is the intersection point of the boundary of $\bar{B}(a,\rho)$ and the minimal geodesic joining x and a.

Proof. Thanks to Lemma 3.14, it suffices to find the global minimum points of $h_{x,z}$ for $p = (\sin\rho\cos\theta, \sin\rho\sin\theta, \cos\rho)$ and $\theta \in [0, 2\pi)$. In this case,

$$\begin{aligned} h_{x,z}(p) &= d(x,p) - d(z,p) \\ &= \arccos(\sin(\rho+t)\sin\rho\cos\theta + \cos(\rho+t)\cos\rho) \\ &\quad - \arccos(\sin u \sin\rho\cos\theta + \cos u\cos\rho) \\ &:= h(\theta). \end{aligned}$$

Hence let $p = (\sin\rho\cos\theta, \sin\rho\sin\theta, \cos\rho)$ be a local minimum point of $h_{x,z}$, then Lemma 3.14 yields that $h'(\theta)$ exists and equals zero. On the other hand, by elementary calculation,

$$\begin{aligned} h'(\theta) &= \sin\rho\sin\theta\bigg(\frac{\sin(\rho+t)}{\sqrt{1-(\sin(\rho+t)\sin\rho\cos\theta+\cos(\rho+t)\cos\rho)^2}} \\ &\qquad - \frac{\sin u}{\sqrt{1-(\sin u\sin\rho\cos\theta+\cos u\cos\rho)^2}}\bigg) \\ &= \sin\rho\sin\theta\bigg(\frac{\sin(\rho+t)}{\sin d(x,p)} - \frac{\sin u}{\sin d(z,p)}\bigg), \end{aligned}$$

where the second equality is because of the spherical law of cosines. Hence we have necessarily

$$\theta = 0,\ \pi \quad \text{or} \quad \frac{\sin(\rho+t)}{\sin d(x,p)} = \frac{\sin u}{\sin d(z,p)}.$$

Firstly, we observe that w, the corresponding point p when $\theta = \pi$, cannot be a minimum point. In fact, let w' be the antipode of w. If $d(x,a) < d(w',a)$, then $h_{x,z}(w) = d(x,z)$. So that w is a maximum point. On the other hand, if $d(x,a) \geq d(w',a)$, then $d(w,x)+d(x,z)+d(z,w) \equiv 2\pi$. Hence Lemma 3.11 and the condition $\rho+t+u < \pi$ imply that w is not a minimum point. So that the assertion holds.

Now assume that $p \neq w, y$ such that

$$\frac{\sin(\rho+t)}{\sin d(x,p)} = \frac{\sin u}{\sin d(z,p)}. \tag{2.3}$$

Let $\beta = \angle\, zpa$. Then by the spherical law of sines, (2.3) is equivalent to $\sin(\beta + \angle\, zpx) = \sin\beta$, i.e. that $\angle\, zpx = \pi - 2\beta$. Applying the spherical law of sines to $\triangle zpx$ we get

$$\frac{\sin\angle\, x}{\sin d(z,p)} = \frac{\sin 2\beta}{\sin(\rho + t - u)}. \tag{2.4}$$

By the spherical law of sines in $\triangle apx$,

$$\frac{\sin\angle\, x}{\sin\rho} = \frac{\sin\beta}{\sin(\rho + t)}. \tag{2.5}$$

Then (2.4)/(2.5) gives that

$$\sin d(z,p)\cos\beta = \frac{\sin\rho\sin(\rho + t - u)}{2\sin(\rho + t)}. \tag{2.6}$$

By the spherical law of cosines in $\triangle azp$,

$$\sin d(z,p)\cos\beta = \frac{\cos u - \cos\rho\cos d(z,p)}{\sin\rho}. \tag{2.7}$$

Then (2.6) and (2.7) give that

$$\cos d(z,p) = \frac{2\cos u\sin(\rho + t) - \sin^2\rho\sin(\rho + t - u)}{2\cos\rho\sin(\rho + t)}. \tag{2.8}$$

Moreover, by (2.8) and spherical law of cosines in $\triangle azp$,

$$\cos\theta = \frac{\tan\rho}{2}(\cot u + \cot(\rho + t)). \tag{2.9}$$

The condition $0 < \rho + t + u < \pi$ and (2.9) give that

$$\cos\theta = \frac{\tan\rho\sin(\rho + t + u)}{2\sin u\sin(\rho + t)} > 0. \tag{2.10}$$

Furthermore, considering $p \neq y$ we must have $\cos\theta < 1$. By (2.9) this is equivalent to

$$\cot u < 2\cot\rho - \cot(\rho + t), \tag{2.11}$$

which is also equivalent to

$$\frac{\sin(\rho - u)}{\sin u} < \frac{\sin t}{\sin(\rho + t)}. \tag{2.12}$$

For the case when $u \geq \rho$ it is easily seen that y is a maximum point of $h_{x,z}$ and hence p must verify (2.3) and the corresponding θ is determined by (2.9). Hence in this case, there are exactly two local minimum points of $h_{x,z}$ and obviously they are also global ones.

Now let $u < \rho$, then easy computation gives

$$h''(0) = \sin\rho\left(\frac{\sin(\rho+t)}{\sin t} - \frac{\sin u}{\sin(\rho-u)}\right). \tag{2.13}$$

Hence if $\frac{\sin(\rho+t)}{\sin t} \geq \frac{\sin u}{\sin(\rho-u)}$, then (2.12) yields that y is the unique global minimum of $h_{x,z}$. In the opposite case, (2.13) implies that y is a local maximum point. Hence the same argument as in the case when $u \geq \rho$ completes the proof of lemma. □

We need the following technical lemma.

Lemma 3.16. *If* $\cot u < 2\cot\rho - \cot(\rho+t)$, *then every* $p \in \operatorname{argmin} h_{x,z}$ *verifies* $0 < d(x,p) - d(z,p) < \pi$.

Proof. It suffices to show $d(x,p) > d(z,p)$. For the case when $u < \rho$, we firstly show that $d(z,p) < \pi/2$. In fact, by (2.8) this is equivalent to show that

$$(1+\cos^2\rho)\cos u\sin(\rho+t) + \sin^2\rho\sin u\cos(\rho+t) > 0 \tag{2.14}$$

If $\rho + t \leq \pi/2$, (2.14) is trivially true. Now assume $\rho + t > \pi/2$. So that $\pi/2 < \rho + t < \pi - u$, which implies $\sin(\rho+t) > \sin u$ and $\cos(\rho+t) > -\cos u$. Hence we get

$$\begin{aligned}
&(1+\cos^2\rho)\cos u\sin(\rho+t) + \sin^2\rho\sin u\cos(\rho+t)\\
>&(1+\cos^2\rho)\cos u\sin u - \sin^2\rho\sin u\cos u\\
=&2\cos^2\rho\cos u\sin u > 0.
\end{aligned}$$

So that $d(z,p) < \pi/2$ holds. Now if $d(x,p) \geq \pi/2$, then obviously $d(x,p) > d(z,p)$. So that assume $d(x,p) < \pi/2$. Observe that $\rho+t+u < \pi$ implies $\sin u < \sin(\rho+t)$, then (2.3) yields $d(x,p) > d(z,p)$.

For the case when $u \geq \rho$, it suffices to show that $\cos d(z,p) > \cos d(x,p)$ for every $p = (\sin\rho\cos\theta, \sin\rho\sin\theta, \cos\rho)$ with $\theta \in [0,\pi]$. Now let

$$\begin{aligned}
g(\theta) &= \sin\rho\cos\theta(\sin u - \sin(\rho+t)) + \cos\rho(\cos u - \cos(\rho+t))\\
&= \cos d(z,p) - \cos d(x,p)
\end{aligned}$$

Then $g'(\theta) = -\sin\rho\sin\theta(\sin u - \sin(\rho+t))$. Observe that $\rho+t+u < \pi$ and $u < \rho+t$ imply that $\sin u < \sin(\rho+t)$, hence $g(\theta) \geq g(0) = \cos d(z,y) - \cos d(x,y) > 0$. The proof is complete. □

Proof of Proposition 3.13. By Lemma 3.15, it suffices to consider the case when $\cot u < 2\cot\rho - \cot(\rho+t)$. Let $p \in \operatorname{argmin} h_{x,z}$, then by (2.9) and the spherical law of cosines in $\triangle apx$,

$$\cos d(x,p) = \frac{2\cos(\rho+t)\sin u + \sin^2\rho\sin(\rho+t-u)}{2\cos\rho\sin u}. \tag{2.15}$$

Now let $u = \rho - v$, then $\rho + t - u = t + v$. So that (2.8) and (2.15) become

$$\cos d(z,p) = \frac{2\cos(\rho - v)\sin(\rho + t) - \sin^2\rho\sin(t+v)}{2\cos\rho\sin(\rho+t)}. \tag{2.16}$$

$$\cos d(x,p) = \frac{2\cos(\rho + t)\sin(\rho - v) + \sin^2\rho\sin(t+v)}{2\cos\rho\sin(\rho-v)}. \tag{2.17}$$

It follows that

$$\begin{aligned}\cos d(z,p)\cos d(x,p) = &(4\cos(\rho - v)\sin(\rho+t)\cos(\rho+t)\sin(\rho-v)\\ &+ 2\sin^2\rho\sin^2(t+v) - \sin^4\rho\sin^2(t+v))\\ &/(4\cos^2\rho\sin(\rho+t)\sin(\rho-v)).\end{aligned} \tag{2.18}$$

On the other hand, (2.3) and (2.16) yield that

$$\begin{aligned}\sin d(z,p)\sin d(x,p) &= (1-\cos^2 d(z,p))\frac{\sin(\rho+t)}{\sin(\rho-v)}\\ &= (4\sin^2(\rho+t)(\cos^2\rho - \cos^2(\rho-v)) - \sin^4\rho\sin^2(t+v)\\ &\quad + 4\sin^2\rho\sin(t+v)\sin(\rho+t)\cos(\rho-v))\\ &\quad /(4\cos^2\rho\sin(\rho+t)\sin(\rho-v)).\end{aligned} \tag{2.19}$$

Then by (2.18) and (2.19) we obtain

$$\begin{aligned}&4\cos^2\rho\sin(\rho+t)\sin(\rho-v)(\cos(d(x,p)-d(z,p)) - \cos(t+v))\\ =&4\cos^2\rho\sin(\rho+t)\sin(\rho-v)(\cos d(x,p)\cos d(z,p) + \sin d(x,p)\sin d(z,p) - \cos(t+v))\\ =&4\cos(\rho-v)\sin(\rho-v)\cos(\rho+t)\sin(\rho+t) + 2\sin^2\rho\cos^2\rho\sin^2(t+v)\\ &+ 4\sin^2(\rho+t)(\cos^2\rho - \cos^2(\rho-v)) + 4\sin^2\rho\sin(\rho+t)\cos(\rho-v)\sin(t+v))\\ &- 4\cos^2\rho\sin(\rho+t)\sin(\rho-v)\cos(t+v)\\ =&(-4\cos^4\rho\cos^2 v\sin^2 t - 4\cos^4\rho\sin^2 t\sin^2 v + 4\cos^4\rho\sin^2 t)\\ &+ (-8\cos^3\rho\cos t\cos^2 v\sin\rho\sin t - 8\cos^3\rho\cos t\sin\rho\sin t\sin^2 v + 8\cos^3\rho\cos t\sin\rho\sin t)\\ &+ (-4\cos^2\rho\cos^2 t\cos^2 v\sin^2\rho - 2\cos^2\rho\cos^2 t\sin^2\rho\sin^2 v + 4\cos^2\rho\cos^2 t\sin^2\rho)\\ &+ 4\cos^2\rho\sin^2\rho\sin v\cos v\sin t\cos t + 2\cos^2\rho\cos^2 v\sin^2\rho\sin^2 t\\ =&2\cos^2\rho\sin^2\rho\cos^2 t\sin^2 v + 4\cos^2\rho\sin^2\rho\sin v\cos v\sin t\cos t + 2\cos^2\rho\cos^2 v\sin^2\rho\sin^2 t\\ =&2\cos^2\rho\sin^2\rho\sin^2(t+v).\end{aligned}$$

As a result,

$$\begin{aligned}\cos(d(x,p)-d(z,p)) &= \cos(t+v) + \frac{2\cos^2\rho\sin^2\rho\sin^2(t+v)}{4\cos^2\rho\sin(\rho+t)\sin(\rho-v)}\\ &= \cos(t+v) + \frac{\sin^2\rho\sin^2(t+v)}{2\sin(\rho+t)\sin(\rho-v)}\\ &= \cos(\rho+t-u) + \frac{\sin^2\rho\sin^2(\rho+t-u)}{2\sin u\sin(\rho+t)}.\end{aligned}$$

Now it suffices to use Lemma 3.16 to finish the proof. □

We also need the following lemma.

Lemma 3.17. *Let κ be real number and $1/2 < \alpha \leq 1$. For $t \in (0, \rho/(2\alpha-1)]$ define*

$$F_{\alpha,\rho,\kappa}(t) = \begin{cases} \cot(\sqrt{\kappa}(2\alpha-1)t) - \cot(\sqrt{\kappa}t) - 2\cot(\sqrt{\kappa}\rho), & \text{if } \kappa > 0; \\ (1-\alpha)\rho - (2\alpha-1)t, & \text{if } \kappa = 0; \\ \coth(\sqrt{-\kappa}(2\alpha-1)t) - \coth(\sqrt{-\kappa}t) - 2\coth(\sqrt{-\kappa}\rho), & \text{if } \kappa < 0. \end{cases}$$

Assume that $1/2 < \alpha < 1$, then there exists a unique $t_\kappa \in (0, \rho/(2\alpha-1))$ such that

$$\left\{ t \in (0, \frac{\rho}{2\alpha-1}] : \ F_{\alpha,\rho,\kappa}(t) \geq 0 \right\} = (0, t_\kappa].$$

In this case, when $\kappa \leq 0$, the function $F_{\alpha,\rho,\kappa}$ is strictly deceasing.

Proof. We only prove the case when $\kappa = 1$, since the proof of the other two cases are similar and easier. Observe that $F_{\alpha,\rho,1}(0+) = +\infty$ (since $1/2 < \alpha < 1$) and $F_{\alpha,\rho,1}(\rho/(2\alpha-1)) < 0$ (since $2\alpha\rho/(2\alpha-1) < \pi$), then there exists some $t_1 \in (0, \rho/(2\alpha-1))$ such that $F_{\alpha,\rho,1}(t_1) = 0$. Moreover,

$$F'_{\alpha,\rho,1}(t) = \frac{1}{\sin^2((2\alpha-1)t)}\Big((\frac{\sin((2\alpha-1)t)}{\sin t})^2 - (2\alpha-1) \Big).$$

Observe that the function $l(t) = \sin((2\alpha-1)t)/\sin t$ is strictly increasing on $(0, \pi/(2\alpha)]$, $l^2(0+) = (2\alpha-1)^2 < 2\alpha-1$ and $l^2(\pi/(2\alpha)) = 1 > 2\alpha-1$. Hence there exists a unique $s \in (0, \pi/(2\alpha))$ such that if $t < s$, then $F'_{\alpha,\rho,1}(t) < 0$; if $t = s$, then $F'_{\alpha,\rho,1}(t) = 0$; if $t > s$, then $F'_{\alpha,\rho,1}(t) > 0$. Hence $F_{\alpha,\rho,1}$ is strictly decreasing on $(0, s]$ and strictly increasing on $[s, \rho/(2\alpha-1)]$. Since $F_{\alpha,\rho,1}(\rho/(2\alpha-1)) < 0$, the point t_1 must be unique. Moreover, it is easily seen that $\{F_{\alpha,\rho,1} \geq 0\} = (0, t_1]$. The proof is complete. □

The main theorem of this section is justified by the lemma below.

Lemma 3.18. *Assumption 3.10 implies that*

$$\frac{\alpha S_\Delta(\rho)}{\sqrt{2\alpha-1}} < S_\Delta(\frac{r_*}{2}), \ \text{where } S_\Delta(t) := \begin{cases} \sin(\sqrt{\Delta}\, t), & \text{if } \Delta > 0; \\ t, & \text{if } \Delta = 0; \\ \sinh(\sqrt{-\Delta}\, t), & \text{if } \Delta < 0. \end{cases}$$

Proof. We only prove the case when $\Delta > 0$, since the proof for the cases when $\Delta \leq 0$ are easier. Without loss of generality, we can assume that $\Delta = 1$. Since

$$\frac{2\alpha\rho}{2\alpha-1} < r_* \Longleftrightarrow \sin\rho < \sin\frac{2\alpha-1}{2\alpha} r_*,$$

it is sufficient to show that

$$\frac{\alpha}{\sqrt{2\alpha-1}}\sin\frac{2\alpha-1}{2\alpha}r_* < \sin\frac{r_*}{2}.$$

To this end, let $c = r_*/2 \in (0, \pi/2]$, we will show that the function

$$f(\alpha) = \frac{\alpha}{\sqrt{2\alpha-1}}\sin\frac{2\alpha-1}{\alpha}c$$

is strictly increasing for $\alpha \in (1/2, 1)$. Easy computation gives that

$$f'(\alpha) > 0 \Longleftrightarrow \frac{\tan(\theta c)}{\theta c} < \frac{2-\theta}{1-\theta},$$

where $\theta = (2\alpha-1)/\alpha \in (0,1)$. Observe that the function $x \longmapsto \tan x/x$ is increasing on $[0, \pi/2)$, hence it suffices to show that

$$\frac{\tan(\theta\pi/2)}{(\theta\pi/2)} < \frac{2-\theta}{1-\theta}.$$

This is true because Becker-Stark inequality (see [22]) yields

$$\frac{\tan(\theta\pi/2)}{(\theta\pi/2)} < \frac{1}{1-\theta^2} < \frac{2-\theta}{1-\theta}.$$

The proof is complete. □

Now we are ready to give the main result of this section.

Theorem 3.19. *The following estimations hold:*

i) If $\Delta > 0$ and $Q_\mu \subset \bar{B}(a, r_*/2)$, then

$$Q_\mu \subset \bar{B}\Big(a, \frac{1}{\sqrt{\Delta}}\arcsin\big(\frac{\alpha\sin(\sqrt{\Delta}\rho)}{\sqrt{2\alpha-1}}\big)\Big).$$

Moreover, any of the two conditions below implies $Q_\mu \subset \bar{B}(a, r_*/2)$:

$$a)\quad \frac{2\alpha\rho}{2\alpha-1} \le \frac{r_*}{2}; \qquad b)\quad \frac{2\alpha\rho}{2\alpha-1} > \frac{r_*}{2} \quad \text{and} \quad F_{\alpha,\rho,\Delta}(\frac{r_*}{2}-\rho) \le 0.$$

ii) If $\Delta = 0$, then

$$Q_\mu \subset \bar{B}\Big(a, \frac{\alpha\rho}{\sqrt{2\alpha-1}}\Big).$$

iii) If $\Delta < 0$, then

$$Q_\mu \subset \bar{B}\Big(a, \frac{1}{\sqrt{-\Delta}}\operatorname{arcsinh}\big(\frac{\alpha\sinh(\sqrt{-\Delta}\rho)}{\sqrt{2\alpha-1}}\big)\Big).$$

Finally, Lemma 3.18 ensures that any of the above three closed balls is contained in the open ball $B(a, r_*/2)$.

Proof. Firstly, we consider the case when $\Delta > 0$. Without loss of generality, we can assume that $\Delta = 1$. For every $x \in B_* \setminus \bar{B}(a,\rho)$, let $t_x = d(a,x) - \rho \in (0, \rho/(2\alpha-1)]$. By Propositions 3.6 and 3.12, if there exists some z on the minimal geodesic joining x and a such that $u_z = d(a,z) \in [0, \rho + t_x)$ verifies $\rho + t_x + u_z < r_*$ and $\min_{\bar{B}(\bar{a},\rho)} \bar{h}_{\bar{x},\bar{z}} > (1-\alpha)(\rho + t_x - u_z)/\alpha$, then $x \notin Q_\mu$. Or equivalently,

$$
\begin{aligned}
&Q_\mu \cap \bar{B}(a,\rho)^c \\
&\subset \Big\{ x \in B_* \setminus \bar{B}(a,\rho):\ t_x \in (0, \frac{\rho}{2\alpha-1}] \text{ has the property that for every } u_z \in [0, \rho + t_x) \\
&\qquad\qquad \text{such that } \rho + t_x + u_z < r_*,\ \min \bar{h}_{\bar{x},\bar{z}} \le \frac{1-\alpha}{\alpha}(\rho + t_x - u_z) \Big\} := A.
\end{aligned}
$$

Since the restrictive condition of the set A is only on t_x, for simplicity and without ambiguity, by dropping the subscripts of t_x and u_z we rewrite A in the following form:

$$
\begin{aligned}
&\Big\{ t \in (0, \frac{\rho}{2\alpha-1}]:\ \text{for every } u \in [0, \rho+t) \text{ such that } \rho + t + u < r_*, \\
&\qquad\qquad \min \bar{h}_{\bar{x},\bar{z}} \le \frac{1-\alpha}{\alpha}(\rho + t - u) \Big\} \\
&= \Big\{ t \in (0, \frac{r_*}{2} - \rho]:\ \text{for every } u \in [0, \rho+t) \text{ such that } \rho + t + u < r_*, \\
&\qquad\qquad \min \bar{h}_{\bar{x},\bar{z}} \le \frac{1-\alpha}{\alpha}(\rho + t - u) \Big\} \\
&\cup \Big\{ t \in (\frac{r_*}{2} - \rho, \frac{\rho}{2\alpha-1}]:\ \text{for every } u \in [0, \rho+t) \text{ such that } \rho + t + u < r_*, \\
&\qquad\qquad \min \bar{h}_{\bar{x},\bar{z}} \le \frac{1-\alpha}{\alpha}(\rho + t - u) \Big\} := B \cup C.
\end{aligned}
$$

Observe that for $t \in (0, r_*/2 - \rho]$ and $u \in [0, \rho+t)$, we always have $\rho + t + u < r_*$, hence by Proposition 3.13 and Lemma 3.16,

$$
\begin{aligned}
B &= \Big\{ t \in (0, \frac{r_*}{2} - \rho]:\ \text{for every } u \in [0, \rho+t),\ \min \bar{h}_{\bar{x},\bar{z}} \le \frac{1-\alpha}{\alpha}(\rho + t - u) \Big\} \\
&= \Big\{ t \in (0, \frac{r_*}{2} - \rho]:\ \text{for every } u \in [0, \rho+t) \text{ such that } \cot u \ge 2\cot\rho - \cot(\rho+t), \\
&\qquad\qquad t - \rho + u \le \frac{1-\alpha}{\alpha}(\rho + t - u) \Big\} \\
&\cap \Big\{ t \in (0, \frac{r_*}{2} - \rho]:\ \text{for every } u \in [0, \rho+t) \text{ such that } \cot u < 2\cot\rho - \cot(\rho+t), \\
&\qquad\qquad \cos(\rho+t-u) + \frac{\sin^2\rho \sin^2(\rho+t-u)}{2\sin u \sin(\rho+t)} \ge \cos\big(\frac{1-\alpha}{\alpha}(\rho+t-u)\big) \Big\}
\end{aligned}
$$

$$
\begin{aligned}
&= \Big\{ t \in (0, \frac{r_*}{2} - \rho] : \text{ for every } u \in [0, \rho + t) \text{ such that } \cot u \geq 2\cot\rho - \cot(\rho+t), \\
&\qquad\qquad u \leq \rho - (2\alpha - 1)t \Big\} \\
&\quad \cap \Big\{ t \in (0, \frac{r_*}{2} - \rho] : \text{ for every } u \in [0, \rho + t) \text{ such that } \cot u < 2\cot\rho - \cot(\rho+t), \\
&\qquad\qquad \sin(\rho + t) \leq \frac{\sin^2\rho}{4\sin u} \frac{\sin(\rho+t-u)}{\sin\frac{\rho+t-u}{2\alpha}} \frac{\sin(\rho+t-u)}{\sin\big(\frac{2\alpha-1}{2\alpha}(\rho+t-u)\big)} \Big\} \\
&= \Big\{ t \in (0, \frac{r_*}{2} - \rho] : \ \cot(\rho - (2\alpha-1)t) \leq 2\cot\rho - \cot(\rho+t) \Big\} \\
&\quad \cap \Big\{ t \in (0, \frac{r_*}{2} - \rho] : \text{ for every } u \in [0, \rho + t) \text{ such that } \cot u < 2\cot\rho - \cot(\rho+t), \\
&\qquad\qquad \sin(\rho + t) \leq \frac{\sin^2\rho}{4\sin u} \frac{\sin(\rho+t-u)}{\sin\frac{\rho+t-u}{2\alpha}} \frac{\sin(\rho+t-u)}{\sin\big(\frac{2\alpha-1}{2\alpha}(\rho+t-u)\big)} \Big\} \\
&\subset \Big\{ t \in (0, \frac{r_*}{2} - \rho] : \text{ for every } u \in (\rho - (2\alpha-1)t, \rho + t), \\
&\qquad\qquad \sin(\rho + t) \leq \frac{\sin^2\rho}{4\sin u} \frac{\sin(\rho+t-u)}{\sin\frac{\rho+t-u}{2\alpha}} \frac{\sin(\rho+t-u)}{\sin\big(\frac{2\alpha-1}{2\alpha}(\rho+t-u)\big)} \Big\} \\
&\subset \Big\{ t \in (0, \frac{r_*}{2} - \rho] : \ \sin(\rho+t) \leq \frac{\sin^2\rho}{4\sin(\rho+t)} \cdot 2\alpha \cdot \frac{2\alpha}{2\alpha - 1} \Big\} \\
&= \Big\{ t \in (0, \frac{r_*}{2} - \rho] : \ \rho + t \leq \arcsin\big(\frac{\alpha\sin\rho}{\sqrt{2\alpha-1}}\big) \Big\}. \qquad (2.20)
\end{aligned}
$$

Hence if $Q_\mu \subset \bar{B}(a, r_*/2)$, then $C = \phi$ and (2.20) says that

$$
Q_\mu \subset \bar{B}\Big(a, \arcsin\big(\frac{\alpha\sin\rho}{\sqrt{2\alpha-1}}\big)\Big),
$$

this completes the proof of the first assertion of i). To show the second one, observe that if a) holds, then Theorem 3.7 implies the desired result. Hence assume that b) holds. By Proposition 3.13 one has

$$
\begin{aligned}
C \subset \Big\{ t \in (\frac{r_*}{2} - \rho, \frac{\rho}{2\alpha-1}] : \ &\text{for every } u \in [0, r_* - (\rho + t)) \text{ such that} \\
&\cot u \geq 2\cot\rho - \cot(\rho+t), \\
&t - \rho + u \leq \frac{1-\alpha}{\alpha}(\rho + t - u) \Big\}
\end{aligned}
$$

$$=\left\{t \in (\frac{r_*}{2} - \rho, \frac{\rho}{2\alpha - 1}] :\ \cot(\rho - (2\alpha - 1)t) \le 2\cot\rho - \cot(\rho + t)\right\}.$$

Observe that for $t \in (r_*/2 - \rho,\ \rho/(2\alpha - 1))$ we have

$$\begin{aligned}
&\cot(\rho - (2\alpha - 1)t) \le 2\cot\rho - \cot(\rho + t) \\
\Longleftrightarrow\ &\cot(\rho - (2\alpha - 1)t) - \cot\rho \le \cot\rho - \cot(\rho + t) \\
\Longleftrightarrow\ &\frac{\sin(\rho - (2\alpha - 1)t)}{\sin((2\alpha - 1)t)} \ge \frac{\sin(\rho + t)}{\sin t} \\
\Longleftrightarrow\ &\cot((2\alpha - 1)t) - \cot\rho \ge \cot\rho + \cot t \\
\Longleftrightarrow\ &F_{\alpha,\rho,1}(t) \ge 0.
\end{aligned}$$

Hence

$$C \subset \left\{t \in (\frac{r_*}{2} - \rho, \frac{\rho}{2\alpha - 1}] :\ F_{\alpha,\rho,1}(t) \ge 0\right\} := D.$$

If $\alpha = 1$, clearly $D = \phi$. Now let $1/2 < \alpha < 1$, then Lemma 3.17 yields that

$$D = (\frac{r_*}{2} - \rho, \frac{\rho}{2\alpha - 1}] \cap (0, t_1] = \phi.$$

Thus $C = \phi$ still holds, that is, $Q_\mu \subset \bar{B}(a, r_*/2)$. The proof of $i)$ is complete.

Now let us turn to the proof of $ii)$ and $iii)$. In fact, the proof for these two cases are essentially the same as that of $i)$ except to note that we no longer need to assume that $Q_\mu \subset \bar{B}(a, r_*/2)$, because this is implied by Assumption 3.10. To see this, if $4\alpha\rho \le (2\alpha - 1)r_*$, then it suffices to use Theorem 3.7. So that let us assume $4\alpha\rho > (2\alpha - 1)r_*$ and show that $F_{\alpha,\rho,\Delta}(r_*/2 - \rho) \le 0$ for $\Delta \in \{-1, 0\}$. This is trivial if $\Delta = 0$ or $\alpha = 1$, hence let $\Delta = -1$ and $\alpha \in (1/2, 1)$. Since $r_*/2 - \rho > (1-\alpha)\rho/(2\alpha - 1)$ and $F_{\alpha,\rho,-1}$ is strictly decreasing, it suffices to show that $F_{\alpha,\rho,-1}((1-\alpha)\rho/(2\alpha-1)) \le 0$. To this end, define $f(\alpha) = F_{\alpha,\rho,-1}((1-\alpha)\rho/(2\alpha-1))$, easy computation gives that

$$f'(\alpha) = \frac{\rho}{\sinh^2((1-\alpha)\rho)} - \frac{\rho}{(2\alpha - 1)^2 \sinh^2(\frac{1-\alpha}{2\alpha-1}\rho)} > 0,$$

because the function $x \longmapsto \sinh x/x$ is strictly increasing. Hence $f(\alpha) < f(1-) = 2(\rho^{-1} - \coth\rho) < 0$. The proof is complete. □

Remark 3.20. When $\Delta > 0$, Assumption 3.10 does not imply the condition $b)$ in $i)$. In fact, in the case when $M = \mathbb{S}^2$, we have $r_* = \pi$ and $\Delta = 1$. Then let $\alpha = 0.51$ and $\rho = 0.99\pi(1 - (2\alpha)^{-1})$, then $2\alpha\rho/(2\alpha - 1) \in (\pi/2 + 1.5393, \pi - 0.0314)$, but $F_{\alpha,\rho,1}(\pi/2 - \rho) \approx 0.2907 > 0$.

Remark 3.21. It is easily seen that if we replace r_* by any $r \in (0, r_*]$ in Assumption 3.10, then Lemma 3.18 still holds when r_* is replaced by r. This observation can be used to reinforce the conclusions of Theorem 3.19. For example, in the case when $\Delta > 0$,

$$\frac{2\alpha\rho}{2\alpha-1} \le \frac{r_*}{2} \text{ implies that } Q_\mu \subset \bar{B}\Big(a, \frac{1}{\sqrt{\Delta}}\arcsin\big(\frac{\alpha\sin(\sqrt{\Delta}\rho)}{\sqrt{2\alpha-1}}\big)\Big) \subset B\Big(a, \frac{r_*}{4}\Big).$$

Remark 3.22. Although we have chosen the framework of this section to be a Riemannian manifold, the essential tool that has been used is the hinge version of the triangle comparison theorem. Consequently, all the results in this section remain true if M is a CAT(Δ) space (see [26, Chapter 2]) and r_* is replaced by $\pi/\sqrt{\Delta}$ in Assumption 3.10.

Remark 3.23. For the case when $\alpha = 1$, Assumption 3.10 becomes

$$\rho < \frac{1}{2}\min\{\frac{\pi}{\sqrt{\Delta}}, \mathrm{inj}\}.$$

Observe that in this case, when $\Delta > 0$, the condition $F_{1,\rho,\Delta}(r_*/2-\rho) \le 0$ is trivially true in case of need. Hence Theorem 3.19 yields that $Q_\mu \subset \bar{B}(a,\rho)$, which is exactly what the Theorem 2.1 in [1] says for medians.

3.3 Uniqueness of Fréchet sample medians in compact Riemannian manifolds

In this section, we shall always assume that M is a complete Riemannian manifold of dimension $l \ge 2$. The Riemannian metric and the Riemannian distance are denoted by $\langle\cdot,\cdot\rangle$ and d, respectively. For each point $x \in M$, S_x denotes the unit sphere in T_xM. Moreover, for a tangent vector $v \in S_x$, the distance between x and its cut point along the geodesic starting from x with velocity v is denoted by $\tau(v)$. Certainly, if there is no cut point along this geodesic, then we define $\tau(v) = +\infty$.

For every point $(x_1,\dots,x_N) \in M^N$, where $N \ge 3$ is a fixed natural number, we write

$$\mu(x_1,\dots,x_N) = \frac{1}{N}\sum_{k=1}^{N}\delta_{x_k}.$$

The set of all the Fréchet medians of $\mu(x_1,\dots,x_N)$ is denoted by $Q(x_1,\dots,x_N)$.

We begin with the basic observation that if one data point is moved towards a median along some minimizing geodesic for a little distance, then the median remains unchanged.

Proposition 3.24. *Let $(x_1,\dots,x_N) \in M^N$ and $m \in Q(x_1,\dots,x_N)$. Fix a normal geodesic $\gamma : [0,+\infty) \to M$ such that $\gamma(0) = x_1$, $\gamma(d(x_1,m)) = m$. Then for every $t \in [0, d(x_1,m)]$ we have*

$$Q(\gamma(t),x_2,\dots,x_N) = \begin{cases} Q(x_1,\dots,x_N)\cap\gamma[t,\tau(\dot\gamma(0))], & \text{if } \tau(\dot\gamma(0)) < +\infty; \\ Q(x_1,\dots,x_N)\cap\gamma_1[t,+\infty), & \text{if } \tau(\dot\gamma(0)) = +\infty. \end{cases}$$

Particularly, $m \in Q(\gamma(t), x_2, \ldots, x_N)$.

Proof. For simplicity, let $\mu = \mu(x_1, \ldots, x_N)$ and $\mu_t = \mu(\gamma(t), x_2, \ldots, x_N)$. Then for every $x \in M$,

$$\begin{aligned} f_{\mu_t}(x) - f_{\mu_t}(m) &= \Big(f_\mu(x) - \frac{1}{N}d(x, x_1) + \frac{1}{N}d(x, \gamma(t))\Big) \\ &\quad - \Big(f_\mu(m) - \frac{1}{N}d(m, x_1) + \frac{1}{N}d(m, \gamma(t))\Big) \\ &= \Big(f_\mu(x) - f_\mu(m)\Big) + \Big(d(x, \gamma(t)) + t - d(x, x_1)\Big) \geq 0. \end{aligned}$$

So that $m \in Q_{\mu_t}$. Combine this with the fact that m is a median of μ, it is easily seen from the above proof that

$$Q_{\mu_t} = Q_\mu \cap \{x \in M : d(x, \gamma(t)) + t = d(x, x_1)\}.$$

Now the conclusion follows from the definition of $\tau(\dot\gamma(0))$. □

The following theorem states that in order to get the uniqueness of Fréchet medians, it suffices to move two data points towards a common median along some minimizing geodesics for a little distance.

Theorem 3.25. *Let* $(x_1, \ldots, x_N) \in M^N$ *and* $m \in Q(x_1, \ldots, x_N)$. *Fix two normal geodesics* $\gamma_1, \gamma_2 : [0, +\infty) \to M$ *such that* $\gamma_1(0) = x_1$, $\gamma_1(d(x_1, m)) = m$, $\gamma_2(0) = x_2$ *and* $\gamma_2(d(x_2, m)) = m$. *Assume that*

$$x_2 \notin \begin{cases} \gamma_1[0, \tau(\dot\gamma_1(0))], & \text{if } \tau(\dot\gamma_1(0)) < +\infty; \\ \gamma_1[0, +\infty), & \text{if } \tau(\dot\gamma_1(0)) = +\infty. \end{cases}$$

Then for every $t \in (0, d(x_1, m)]$ *and* $s \in (0, d(x_2, m)]$ *we have*

$$Q(\gamma_1(t), \gamma_2(s), x_3, \ldots, x_N) = \{m\}.$$

Proof. Without loss of generality, we may assume that both $\tau(\dot\gamma_1(0))$ and $\tau(\dot\gamma_2(0))$ are finite. Applying Proposition 3.24 two times we get

$$Q(\gamma_1(t), \gamma_2(s), x_3, \ldots, x_N) \subset Q(x_1, \ldots, x_N) \cap \gamma_1[t, \tau(\dot\gamma_1(0))] \cap \gamma_2[s, \tau(\dot\gamma_2(0))].$$

Since $x_2 \notin \gamma_1[0, \tau(\dot\gamma_1(0))]$, the definition of cut point yields $\gamma_1[t, \tau(\dot\gamma_1(0))] \cap \gamma_2[s, \tau(\dot\gamma_2(0))] = \{m\}$. The proof is complete. □

We need the following necessary conditions of Fréchet medians.

Proposition 3.26. *Let* $(x_1, \ldots, x_N) \in M^N$ *and* $m \in Q(x_1, \ldots, x_N)$. *For every* $k = 1, \ldots, N$ *let* $\gamma_k : [0, d(m, x_k)] \to M$ *be a normal geodesic such that* $\gamma_k(0) = m$ *and* $\gamma_k(d(m, x_k)) = x_k$.
i) If m *does not coincide with any* x_k*, then*

$$\sum_{k=1}^{N} \dot{\gamma}_k(0) = 0. \tag{3.21}$$

In this case, the minimizing geodesics $\gamma_1, \ldots, \gamma_N$ *are uniquely determined.*
ii) If m *coincides with some* x_{k_0}*, then*

$$\Bigg| \sum_{x_k \neq x_{k_0}} \dot{\gamma}_k(0) \Bigg| \leq \sum_{x_k = x_{k_0}} 1. \tag{3.22}$$

Proof. For sufficiently small $\varepsilon > 0$, Proposition 3.24 yields that m is a median of $\mu(\gamma_1(\varepsilon), \ldots, \gamma_N(\varepsilon))$. Hence [91, Theorem 2.2] gives (3.21) and (3.22). Now assume that m does not coincide with any x_k and, without loss of generality, there is another normal geodesic $\zeta_1 : [0, d(m, x_1)] \to M$ such that $\zeta_1(0) = m$ and $\zeta_1(d(m, x_1)) = x_1$. Then (3.21) yields $\dot{\zeta}_1(0) + \sum_{k=2}^{N} \dot{\gamma}_k(0) = 0$. So that $\dot{\zeta}_1(0) = \dot{\gamma}_1(0)$, that is to say, $\zeta_1 = \gamma_1$. The proof is complete. □

From now on, we will only consider the case when M is a compact Riemannian manifold. As a result, let the following assumption hold in the rest part of this section:

Assumption 3.27. M is a compact Riemannian manifold with diameter L.

In what follows, all the measure-theoretic statements should be understood to be with respect to the canonical Lebesgue measure of the underlying manifold. Let λ_M denote the canonical Lebesgue measure of M.

The following lemma gives a simple observation on dimension.

Lemma 3.28. *The manifold*

$$V = \Big\{ (x, n_1, \ldots, n_N) : x \in M,\ (n_k)_{1 \leq k \leq N} \subset S_x,\ \sum_{k=1}^{N} n_k = 0 \textit{ and there exist } k_1 \neq k_2 \textit{ such that } n_{k_1} \textit{ and } n_{k_2} \textit{ are linearly independent} \Big\}$$

is of dimension $N(l-1)$.

Proof. Consider the manifold

$$W = \Big\{ (x, n_1, \ldots, n_N) : x \in M,\ (n_k)_{1 \leq k \leq N} \subset S_x \text{ and there exist } k_1 \neq k_2 \text{ such that } n_{k_1} \text{ and } n_{k_2} \text{ are linearly independent} \Big\},$$

then the smooth map

$$f: \quad W \longrightarrow TM, \quad (x, n_1, \ldots, n_N) \longmapsto \sum_{k=1}^{N} n_k$$

has full rank l everywhere on W. Now the constant rank level set theorem (see [64, Theorem 8.8]) yields the desired result. $\square$

Our method of studying the uniqueness of Fréchet medians is based on the regularity properties of the following function:

$$\begin{aligned} \varphi: \quad & V \times \mathbf{R}^N \longrightarrow M^N, \\ & (x, n_1, \ldots, n_N, r_1, \ldots, r_N) \longmapsto (\exp_x(r_1 n_1), \ldots, \exp_x(r_N n_N)). \end{aligned}$$

For a closed interval $[a, b] \subset \mathbf{R}$, the restriction of φ to $V \times [a, b]^N$ will be denoted by $\varphi_{a,b}$. The canonical projection of $V \times \mathbf{R}^N$ onto M and $\mathbf{R}^N$ will be denoted by σ and ζ, respectively.

Generally speaking, the non uniqueness of Fréchet medians is due to some symmetric properties of data points. As a result, generic data points should have a unique Fréchet median. In mathematical language, this means that the set of all the particular positions of data points is of measure zero. Now our aim is to find all these particular cases. Firstly, in view of the uniqueness result of Riemannian medians (see [91, Theorem 3.1]), the first null set that should be eliminated is

$$C_1 = \Big\{(x_1, \ldots, x_N) \in M^N : x_1, \ldots, x_N \text{ are contained in a single geodesic}\Big\}.$$

Observe that C_1 is a closed subset of M^N. The second null set coming into our sight is the following one:

$$C_2 = \Big\{(x_1, \ldots, x_N) \in M^N : \ (x_1, \ldots, x_N) \text{ is a critical value of } \varphi_{0,L}\Big\}.$$

Since φ is smooth, Sard's theorem implies that C_2 is of measure zero. Moreover, it is easily seen that $(M^N \setminus C_1) \cap C_2$ is closed in $M^N \setminus C_1$.

The following proposition says that apart form $C_1 \cup C_2$ one can only have a finite number of Fréchet medians.

Proposition 3.29. *$Q(x_1, \ldots, x_N)$ is a finite set for every $(x_1, \ldots, x_N) \in M^N \setminus (C_1 \cup C_2)$.*

Proof. Let $(x_1, \ldots, x_N) \in M^N \setminus (C_1 \cup C_2)$ and $A \subset Q(x_1, \ldots, x_N)$ be the set of medians that do not coincide with any x_k. If $A = \phi$, then there is nothing to prove. Now assume that $A \neq \phi$, then Proposition 3.26 implies that $A \subset \sigma \circ \varphi_{0,L}^{-1}(x_1, \ldots, x_N)$. Moreover, Lemma 3.28 and the constant rank level set theorem imply that $\varphi_{0,L}^{-1}(x_1, \ldots, x_N)$ is a zero dimensional regular submanifold of $V \times [0, L]^N$, that is, some isolated points. Since $(x_1, \ldots, x_N) \notin C_1$, $\varphi_{0,L}^{-1}(x_1, \ldots, x_N)$ is also compact, hence it is a finite set. So that A is also finite, as desired. $\square$

The following two lemmas enable us to avoid the problem of cut locus.

Lemma 3.30. *Let U be a bounded open subset of $V\times\mathbf{R}^N$ such that $\varphi:\ U\longmapsto\varphi(U)$ is a diffeomorphism, then*

$$\lambda_M^{\otimes N}\left\{(x_1,\dots,x_N)\in\varphi(U):\sigma\circ\varphi^{-1}(x_1,\dots,x_N)\in\bigcup_{k=1}^N \mathrm{Cut}(x_k)\right\}=0.$$

Proof. Without loss of generality, we will show that $\lambda_M^{\otimes N}\{(x_1,\dots,x_N)\in M^N:\sigma\circ\varphi^{-1}(x_1,\dots,x_N)\in\mathrm{Cut}(x_N)\}=0$. In fact, letting $(x_1,\dots,x_N)=\varphi(x,n_1,\dots n_N,r_1,\dots,r_N)$, $\mathbf{n}=(n_1,\dots n_N)$ and $\det(D\varphi)\le c$ on U for some $c>0$, then the change of variable formula and Fubini's theorem yield that

$$\begin{aligned}
&\lambda_M^{\otimes N}\{(x_1,\dots,x_N)\in\varphi(U):\sigma\circ\varphi^{-1}(x_1,\dots,x_N)\in\mathrm{Cut}(x_N)\}\\
=&\lambda_M^{\otimes N}\{(x_1,\dots,x_N)\in\varphi(U):x_N\in\mathrm{Cut}(\sigma\circ\varphi^{-1}(x_1,\dots,x_N))\}\\
=&\int_{\varphi(U)}\mathbf{1}_{\{x_N\in\mathrm{Cut}(\sigma\circ\varphi^{-1}(x_1,\dots,x_N))\}}dx_1\dots dx_N\\
=&\int_U\mathbf{1}_{\{\exp_x(r_Nn_N)\in\mathrm{Cut}(x)\}}\det(D\varphi)dx\,d\mathbf{n}\,dr_1\dots dr_N\\
\le& c\int_{V\times\mathbf{R}^N}\mathbf{1}_{\{\exp_x(r_Nn_N)\in\mathrm{Cut}(x)\}}dx\,d\mathbf{n}\,dr_1\dots dr_N\\
=& c\int_{V\times\mathbf{R}^{N-1}}dx\,d\mathbf{n}\,dr_1\dots dr_{N-1}\int_{\mathbf{R}}\mathbf{1}_{\{\exp_x(r_Nn_N)\in\mathrm{Cut}(x)\}}dr_N\\
=&0.
\end{aligned}$$

The proof is complete. $\square$

In order to tackle the cut locus, it is easily seen that the following null set should be eliminated:

$$C_3=\left\{(x_1,\dots,x_N)\in M^N:\ x_i\in\{x_j\}\cup\mathrm{Cut}(x_j),\ \text{for some } i\neq j\right\}.$$

Observe that C_3 is also closed because the set $\{(x,y)\in M^2:\ x\in\mathrm{Cut}(y)\}$ is closed.

Lemma 3.31. *For every $(x_1,\dots,x_N)\in M^N\setminus(C_1\cup C_2\cup C_3)$, there exists $\delta>0$ such that*

$$\lambda_M^{\otimes N}\left\{(y_1,\dots,y_N)\in B(x_1,\delta)\times\cdots\times B(x_N,\delta):\ Q(y_1,\dots,y_N)\cap\bigcup_{k=1}^N\mathrm{Cut}(y_k)\neq\phi\right\}=0.$$

Proof. If $Q(x_1,\dots,x_N)\subset\{x_1,\dots,x_N\}$, then the assertion is trivial by Theorem 3.3. Now assume that $Q(x_1,\dots,x_N)\setminus\{x_1,\dots,x_N\}\neq\phi$, then the proof of Proposition 3.29 yields that $\varphi_{0,L}^{-1}(x_1,\dots,x_N)$ is finite. Hence we can choose $\varepsilon,\eta>0$

and O a relatively compact open subset of V such that $\varepsilon < \min\{\operatorname{inj} M/2, \min\{r_k : (r_1,\dots,r_N) \in \zeta \circ \varphi_{0,L}^{-1}(x_1,\dots,x_N), k=1,\dots,N\}\}$, $B(x_i, 2\varepsilon) \cap \operatorname{Cut}(B(x_j, 2\varepsilon)) = \phi$ and $\varphi_{\varepsilon,L+\eta}^{-1}(x_1,\dots,x_N) \subset O \times (\varepsilon, L+\eta)^N$. Then by the stack of records theorem (see [50, Exercise 7, Chapter 1, Section 4]), there exists $\delta \in (0,\varepsilon)$ such that $U = B(x_1,\delta) \times \cdots \times B(x_N,\delta)$ verifies $\varphi_{\varepsilon,L+\eta}^{-1}(U) = V_1 \cup \cdots \cup V_h$, $V_i \cap V_j = \phi$ for $i \neq j$ and $\varphi_{\varepsilon,L+\eta} : V_i \to U$ is a diffeomorphism for every i. Lemma 3.30 yields that there exists a null set $A \subset U$, such that for every $(y_1,\dots,y_N) \in U \setminus A$ and for every $(y, n_1,\dots,n_N, r_1,\dots,r_N) \in \varphi_{\varepsilon,L+\eta}^{-1}(y_1,\dots,y_N)$ one always has $y \notin \bigcup_{k=1}^N \operatorname{Cut}(y_k)$. Particularly, for $m \in Q(y_1,\dots,y_N)$ such that $d(m,y_k) \geq \varepsilon$ for every k, we have $m \notin \bigcup_{k=1}^N \operatorname{Cut}(y_k)$. Now let $m \in Q(y_1,\dots,y_N)$ such that $d(m,y_{k_0}) < \varepsilon$ for some k_0, then $d(m,x_{k_0}) \leq d(m,y_{k_0}) + d(y_{k_0},x_{k_0}) < 2\varepsilon$. So that $m \notin \bigcup_{k=1}^N \operatorname{Cut}(y_k)$ since $y_k \in B(x_k, 2\varepsilon)$. This completes the proof. □

Now the cut locus can be eliminated without difficulty.

Proposition 3.32. *The set*

$$C_4 = \left\{(x_1,\dots,x_N) \in M^N : \ Q(x_1,\dots,x_N) \cap \bigcup_{k=1}^N \operatorname{Cut}(x_k) \neq \phi\right\}$$

is of measure zero and is closed.

Proof. It suffices to show that C_4 is of measure zero. This is a direct consequence of Lemma 3.31 and the fact that $M \setminus (C_1 \cup C_2 \cup C_3)$ is second countable. □

Let $x, y \in M$ such that $y \notin \{x\} \cup \operatorname{Cut}(x)$, we denote $\gamma_{xy} : [0, d(x,y)] \to M$ the unique minimizing geodesic such that $\gamma_{xy}(0) = x$ and $\gamma_{xy}(d(x,y)) = y$. For every $u \in T_xM$ and $v \in T_yM$, let $J(v,u)(\cdot)$ be the unique Jacobi field along γ_{xy} with boundary condition $J(u,v)(0) = u$ and $J(u,v)(d(x,y)) = v$.

Lemma 3.33. *Let $x, y \in M$ such that $y \notin \{x\} \cup \operatorname{Cut}(x)$. Then for every $v \in T_yM$, we have*

$$\nabla_v \frac{\exp_x^{-1}(\cdot)}{d(x,\cdot)} = \dot{J}(0_x, v^{\mathrm{nor}})(0),$$

where v^{nor} is the normal component of v with respect to $\dot{\gamma}_{xy}(d(x,y))$.

Proof. By [9, p. 1517],

$$\begin{aligned}
\nabla_v \frac{\exp_x^{-1}(\cdot)}{d(x,\cdot)} =& \frac{\nabla_v \exp_x^{-1}(\cdot)}{d(x,y)} - \frac{\exp_x^{-1} y \nabla_v d(x,\cdot)}{d(x,y)^2} \\
=& \frac{\nabla_v \exp_x^{-1}(\cdot)}{d(x,y)} - \left\langle v, \frac{-\exp_y^{-1} x}{d(x,y)} \right\rangle \frac{\exp_x^{-1} y}{d(x,y)^2} \\
=& \dot{J}(0_x, v)(0) - \dot{J}(0_x, v^{\tan})(0) \\
=& \dot{J}(0_x, v^{\mathrm{nor}})(0),
\end{aligned}$$

where $v^{\tan}$ is the tangent component of u with respect to $\dot{\gamma}_{xy}(d(x,y))$. □

With the this differential formula, another particular case can be eliminated now.

Proposition 3.34. *The set*

$$C_5 = \left\{ (x_1, \ldots x_N) \in M^N \setminus (C_1 \cup C_3) : \left| \sum_{k \neq k_0} \frac{\exp_{x_{k_0}}^{-1} x_k}{d(x_{k_0}, x_k)} \right| = 1 \text{ for some } k_0 \right\}$$

is of measure zero and is closed in $M^N \setminus (C_1 \cup C_3)$.

Proof. Without loss of generality, let us show that

$$C_5' = \left\{ (x_1, \ldots, x_N) \in M^N \setminus (C_1 \cup C_3) : h(x_1, \ldots, x_N) = 1 \right\}$$

is of measure zero, where

$$h(x_1, \ldots, x_N) = \left| \sum_{k=1}^{N-1} \frac{\exp_{x_N}^{-1} x_k}{d(x_N, x_k)} \right|^2 .$$

By the constant rank level set theorem, it suffices to show that $\operatorname{grad} h$ is nowhere vanishing on $M^N \setminus (C_1 \cup C_3)$. To this end, let $(x_1, \ldots, x_N) \in M^N \setminus (C_1 \cup C_3)$ and $u = \sum_{k=1}^{N-1} \exp_{x_N}^{-1} x_k / d(x_N, x_k)$. Since $N \geq 3$, without loss of generality, we can assume that u and $\exp_{x_N}^{-1} x_1$ are not parallel. Then for each $v \in T_{x_1} M$, by lemma 3.33 we have

$$\nabla_v h(\cdot, x_2, \ldots, x_N) = \left\langle \nabla_v \frac{\exp_{x_N}^{-1} x_1}{d(x_N, x_1)}, u \right\rangle = 2 \langle \dot{J}(0_{x_N}, v^{\mathrm{nor}})(0), u \rangle = 2 \langle \psi(v), u \rangle,$$

where the linear map ψ is defined by

$$\psi : \quad T_{x_1} M \longrightarrow T_{x_N} M, \quad v \longmapsto \dot{J}(0_{x_N}, v^{\mathrm{nor}})(0),$$

v^{nor} is the normal component of v with respect to $\exp_{x_1}^{-1} x_N$. Hence we have $\operatorname{grad}_{x_1} h(\cdot, x_2, \ldots, x_N) = \psi^*(u)$, where ψ^* is the adjoint of ψ. Since the range space of ψ is the orthogonal complement of $\exp_{x_N}^{-1} x_1$, one has necessarily $\psi^*(u) \neq 0$, this completes the proof. □

The reason why the set C_5 should be eliminated is given by the following simple lemma.

Lemma 3.35. *Let* $(x_1, \ldots, x_N), (x_1^i, \ldots, x_N^i) \in M^N \setminus C_3$ *for every* $i \in \mathbf{N}$ *and* $(x_1^i, \ldots, x_N^i) \longrightarrow (x_1, \ldots, x_N)$, *when* $i \longrightarrow \infty$. *Assume that* $m_i \in Q(x_1^i, \ldots, x_N^i) \setminus \{x_1^i, \ldots, x_N^i\}$ *and* $m_i \longrightarrow x_{k_0}$, *then*

$$\left| \sum_{k \neq k_0} \frac{\exp_{x_{k_0}}^{-1} x_k}{d(x_{k_0}, x_k)} \right| = 1.$$

Proof. It suffices to note that for i sufficiently large, Proposition 3.26 gives

$$\left| \sum_{k\neq k_0} \frac{\exp_{m_i}^{-1} x_k^i}{d(m_i, x_k^i)} \right| = \left| \frac{\exp_{m_i}^{-1} x_{k_0}^i}{d(m_i, x_{k_0}^i)} \right| = 1.$$

Then letting $i \to \infty$ gives the result. $\square$

As a corollary to Proposition 3.34, the following proposition tells us that for generic data points, there cannot exist two data points which are both Fréchet medians.

Proposition 3.36. *The set*

$$C_6 = \Big\{ (x_1, \ldots, x_N) \in M^N \setminus (C_1 \cup C_3 \cup C_5) : \\ \textit{there exist } i \neq j \textit{ such that } f_\mu(x_i) = f_\mu(x_j), \textit{ where } \mu = \mu(x_1 \ldots, x_N) \Big\}$$

is of measure zero and is closed in $M^N \setminus (C_1 \cup C_3 \cup C_5)$.

Proof. For every $(x_1, \ldots, x_N) \in M^N \setminus (C_1 \cup C_3 \cup C_5)$, let $f(x_1, \ldots, x_N) = f_\mu(x_{N-1})$ and $g(x_1, \ldots, x_N) = f_\mu(x_N)$, where $\mu = \mu(x_1 \ldots, x_N)$. Without loss of generality, we will show that $\{(x_1, \ldots, x_N) \in M^N \setminus (C_1 \cup C_3 \cup C_5) : f(x_1, \ldots, x_N) = g(x_1, \ldots, x_N)\}$ is of measure zero. Always by the constant rank level set theorem, it suffices to show that grad f and grad g are nowhere identical on $M^N \setminus (C_1 \cup C_3 \cup C_5)$. In fact,

$$\begin{aligned} \operatorname{grad}_{x_N} f(x_1, \ldots, x_N) &= \frac{-\exp_{x_N}^{-1} x_{N-1}}{d(x_N, x_{N-1})} \\ &\neq \sum_{k=1}^{N-1} \frac{-\exp_{x_N}^{-1} x_k}{d(x_N, x_k)} = \operatorname{grad}_{x_N} g(x_1, \ldots, x_N), \end{aligned}$$

because C_5 is eliminated, as desired. $\square$

As needed in the following proofs, the restriction of φ on the set

$$E = \Big\{ (x, n_1, \ldots, n_N, r_1, \ldots, r_N) \in V \times \mathbf{R}^N : 0 < r_k < \tau(n_k) \text{ for every } k \Big\}$$

is denoted by let $\hat{\varphi}$. Clearly, $\hat{\varphi}$ is smooth.

The lemma below is a final preparation for the main result of this section.

Lemma 3.37. *Let U be an open subset of $M^N \setminus \bigcup_{k=1}^6 C_k$. Assume that $U_1 \cup U_2 \subset \hat{\varphi}^{-1}(U)$ such that for $i = 1, 2$, $\hat{\varphi}_i = \hat{\varphi}|_{U_i} : U_i \to U$ is a diffeomorphism and $\sigma(U_1) \cap \sigma(U_2) = \phi$. For simplicity, when $(x_1, \ldots, x_N) \in U$, we write $x =$*

$\sigma \circ \hat{\varphi}_1^{-1}(x_1, \ldots, x_N)$, $y = \sigma \circ \hat{\varphi}_2^{-1}(x_1, \ldots, x_N)$ and $\mu = \mu(x_1, \ldots, x_N)$. Then the following two sets are of measure zero:

$$\Big\{(x_1, \ldots, x_N) \in U : f_\mu(x) = f_\mu(y)\Big\} \text{ and}$$
$$\Big\{(x_1, \ldots, x_N) \in U : \text{ there exists } k_0 \text{ such that } f_\mu(x) = f_\mu(x_{k_0})\Big\}.$$

Proof. We only show the first set is null, since the proof for the second one is similar. Let $f_1(x_1, \ldots, x_N) = f_\mu(x)$, $f_2(x_1, \ldots, x_N) = f_\mu(y)$ and $w_k \in T_{x_k}M$. Then the first variational formula of arc length (see [33, p. 5]) yields that

$$\begin{aligned}
&\frac{d}{dt}\Big|_{t=0} f_1(\exp_{x(t)}(tw_1), \ldots \exp_{x(t)}(tw_N)) \\
&= \sum_{k=1}^{N} \left(\left\langle \frac{-\exp_{x_k}^{-1} x}{d(x_k, x)}, w_k \right\rangle - \left\langle \dot{x}(0), \frac{\exp_x^{-1} x_k}{d(x, x_k)} \right\rangle \right) \\
&= \sum_{k=1}^{N} \left\langle \frac{-\exp_{x_k}^{-1} x}{d(x_k, x)}, w_k \right\rangle - \left\langle \dot{x}(0), \sum_{k=1}^{N} \frac{\exp_x^{-1} x_k}{d(x, x_k)} \right\rangle \\
&= \sum_{k=1}^{N} \left\langle \frac{-\exp_{x_k}^{-1} x}{d(x_k, x)}, w_k \right\rangle.
\end{aligned}$$

Hence

$$\operatorname{grad} f_1(x_1, \ldots, x_N) = \left(\frac{-\exp_{x_1}^{-1} x}{d(x_1, x)}, \ldots, \frac{-\exp_{x_N}^{-1} x}{d(x_N, x)} \right).$$

Observe that $(x_1, \ldots, x_N) \notin C_1$, $N \geq 3$ and $x \neq y$, we have $\operatorname{grad} f_1 \neq \operatorname{grad} f_2$ on U. Then the constant rank level set theorem yields that $\{f_1 = f_2\}$ is a regular submanifold of U of codimension 1, hence it is of measure zero. The proof is complete. □

The following theorem is the main result of this section.

Theorem 3.38. *$\mu(x_1, \ldots, x_N)$ has a unique Fréchet median for almost every $(x_1, \ldots, x_N) \in M^N$.*

Proof. Since $M^N \setminus \bigcup_{k=1}^6 C_k$ is second countable, it suffices to show that for every $(x_1, \ldots, x_N) \in M^N \setminus \bigcup_{k=1}^6 C_k$, there exists $\delta > 0$ such that $\mu(y_1, \ldots, y_N)$ has a unique Fréchet median for almost every $(y_1, \ldots, y_N) \in B(x_1, \delta) \times \cdots \times B(x_N, \delta)$. In fact, let $(x_1, \ldots, x_N) \in M^N \setminus \bigcup_{k=1}^6 C_k$, without loss of generality, we can assume that $Q(x_1, \ldots, x_N) = \{y, z, x_N\}$, where $y, z \notin \{x_1, \ldots, x_N\}$. Assume that $Y = (y, n_1, \ldots, n_N, r_1, \ldots, r_N)$ and $Z = (z, v_1, \ldots, v_N, t_1, \ldots, t_N) \in \hat{\varphi}^{-1}(x_1, \ldots, x_N)$. Since $(x_1, \ldots, x_N)$ is a regular value of $\hat{\varphi}$, we can choose a $\delta > 0$ such that there exist neighborhoods U_1 of Y and U_2 of Z such that for $i = 1, 2$, $\hat{\varphi}_i = \hat{\varphi}|_{U_i} : U_i \to$

U is diffeomorphism, $\sigma(U_1) \cap \sigma(U_2) = \phi$ and $B(x_N, \delta) \cap (\sigma(U_1) \cup \sigma(U_2)) = \phi$, where $U = B(x_1, \delta) \times \cdots \times B(x_N, \delta)$. Furthermore, by Theorem 3.3 and Lemma 3.35, we can also assume that for every $(y_1, \ldots, y_N) \in B(x_1, \delta) \times \cdots \times B(x_N, \delta)$, $Q(y_1, \ldots, y_N) \subset B(x_N, \delta) \cup \sigma(U_1) \cup \sigma(U_2)$ and $Q(y_1, \ldots, y_N) \cap B(x_N, \delta) \subset \{y_N\}$. Now it suffices to use Lemma 3.37 to complete the proof. □

Remark 3.39. In probability language, Theorem 3.38 is equivalent to say that if $(X_1, \ldots, X_N)$ is an M^N-valued random variable with density, then $\mu(X_1, \ldots, X_N)$ has a unique Fréchet median almost surely. Clearly, the same statement is also true if $X_1, \ldots, X_N$ are independent and M-valued random variables with desity.

3.4 Appendix

In this appendix, under the framework of section 3.2, we give another estimation for the position of Q_μ using a different method.

Lemma 3.40. *Let $r < r_*/2$ and $\triangle ABC$ be a geodesic triangle in $\bar{B}(a, r)$ such that $A = a$, $B \in \bar{B}(a, \rho)$ and $C \in \bar{B}(a, r) \setminus \bar{B}(a, \rho)$. Then*

$$\cos \angle C \geq \frac{\sin(\sqrt{\Delta}(d(C, A) - \rho))}{\sin(\sqrt{\Delta}(d(C, A) + \rho))}.$$

Proof. The proof of this lemma will be the same as that of Lemma 2.11 as long as the existence of a comparison triangle of $\triangle ABC$ in model space M^2_Δ is verified. To this end, $d(A, B) + d(B, C) + d(C, A) \leq \rho + (\rho + r) + r < 4r < 2\pi/\sqrt{\Delta}$. The proof is complete. □

Now we give an estimation of Q_μ.

Proposition 3.41. *Assume that*

$$\frac{2\alpha\rho}{2\alpha - 1} < \frac{r_*}{2},$$

then

$$Q_\mu \subset \bar{B}\Big(a, \frac{\rho}{2\alpha - 1}\Big).$$

Proof. Let $\eta = (2\alpha - 1)^{-1}$, then $\bar{B}(a, \rho) \subset \bar{B}(a, \eta\rho) \subset B(a, r)$ since $\alpha > 1/2$. Fix a point $x \in \bar{B}(a, r) \setminus \bar{B}(a, \eta\rho)$ and let $\gamma : [0, b] \to \bar{B}(a, r)$ the minimal geodesic parametrized by arc length such that $\gamma(0) = a$ et $\gamma(b) = x$. Then there exists $t_0 \in (0, b)$ such that $y = \gamma(t_0) \in \partial\bar{B}(a, \eta\rho)$. For every $p \in \bar{B}(a, \rho)$, by intermediate value theorem, there exists $\xi \in (t_0, b)$ such that

$$\begin{aligned} d(x,p) - d(y,p) &= \frac{d}{dt} d(\gamma(t),p)\Big|_{t=\xi} (b - t_0) = \langle \dot\gamma(\xi), \frac{-\exp^{-1}_{\gamma(\xi)} p}{d(\gamma(\xi),p)} \rangle (b - t_0) \\ &= \langle \frac{-\exp^{-1}_{\gamma(\xi)} y}{d(\gamma(\xi),y)}, \frac{-\exp^{-1}_{\gamma(\xi)} p}{d(\gamma(\xi),p)} \rangle d(x,y) \\ &= d(x,y) \cos \angle y\gamma(\xi)p = d(x,y) \cos \angle a\gamma(\xi)p. \end{aligned}$$

By Lemma 3.40,

$$\cos \angle a\gamma(\xi)p \geq \frac{\sin(\sqrt{\Delta}(d(\gamma(\xi),a) - \rho))}{\sin(\sqrt{\Delta}(d(\gamma(\xi),a) + \rho))} > \frac{\sin(\sqrt{\Delta}(\eta\rho - \rho))}{\sin(\sqrt{\Delta}(\eta\rho + \rho))} \geq \frac{\eta - 1}{\eta + 1} = \frac{1-\alpha}{\alpha},$$

since for $u_0 \in (0, \pi/2)$, the function $u \mapsto \sin(u - u_0)/\sin(u + u_0)$ is strictly increasing on $(0, \pi/2)$. Moreover, the inequality $\sin u/\sin v \geq u/v$ if $0 \leq u < v \leq \pi/2$ can also be applied because $0 < \sqrt{\Delta}(\eta\rho + \rho) = r\sqrt{\Delta} < \pi/2$.

By the preceding argument we get for every $p \in \bar{B}(a, \rho)$,

$$d(x,p) - d(y,p) > \frac{1-\alpha}{\alpha} d(x,y).$$

Hence

$$\begin{aligned} f_\mu(x) - f_\mu(y) &= \int_{\bar{B}(a,\rho)} (d(x,p) - d(y,p))\mu(dp) + \int_{M \setminus \bar{B}(a,\rho)} (d(x,p) - d(y,p))\mu(dp) \\ &> \int_{\bar{B}(a,\rho)} \frac{1-\alpha}{\alpha} d(x,y)\mu(dp) - \int_{M \setminus \bar{B}(a,\rho)} d(x,y)\mu(dp) \\ &= 0. \end{aligned}$$

Particularly, we obtain $Q_\mu \subset \bar{B}(a, \eta\rho)$. $\square$

Chapter 4

Stochastic and deterministic algorithms for computing means of probability measures

Abstract

This chapter is a collaborative work of Marc Arnaudon, Clément Dombry, Anthony Phan and me. We consider a probability measure μ supported by a regular geodesic ball in a manifold and, for any $p \geq 1$, define a stochastic algorithm which converges almost surely to the p-mean e_p of μ. Assuming furthermore that the functional to minimize is regular around e_p, we prove that a natural renormalization of the inhomogeneous Markov chain converges in law into an inhomogeneous diffusion process. We give the explicit expression of this process, as well as its local characteristic. After that, the performance of the stochastic algorithms are illustrated by simulations. Finally, we show that the p-mean of μ can also be computed by the method of gradient descent. The questions concerning the choice of stepsizes and error estimates of this deterministic method are also considered.

4.1 p-means in regular geodesic balls

Let M be a Riemannian manifold with pinched sectional curvatures. Let $\alpha, \beta > 0$ such that α^2 is a positive upper bound for sectional curvatures on M, and $-\beta^2$ is a negative lower bound for sectional curvatures on M. Denote by ρ the Riemannian distance on M.

In M consider a geodesic ball $B(a, r)$ with $a \in M$. Let μ be a probability measure with support included in a compact convex subset K_μ of $B(a, r)$. Fix $p \in [1, \infty)$. We will always make the following assumptions on (r, p, μ):

Assumption 4.1. The support of μ is not reduced to one point. Either $p > 1$ or the support of μ is not contained in a line, and the radius r satisfies

$$r < r_{\alpha,p} \quad \text{with} \left\{ \begin{array}{ll} r_{\alpha,p} & = \frac{1}{2}\min\left\{\text{inj}(M), \frac{\pi}{2\alpha}\right\} \quad \text{if } p \in [1,2) \\ r_{\alpha,p} & = \frac{1}{2}\min\left\{\text{inj}(M), \frac{\pi}{\alpha}\right\} \quad \text{if } p \in [2,\infty) \end{array} \right. \tag{1.1}$$

Note that $B(a,r)$ is convex if $r < \frac{1}{2}\min\left\{\text{inj}(M), \frac{\pi}{\alpha}\right\}$.

Under assumption 4.1, it has been proved in [1, Theorem 2.1] that the function

$$\begin{aligned} H_p : M &\longrightarrow \mathbb{R}_+ \\ x &\longmapsto \int_M \rho^p(x,y)\mu(dy) \end{aligned} \tag{1.2}$$

has a unique minimizer e_p in M, the p-mean of μ, and moreover $e_p \in B(a,r)$. If $p = 1$, e_1 is the median of μ.

It is easily checked that if $p \in [1,2)$, then H_p is strictly convex on $B(a,r)$. On the other hand, if $p \geq 2$ then H_p is of class C^2 on $B(a,r)$.

4.2 Stochastic algorithms for computing p-means

The following proposition gives the fundamental estimations for the convergence of our stochastic algorithms.

Proposition 4.2. *Let K be a convex subset of $B(a,r)$ containing the support of μ. Then there exists $C_{p,\mu,K} > 0$ such that for all $x \in K$,*

$$H_p(x) - H_p(e_p) \geq \frac{C_{p,\mu,K}}{2}\rho(x,e_p)^2. \tag{2.1}$$

Moreover if $p \geq 2$ then we can choose $C_{p,\mu,K}$ so that for all $x \in K$,

$$\|\operatorname{grad}_x H_p\|^2 \geq C_{p,\mu,K}\left(H_p(x) - H_p(e_p)\right). \tag{2.2}$$

Proof. For simplicity, let us write shortly $e = e_p$ in the proofs. For $p = 1$ this is a direct consequence of [91, Theorem 3.7]

Next we consider the case $p \in (1,2)$.

Let $K \subset B(a,r)$ be a compact convex set containing the support of μ. Let $x \in K\backslash\{e\}$, $t = \rho(e,x)$, $u \in T_eM$ the unit vector such that $\exp_e(\rho(e,x)u) = x$, and γ_u the geodesic with initial speed u : $\dot{\gamma}_u(0) = u$. For $y \in K$, letting $h_y(s) = \rho(\gamma_u(s),y)^p$, $s \in [0,t]$, we have since $p > 1$

$$h_y(t) = h_y(0) + th_y'(0) + \int_0^t (t-s)h_y''(s)\,ds$$

with the convention $h_y''(s) = 0$ when $\gamma_u(s) = y$. Indeed, if $y \notin \gamma([0,t])$ then h_y is smooth, and if $y \in \gamma([0,t])$, say $y = \gamma(s_0)$ then $h_y(s) = |s - s_0|^p$ and the formula can easily be checked.

By standard calculation,

$$\begin{aligned} & h_y''(s) \\ & \geq p\rho(\gamma_u(s), y)^{p-2} \\ & \quad \times \Big((p-1)\|\dot\gamma_u(s)^{T(y)}\|^2 + \|\dot\gamma_u(s)^{N(y)}\|^2 \alpha\rho(\gamma_u(s), y)\cot\left(\alpha\rho(\gamma_u(s), y)\right)\Big) \end{aligned} \tag{2.3}$$

with $\dot\gamma_u(s)^{T(y)}$ (resp. $\dot\gamma_u(s)^{N(y)}$) the tangential (resp. the normal) part of $\dot\gamma_u(s)$ with respect to $n(\gamma_u(s), y) = \dfrac{1}{\rho(\gamma_u(s), y)} \exp^{-1}_{\gamma_u(s)}(y)$:

$$\dot\gamma_u(s)^{T(y)} = \langle \dot\gamma_u(s), n(\gamma_u(s), y)\rangle n(\gamma_u(s), y), \quad \dot\gamma_u(s)^{N(y)} = \dot\gamma_u(s) - \dot\gamma_u(s)^{T(y)}.$$

From this we get

$$h_y''(s) \geq p\rho(\gamma_u(s), y)^{p-2}\left(\min\left(p-1, 2\alpha r\cot\left(2\alpha r\right)\right)\right). \tag{2.4}$$

Now

$$\begin{aligned} & H_p(\gamma_u(t')) \\ & = \int_{B(a,r)} h_y(\gamma_u(t'))\,\mu(dy) \\ & = \int_{B(a,r)} h_y(0)\,\mu(dy) + t'\int_{B(a,r)} h_y'(0)\,\mu(dy) + \int_0^{t'} (t'-s)\left(\int_{B(a,r)} h_y(s)''\,\mu(dy)\right) ds \end{aligned}$$

and $H_p(\gamma_u(t'))$ attains its minimum at $t' = 0$, so $\displaystyle\int_{B(a,r)} h_y'(0)\,\mu(dy) = 0$ and we have

$$H_p(x) = H_p(\gamma_u(t)) = H_p(e) + \int_0^t (t-s)\left(\int_{B(a,r)} h_y(s)''\,\mu(dy)\right) ds.$$

Using Equation (2.4) we get

$$\begin{aligned} & H_p(x) \geq H_p(e) \\ & + \int_0^t \left((t-s)\int_{B(a,r)} p\rho(\gamma_u(s), y)^{p-2}\left(\min\left(p-1, 2\alpha r\cot\left(2\alpha r\right)\right)\right)\,\mu(dy)\right) ds. \end{aligned} \tag{2.5}$$

Since $p \leq 2$ we have $\rho(\gamma_u(s), y)^{p-2} \geq (2r)^{p-2}$ and

$$H_p(x) \geq H_p(e) + \frac{t^2}{2} p(2r)^{p-2}\left(\min\left(p-1, 2\alpha r\cot\left(2\alpha r\right)\right)\right). \tag{2.6}$$

So letting

$$C_{p,\mu,K} = p(2r)^{p-2}\left(\min\left(p-1, 2\alpha r\cot\left(2\alpha r\right)\right)\right)$$

we obtain

$$H_p(x) \geq H_p(e) + \frac{C_{p,\mu,K}\rho(e,x)^2}{2}. \tag{2.7}$$

To finish let us consider the case $p \geq 2$.

In the proof of [1, Theorem 2.1], it is shown that e is the only zero of the maps $x \mapsto \operatorname{grad}_x H_p$ and $x \mapsto H_p(x) - H_p(e)$, and that $\nabla dH_p(e)$ is strictly positive. This implies that (2.1) and (2.2) hold on some neighbourhood $B(e,\varepsilon)$ of e. By compactness and the fact that $H_p - H_p(e)$ and $\operatorname{grad} H_p$ do not vanish on $K\backslash B(e,\varepsilon)$ and $H_p - H_p(e)$ is bounded, possibly modifying the constant $C_{p,\mu,K}$, (2.1) and (2.2) also holds on $K\backslash B(e,\varepsilon)$.

□

We now state our main result: we define a stochastic gradient algorithm $(X_k)_{k\geq 0}$ to approximate the p-mean e_p and prove its convergence. In the sequel, we fix

$$K = \bar{B}(a, r-\varepsilon) \quad \text{with} \quad \varepsilon = \frac{\rho(K_\mu, B(a,r)^c)}{2}. \tag{2.8}$$

Theorem 4.3. *Let $(P_k)_{k\geq 1}$ be a sequence of independent $B(a,r)$-valued random variables, with law μ. Let $(t_k)_{k\geq 1}$ be a sequence of positive numbers satisfying*

$$\forall k \geq 1, \quad t_k \leq \min\left(\frac{1}{C_{p,\mu,K}}, \frac{\rho(K_\mu, B(a,r)^c)}{2p(2r)^{p-1}}\right), \tag{2.9}$$

$$\sum_{k=1}^{\infty} t_k = +\infty \quad \textit{and} \quad \sum_{k=1}^{\infty} t_k^2 < \infty. \tag{2.10}$$

Letting $x_0 \in K$, define inductively the random walk $(X_k)_{k\geq 0}$ by

$$X_0 = x_0 \quad \textit{and for } k \geq 0 \quad X_{k+1} = \exp_{X_k}\left(-t_{k+1} \operatorname{grad}_{X_k} F_p(\cdot, P_{k+1})\right) \tag{2.11}$$

where $F_p(x,y) = \rho^p(x,y)$, with the convention $\operatorname{grad}_x F_p(\cdot, x) = 0$.

The random walk $(X_k)_{k\geq 1}$ converges in L^2 and almost surely to e_p.

Proof. Note that, for $x \neq y$,

$$\operatorname{grad}_x F(\cdot, y) = p\rho^{p-1}(x,y)\frac{-\exp_x^{-1}(y)}{\rho(x,y)} = -p\rho^{p-1}(x,y)n(x,y),$$

whith $n(x,y) := \dfrac{\exp_x^{-1}(y)}{\rho(x,y)}$ a unit vector. So, with the condition (2.9) on t_k, the random walk $(X_k)_{k\geq 0}$ cannot exit K: if $X_k \in K$ then there are two possibilities for X_{k+1}:

- either X_{k+1} is in the geodesic between X_k and P_{k+1} and belongs to K by convexity of K;

- or X_{k+1} is after P_{k+1}, but since

$$\begin{aligned}\|t_{k+1}\operatorname{grad}_{X_k}F_p(\cdot,P_{k+1})\| &= t_{k+1}p\rho^{p-1}(X_k,P_{k+1})\\ &\le \frac{\rho(K_\mu,B(a,r)^c)}{2p(2r)^{p-1}}p\rho^{p-1}(X_k,P_{k+1})\\ &\le \frac{\rho(K_\mu,B(a,r)^c)}{2},\end{aligned}$$

  we have in this case

$$\rho(P_{k+1},X_{k+1})\le\frac{\rho(K_\mu,B(a,r)^c)}{2}$$

  which implies that $X_{k+1}\in K$.

First consider the case $p\in[1,2)$.
For $k\ge 0$ let

$$t\mapsto E(t):=\frac{1}{2}\rho^2\left(e,\gamma(t)\right),$$

$\gamma(t)_{t\in[0,t_{k+1}]}$ the geodesic satisfying $\dot\gamma(0)=-\operatorname{grad}_{X_k}F_p(\cdot,P_{k+1})$. We have for all $t\in[0,t_{k+1}]$

$$E''(t)\le C(\beta,r,p):=p^2(2r)^{2p-1}\beta\coth(2\beta r) \tag{2.12}$$

(see e.g. [91]). By Taylor formula,

$$\begin{aligned}&\rho(X_{k+1},e)^2\\ &=2E(t_{k+1})\\ &=2E(0)+2t_{k+1}E'(0)+t_{k+1}^2E''(t)\quad\text{for some } t\in[0,t_{k+1}]\\ &\le\rho(X_k,e)^2+2t_{k+1}\langle\operatorname{grad}_{X_k}F_p(\cdot,P_{k+1}),\exp_{X_k}^{-1}(e)\rangle+t_{k+1}^2C(\beta,r,p).\end{aligned}$$

Now from the convexity of $x\mapsto F_p(x,y)$ we have for all $x,y\in B(a,r)$

$$F_p(e,y)-F_p(x,y)\ge\left\langle\operatorname{grad}_x F_p(\cdot,y),\exp_x^{-1}(e)\right\rangle. \tag{2.13}$$

This applied with $x=X_k$, $y=P_{k+1}$ yields

$$\rho(X_{k+1},e)^2\le\rho(X_k,e)^2-2t_{k+1}\left(F_p(X_k,P_{k+1})-F_p(e,P_{k+1})\right)+C(\beta,r,p)t_{k+1}^2. \tag{2.14}$$

Letting for $k\ge 0$ $\mathscr{F}_k=\sigma(X_\ell,0\le\ell\le k)$, we get

$$\begin{aligned}&\mathbb{E}\left[\rho(X_{k+1},e)^2|\mathscr{F}_k\right]\\ &\le\rho(X_k,e)^2-2t_{k+1}\int_{B(a,r)}\left(F_p(X_k,y)-F_p(e,y)\right)\mu(dy)+C(\beta,r,p)t_{k+1}^2\\ &=\rho(X_k,e)^2-2t_{k+1}\left(H_p(X_k)-H_p(e)\right)+C(\beta,r,p)t_{k+1}^2\\ &\le\rho(X_k,e)^2+C(\beta,r,p)t_{k+1}^2\end{aligned}$$

so that the process $(Y_k)_{k\geq 0}$ defined by

$$Y_0 = \rho(X_0, e)^2 \quad \text{and for } k \geq 1 \quad Y_k = \rho(X_k, e)^2 - C(\beta, r, p) \sum_{j=1}^{k} t_j^2 \tag{2.15}$$

is a bounded supermartingale. So it converges in L^1 and almost surely. Consequently $\rho(X_k, e)^2$ also converges in L^1 and almost surely.

Let

$$a = \lim_{k\to\infty} \mathbb{E}\left[\rho(X_k, e)^2\right]. \tag{2.16}$$

We want to prove that $a = 0$. We already proved that

$$\mathbb{E}\left[\rho(X_{k+1}, e)^2 | \mathscr{F}_k\right] \leq \rho(X_k, e)^2 - 2t_{k+1}\left(H_p(X_k) - H_p(e)\right) + C(\beta, r, p)t_{k+1}^2. \tag{2.17}$$

Taking the expectation and using Proposition 4.2, we obtain

$$\mathbb{E}\left[\rho(X_{k+1}, e)^2\right] \leq \mathbb{E}\left[\rho(X_k, e)^2\right] - t_{k+1} C_{p,\mu,K} \mathbb{E}\left[\rho(X_k, e)^2\right] + C(\beta, r, p)t_{k+1}^2. \tag{2.18}$$

An easy induction proves that for $\ell \geq 1$,

$$\mathbb{E}\left[\rho(X_{k+\ell}, e)^2\right] \leq \prod_{j=1}^{\ell}(1 - C_{p,\mu,K} t_{k+j}) \mathbb{E}\left[\rho(X_k, e)^2\right] + C(\beta, r, p) \sum_{j=1}^{\ell} t_{k+j}^2. \tag{2.19}$$

Letting $\ell \to \infty$ and using the fact that $\sum_{j=1}^{\infty} t_{k+j} = \infty$ which implies

$$\prod_{j=1}^{\infty}(1 - C_{p,\mu,K} t_{k+j}) = 0,$$

we get

$$a \leq C(\beta, r, p) \sum_{j=1}^{\infty} t_{k+j}^2. \tag{2.20}$$

Finally using $\sum_{j=1}^{\infty} t_j^2 < \infty$ we obtain that $\lim_{k\to\infty} \sum_{j=1}^{\infty} t_{k+j}^2 = 0$, so $a = 0$. This proves L^2 and almost sure convergence.

Next assume that $p \geq 2$.

For $k \geq 0$ let

$$t \mapsto E_p(t) := H_p(\gamma(t)),$$

$\gamma(t)_{t\in[0,t_{k+1}]}$ the geodesic satisfying $\dot\gamma(0) = -\operatorname{grad}_{X_k} F_p(\cdot, P_{k+1})$. With a calculation similar to (2.12) we get for all $t \in [0, t_{k+1}]$

$$E_p''(t) \leq 2C(\beta, r, p) := p^3 (2r)^{3p-4} \left(2\beta r \coth(2\beta r) + p - 2\right). \tag{2.21}$$

(see e.g. [91]). By Taylor formula,

$$\begin{aligned}
H_p(X_{k+1}) &= E_p(t_{k+1}) \\
&= E_p(0) + t_{k+1}E_p'(0) + \frac{t_{k+1}^2}{2}E_p''(t) \quad \text{for some } t \in [0, t_{k+1}] \\
&\le H_p(X_k) + t_{k+1}\langle d_{X_k}H_p, \operatorname{grad}_{X_k} F_p(\cdot, P_{k+1})\rangle + t_{k+1}^2 C(\beta, r, p).
\end{aligned}$$

We get

$$\begin{aligned}
&\mathbb{E}\left[H_p(X_{k+1})|\mathscr{F}_k\right] \\
&\le H_p(X_k) - t_{k+1}\left\langle d_{X_k}H_p, \int_{B(a,r)} \operatorname{grad}_{X_k} F_p(\cdot, y)\mu(dy)\right\rangle + C(\beta, r, p)t_{k+1}^2 \\
&= H_p(X_k) - t_{k+1}\left\langle d_{X_k}H_p, \operatorname{grad}_{X_k} H_p(\cdot)\right\rangle + C(\beta, r, p)t_{k+1}^2 \\
&= H_p(X_k) - t_{k+1}\left\|\operatorname{grad}_{X_k} H_p(\cdot)\right\|^2 + C(\beta, r, p)t_{k+1}^2 \\
&\le H_p(X_k) - C_{p,\mu,K}t_{k+1}\left(H_p(X_k) - H_p(e)\right) + C(\beta, r, p)t_{k+1}^2
\end{aligned}$$

(by Proposition 4.2) so that the process $(Y_k)_{k\ge 0}$ defined by

$$Y_0 = H_p(X_0) - H_p(e) \quad \text{and for } k \ge 1 \quad Y_k = H_p(X_k) - H_p(e) - C(\beta, r, p)\sum_{j=1}^{k} t_j^2 \tag{2.22}$$

is a bounded supermartingale. So it converges in L^1 and almost surely. Consequently $H_p(X_k) - H_p(e)$ also converges in L^1 and almost surely.

Let

$$a = \lim_{k\to\infty} \mathbb{E}\left[H_p(X_k) - H_p(e)\right]. \tag{2.23}$$

We want to prove that $a = 0$. We already proved that

$$\begin{aligned}
&\mathbb{E}\left[H_p(X_{k+1}) - H_p(e)|\mathscr{F}_k\right] \\
&\le H_p(X_k) - H_p(e) - C_{p,\mu,K}t_{k+1}\left(H_p(X_k) - H_p(e)\right) + C(\beta, r, p)t_{k+1}^2.
\end{aligned} \tag{2.24}$$

Taking the expectation we obtain

$$\mathbb{E}\left[H_p(X_{k+1}) - H_p(e)\right] \le (1 - t_{k+1}C_{p,\mu,K})\mathbb{E}\left[H_p(X_k) - H_p(e)\right] + C(\beta, r, p)t_{k+1}^2 \tag{2.25}$$

so that proving that $a = 0$ is similar to the previous case.

Finally (2.1) proves that $\rho(X_k, e)^2$ converges in L^1 and almost surely to 0. □

In the following example, we focus on the case $M = \mathbb{R}^d$ and $p = 2$ where drastic simplifications occur.

Example 4.4. In the case when $M = \mathbb{R}^d$ and μ is a compactly supported probability measure on $\mathbb{R}^d$, the stochastic gradient algorithm (2.11) simplifies into

$$X_0 = x_0 \quad \text{and for } k \geq 0 \quad X_{k+1} = X_k - t_{k+1} \operatorname{grad}_{X_k} F_p(\cdot, P_{k+1}).$$

If furthermore $p = 2$, clearly $e_2 = \mathbb{E}[P_1]$ and $\operatorname{grad}_x F_p(\cdot, y) = 2(x - y)$, so that the linear relation

$$X_{k+1} = (1 - 2t_{k+1})X_k + 2t_{k+1}P_{k+1}, \quad k \geq 0$$

holds true and an easy induction proves that

$$X_k = x_0 \prod_{j=0}^{k-1}(1 - 2t_{k-j}) + 2\sum_{j=0}^{k-1} P_{k-j}t_{k-j} \prod_{\ell=0}^{j-1}(1 - 2t_{k-\ell}), \quad k \geq 1. \tag{2.26}$$

Now, taking $t_k = \dfrac{1}{2k}$, we have

$$\prod_{j=0}^{k-1}(1 - 2t_{k-j}) = 0 \quad \text{and} \quad \prod_{\ell=0}^{j-1}(1 - 2t_{k-\ell}) = \frac{k - j}{k}$$

so that

$$X_k = \sum_{j=0}^{k-1} P_{k-j}\frac{1}{k} = \frac{1}{k}\sum_{j=1}^{k} P_j.$$

The stochastic gradient algorithm estimating the mean e_2 of μ is given by the empirical mean of a growing sample of independent random variables with distribution μ. In this simple case, the result of Theorem 4.3 is nothing but the strong law of large numbers. Moreover, fluctuations around the mean are given by the central limit theorem and Donsker's theorem.

4.3 Fluctuations of the stochastic gradient algorithm

The notations are the same as in the beginning of section 4.1. We still make assumption 4.1. Let us define K and ε as in (2.8) and let

$$\delta_1 = \min\left(\frac{1}{C_{p,\mu,K}}, \frac{\rho(K_\mu, B(a,r)^c)}{2p(2r)^{p-1}}\right). \tag{3.1}$$

We consider the time inhomogeneous M-valued Markov chain (2.11) in the particular case when

$$t_k = \min\left(\frac{\delta}{k}, \delta_1\right), \quad k \geq 1 \tag{3.2}$$

for some $\delta > 0$. The particular sequence $(t_k)_{k\geq 1}$ defined by (3.2) satisfies (2.9) and (2.10), so Theorem 4.3 holds true and the stochastic gradient algorithm $(X_k)_{k\geq 0}$ converges a.s. and in L^2 to the p-mean e_p.

In order to study the fluctuations around the p-mean e_p, we define for $n \geq 1$ the rescaled $T_{e_p}M$-valued Markov chain $(Y_k^n)_{k\geq 0}$ by

$$Y_k^n = \frac{k}{\sqrt{n}} \exp_{e_p}^{-1} X_k. \tag{3.3}$$

We will prove convergence of the sequence of process $(Y_{[nt]}^n)_{t\geq 0}$ to a non-homogeneous diffusion process. The limit process is defined in the following proposition:

Proposition 4.5. *Assume that H_p is C^2 in a neighborhood of e_p, and that $\delta > C_{p,\mu,K}^{-1}$. Define*

$$\Gamma = \mathbb{E}\left[\operatorname{grad}_{e_p} F_p(\cdot, P_1) \otimes \operatorname{grad}_{e_p} F_p(\cdot, P_1)\right]$$

and $G_\delta(t)$ the generator

$$G_\delta(t)f(y) := \langle d_y f, t^{-1}(y - \delta\nabla dH_p(y,\cdot)^\sharp)\rangle + \frac{\delta^2}{2}\operatorname{Hess}_y f\,(\Gamma) \tag{3.4}$$

where $\nabla dH_p(y,\cdot)^\sharp$ denotes the dual vector of the linear form $\nabla dH_p(y,\cdot)$.

There exists a unique inhomogeneous diffusion process $(y_\delta(t))_{t>0}$ on $T_{e_p}M$ with generator $G_\delta(t)$ and converging in probability to 0 as $t \to 0^+$.

The process y_δ is continuous, converges a.s. to 0 as $t \to 0^+$ and has the following integral representation:

$$y_\delta(t) = \sum_{i=1}^d t^{1-\delta\lambda_i}\int_0^t s^{\delta\lambda_i - 1}\langle \delta\sigma\, dB_s, e_i\rangle e_i, \quad t \geq 0, \tag{3.5}$$

where B_t is a standard Brownian motion on $T_{e_p}M$, $\sigma \in \operatorname{End}(T_{e_p}M)$ satisfies $\sigma\sigma^* = \Gamma$, $(e_i)_{1\leq i\leq d}$ is an orthonormal basis diagonalizing the symmetric bilinear form $\nabla dH_p(e_p)$ and $(\lambda_i)_{1\leq i\leq d}$ are the associated eigenvalues.

Note that the integral representation (3.5) implies that y_δ is the centered Gaussian process with covariance

$$\mathbb{E}\left[y_\delta^i(t_1)y_\delta^j(t_2)\right] = \frac{\delta^2\Gamma(e_i^*\otimes e_j^*)}{\delta(\lambda_i+\lambda_j)-1} t_1^{1-\delta\lambda_i} t_2^{1-\delta\lambda_j}(t_1\wedge t_2)^{\delta(\lambda_i+\lambda_j)-1}, \tag{3.6}$$

where $y_\delta^i(t) = \langle y_\delta(t), e_i\rangle$, $1 \leq i,j \leq d$ and $t_1, t_2 \geq 0$.

Proof. Fix $\varepsilon > 0$. Any diffusion process on $[\varepsilon,\infty)$ with generator $G_\delta(t)$ is solution of a sde of the type

$$dy_t = \frac{1}{t}L_\delta(y_t)\,dt + \delta\sigma\, dB_t \tag{3.7}$$

where $L_\delta(y) = y - \delta\nabla dH_p(y,\cdot)^\sharp$ and B_t and σ are as in Proposition 4.5. This sde can be solved explicitely on $[\varepsilon,\infty)$. The symmetric endomorphism $y \mapsto \nabla dH_p(y,\cdot)^\sharp$ is diagonalisable in the orthonormal basis $(e_i)_{1\leq i\leq d}$ with eigenvalues $(\lambda_i)_{1\leq i\leq d}$. The endomorphism $L_\delta = \mathrm{id} - \delta \nabla dH_p(e)(\mathrm{id},\cdot)^\sharp$ is also diagonalisable in this basis with eigenvalues $(1-\delta\lambda_i)_{1\le i\le d}$. The solution $y_t = \sum_{i=1}^d y_t^i e_i$ of (3.7) started at $y_\varepsilon = \sum_{i=1}^d y_\varepsilon^i e_i$ is given by

$$y_t = \sum_{i=1}^d \left(y_\varepsilon^i \varepsilon^{\delta\lambda_i - 1} + \int_\varepsilon^t s^{\delta\lambda_i - 1} \langle \delta\sigma\, dB_s, e_i\rangle \right) t^{1-\delta\lambda_i} e_i, \quad t \ge \varepsilon. \tag{3.8}$$

Now by definition of $C_{p,\mu,K}$ we clearly have

$$C_{p,\mu,K} \le \min_{1\le i\le d} \lambda_i. \tag{3.9}$$

So the condition $\delta C_{p,\mu,K} > 1$ implies that for all i, $\delta\lambda_i - 1 > 0$, and as $\varepsilon \to 0$,

$$\int_\varepsilon^t s^{\delta\lambda_i - 1}\langle \delta\sigma\, dB_s, e_i\rangle \to \int_0^t s^{\delta\lambda_i - 1}\langle \delta\sigma\, dB_s, e_i\rangle \quad \text{in probability.} \tag{3.10}$$

Assume that a continuous solution y_t converging in probability to 0 as $t \to 0^+$ exists. Since $y_\varepsilon^i \varepsilon^{\delta\lambda_i - 1} \to 0$ in probability as $\varepsilon \to 0$, we necessarily have using (3.10)

$$y_t = \sum_{i=1}^d t^{1-\delta\lambda_i} \int_0^t s^{\delta\lambda_i - 1}\langle \delta\sigma\, dB_s, e_i\rangle e_i, \quad t \ge 0. \tag{3.11}$$

Note y_δ^i is Gaussian with variance $\dfrac{t\delta^2\Gamma(e_i^* \otimes e_i^*)}{2\delta\lambda_i - 1}$, so it converges in L^2 to 0 as $t \to 0$. Conversely, it is easy to check that equation (3.11) defines a solution to (3.7).

To prove the a.s. convergence to 0 we use the representation

$$\int_0^t s^{\delta\lambda_i - 1}\langle \delta\sigma\, dB_s, e_i\rangle = B^i_{\varphi_i(t)}$$

where B_s^i is a Brownian motion and $\varphi_i(t) = \dfrac{\delta^2\Gamma(e_i^*\otimes e_i^*)}{2\delta\lambda_i - 1} t^{2\delta\lambda_i - 1}$. Then by the law of iterated logarithm

$$\limsup_{t\downarrow 0} t^{1-\delta\lambda_i} B^i_{\varphi_i(t)} \le \limsup_{t\downarrow 0} t^{1-\delta\lambda_i}\sqrt{2\varphi_i(t)\ln\ln\left(\varphi_i^{-1}(t)\right)}$$

But for t small we have

$$\sqrt{2\varphi_i(t)\ln\ln\left(\varphi_i^{-1}(t)\right)} \le t^{\delta\lambda_i - 3/4}$$

so

$$\limsup_{t\downarrow 0} t^{1-\delta\lambda_i} B^i_{\varphi_i(t)} \le \lim_{t\downarrow 0} t^{1/4} = 0.$$

This proves a.s. convergence to 0. Continuity is easily checked using the integral representation (3.11). $\square$

Our main result on the fluctuations of the stochastic gradient algorithm is the following:

Theorem 4.6. *Assume that either e_p does not belong to the support of μ or $p \geq 2$. Assume furthermore that $\delta > C_{p,\mu,K}^{-1}$. The sequence of processes $\left(Y_{[nt]}^n\right)_{t\geq 0}$ weakly converges in $\mathbb{D}((0,\infty), T_{e_p}M)$ to y_δ.*

Remark 4.7. The assumption on e_p implies that H_p is of class C^2 in a neighbourhood of e_p. In the case $p > 1$, in the "generic" situation for applications, μ is a discrete measure and e_p does not belong to its support. For $p = 1$ one has to be more careful since if μ is equidistributed in a random set of points, then with positive probability e_1 belongs to the support of μ.

Remark 4.8. From section 4.1 we know that, when $p \in (1, 2]$, the constant

$$C_{p,\mu,K} = p(2r)^{p-2} \left(\min\left(p-1, 2\alpha r \cot\left(2\alpha r\right)\right)\right)$$

is explicit. The constraint $\delta > C_{p,\mu,K}^{-1}$ can easily be checked in this case.

Remark 4.9. In the case $M = \mathbb{R}^d$, $Y_k^n = \frac{k}{\sqrt{n}}(X_k - e_p)$ and the tangent space $T_{e_p}M$ is identified to $\mathbb{R}^d$. Theorem 4.6 holds and, in particular, when $t = 1$, we obtain a central limit Theorem: $\sqrt{n}(X_n - e_p)$ converges as $n \to \infty$ to a centered Gaussian d-variate distribution (with covariance structure given by (3.6) with $t_1 = t_2 = 1$). This is a central limit theorem: the fluctuations of the stochastic gradient algorithm are of scale $n^{-1/2}$ and asymptotically Gaussian.

Now let us introduce some notations needed in the proof of Theorem 4.6. Consider the time homogeneous Markov chain $(Z_k^n)_{k\geq 0}$ with state space $[0,\infty) \times T_eM$ defined by

$$Z_k^n = \left(\frac{k}{n}, Y_k^n\right). \tag{3.12}$$

The first component has a deterministic evolution and will be denoted by t_k^n; it satisfies

$$t_{k+1}^n = t_k^n + \frac{1}{n}, \quad k \geq 0. \tag{3.13}$$

Let k_0 be such that

$$\frac{\delta}{k_0} < \delta_1. \tag{3.14}$$

Using equations (2.11), (3.3) and (3.2), we have for $k \geq k_0$,

$$Y_{k+1}^n = \frac{nt_k^n + 1}{\sqrt{n}} \exp_e^{-1}\left(\exp_{\exp_e \frac{1}{\sqrt{n}t_k^n} Y_k^n}\left(-\frac{\delta}{nt_k^n + 1} \operatorname{grad}_{\frac{1}{\sqrt{n}t_k^n} Y_k^n} F_p(\cdot, P_{k+1})\right)\right). \tag{3.15}$$

Consider the transition kernel $P^n(z,dz')$ on $(0,\infty)\times T_eM$ defined for $z=(t,y)$ by

$$P^n(z,A)=$$
$$\mathbb{P}\left[\left(t+\frac{1}{n},\frac{nt+1}{\sqrt{n}}\exp_e^{-1}\left(\exp_{\exp_e\frac{1}{\sqrt{nt}}y}\left(-\frac{\delta}{nt+1}\operatorname{grad}_{\exp_e\frac{1}{\sqrt{nt}}y}F_p(\cdot,P_1)\right)\right)\right)\in A\right] \tag{3.16}$$

where $A\in\mathscr{B}((0,\infty)\times T_eM)$. Clearly this transition kernel drives the evolution of the Markov chain $(Z_k^n)_{k\geq k_0}$.

For the sake of clarity, we divide the proof of Theorem 4.6 into four lemmas.

Lemma 4.10. *Assume that either $p\geq 2$ or e does not belong to the support* $\operatorname{supp}(\mu)$ *of μ (note this implies that for all $x\in\operatorname{supp}(\mu)$ the function $F_p(\cdot,x)$ is of class C^2 in a neighbourhood of e). Fix $\delta>0$. Let B be a bounded set in T_eM and let $0<\varepsilon<T$. We have for all C^2 function f on T_eM*

$$\begin{aligned}&n\left(f\left(\frac{nt+1}{\sqrt{n}}\exp_e^{-1}\left(\exp_{\exp_e\frac{1}{\sqrt{nt}}y}\left(-\frac{\delta}{nt+1}\operatorname{grad}_{\exp_e\frac{1}{\sqrt{nt}}y}F_p(\cdot,x)\right)\right)\right)-f(y)\right)\\&=\left\langle d_yf,\frac{y}{t}\right\rangle-\sqrt{n}\langle d_yf,\delta\operatorname{grad}_eF_p(\cdot,x)\rangle-\delta\nabla dF_p(\cdot,x)\left(\operatorname{grad}_yf,\frac{y}{t}\right)\\&+\frac{\delta^2}{2}\operatorname{Hess}_yf\left(\operatorname{grad}_eF_p(\cdot,x)\otimes\operatorname{grad}_eF_p(\cdot,x)\right)+O\left(\frac{1}{\sqrt{n}}\right)\end{aligned}\tag{3.17}$$

uniformly in $y\in B$, $x\in\operatorname{supp}(\mu)$, $t\in[\varepsilon,T]$.

Proof. Let $x\in\operatorname{supp}(\mu)$, $y\in T_eM$, $u,v\in\mathbb{R}$ sufficiently close to 0, and $q=\exp_e\left(\dfrac{uy}{t}\right)$. For $s\in[0,1]$ denote by $a\mapsto c(a,s,u,v)$ the geodesic with endpoints $c(0,s,u,v)=e$ and

$$c(1,s,u,v)=\exp_{\exp_e\left(\frac{uy}{t}\right)}\left(-vs\operatorname{grad}_{\exp_e\left(\frac{uy}{t}\right)}F_p(\cdot,x)\right):$$

$$c(a,s,u,v)=\exp_e\left\{a\exp_e^{-1}\left[\exp_{\exp_e\left(\frac{uy}{t}\right)}\left(-sv\operatorname{grad}_{\exp_e\left(\frac{uy}{t}\right)}F_p(\cdot,x)\right)\right]\right\}.$$

This is a C^2 function of $(a,s,u,v)\in[0,1]^2\times(-\eta,\eta)^2$, η sufficiently small. It also depends in a C^2 way of x and y. Letting $c(a,s)=c\left(a,s,\dfrac{1}{\sqrt{n}},\dfrac{\delta}{nt+1}\right)$, we have

$$\exp_e^{-1}\left(\exp_{\exp_e\frac{1}{\sqrt{nt}}y}\left(-\frac{\delta}{nt+1}\operatorname{grad}_{\exp_e\frac{1}{\sqrt{nt}}y}F_p(\cdot,x)\right)\right)=\partial_ac(0,1).$$

So we need a Taylor expansion up to order n^{-1} of $\dfrac{nt+1}{\sqrt{n}}\partial_ac(0,1)$.

We have $c(a,s,0,1) = \exp_e(-as\,\mathrm{grad}_e F_p(\cdot,x))$ and this implies

$$\partial_s^2\partial_a c(0,s,0,1) = 0, \quad \text{so} \quad \partial_s^2\partial_a c(0,s,u,1) = O(u).$$

On the other hand the identities $c(a,s,u,v) = c(a,sv,u,1)$ yields $\partial_s^2\partial_a c(a,s,u,v) = v^2\partial_s^2\partial_a c(a,s,u,1)$, so we obtain

$$\partial_s^2\partial_a c(0,s,u,v) = O(uv^2)$$

and this yields

$$\partial_s^2\partial_a c(0,s) = O(n^{-5/2}),$$

uniformly in s,x,y,t. But since

$$\|\partial_a c(0,1) - \partial_a c(0,0) - \partial_s\partial_a c(0,0)\| \le \frac{1}{2}\sup_{s\in[0,1]}\|\partial_s^2\partial_a c(0,s)\|$$

we only need to estimate $\partial_a c(0,0)$ and $\partial_s\partial_a c(0,0)$.

Denoting by $J(a)$ the Jacobi field $\partial_s c(a,0)$ we have

$$\frac{nt+1}{\sqrt{n}}\partial_a c(0,1) = \frac{nt+1}{\sqrt{n}}\partial_a c(0,0) + \frac{nt+1}{\sqrt{n}}\dot{J}(0) + O\left(\frac{1}{n^2}\right).$$

On the other hand

$$\frac{nt+1}{\sqrt{n}}\partial_a c(0,0) = \frac{nt+1}{\sqrt{n}}\frac{y}{\sqrt{nt}} = y + \frac{y}{nt}$$

so it remains to estimate $\dot{J}(0)$.

The Jacobi field $a \mapsto J(a,u,v)$ with endpoints $J(0,u,v) = 0_e$ and

$$J(1,u,v) = -v\,\mathrm{grad}_{\exp_e\left(\frac{uy}{t}\right)} F_p(\cdot,x)$$

satisfies

$$\nabla_a^2 J(a,u,v) = -R(J(a,u,v),\partial_a c(a,0,u,v))\partial_a c(a,0,u,v) = O(u^2 v).$$

This implies that

$$\nabla_a^2 J(a) = O(n^{-2}).$$

Consequently, denoting by $P_{x_1,x_2} : T_{x_1}M \to T_{x_2}M$ the parallel transport along the minimal geodesic from x_1 to x_2 (whenever it is unique) we have

$$P_{c(1,0),e}J(1) = J(0) + \dot{J}(0) + O(n^{-2}) = \dot{J}(0) + O(n^{-2}). \tag{3.18}$$

But we also have

$$\begin{aligned} P_{c(1,0,u,v),e}J(1,u,v) &= P_{c(1,0,u,v),e}\Big(-v\,\mathrm{grad}_{c(1,0,u,v)} F_p(\cdot,x)\Big) \\ &= -v\,\mathrm{grad}_e F_p(\cdot,x) - v\nabla_{\partial_a c(0,0,u,v)}\,\mathrm{grad}_{\cdot} F_p(\cdot,x) + O(vu^2) \\ &= -v\,\mathrm{grad}_e F_p(\cdot,x) - v\nabla dF_p(\cdot,x)\left(\frac{uy}{t},\cdot\right)^{\sharp} + O(vu^2) \end{aligned}$$

where we used $\partial_a c(0,0,u,v) = \frac{uy}{t}$ and for vector fields A, B on TM and a C^2 function f_1 on M

$$\begin{aligned}\langle \nabla_{A_e} \operatorname{grad} f_1, B_e\rangle &= A_e\langle \operatorname{grad} f_1, B_e\rangle - \langle \operatorname{grad} f_1, \nabla_{A_e} B\rangle \\ &= A_e\langle df_1, B_e\rangle - \langle df_1, \nabla_{A_e} B\rangle \\ &= \nabla df_1(A_e, B_e)\end{aligned}$$

which implies

$$\nabla_{A_e} \operatorname{grad} f_1 = \nabla df_1(A_e, \cdot)^\sharp.$$

We obtain

$$P_{c(1,0),e} J(1) = -\frac{\delta}{nt+1} \operatorname{grad}_e F_p(\cdot, x) - \frac{\delta}{\sqrt{n}(nt+1)} \nabla dF_p(\cdot, x)\left(\frac{y}{t}, \cdot\right)^\sharp + O(n^{-2}).$$

Combining with (3.18) this gives

$$\dot{J}(0) = -\frac{\delta}{nt+1} \operatorname{grad}_e F_p(\cdot, x) - \frac{\delta}{nt+1} \nabla dF_p(\cdot, x)\left(\frac{y}{\sqrt{n}t}, \cdot\right)^\sharp + O\left(\frac{1}{n^2}\right).$$

So finally

$$\frac{nt+1}{\sqrt{n}} \partial_a c(0,1) = y + \frac{y}{nt} - \frac{\delta}{\sqrt{n}} \operatorname{grad}_e F_p(\cdot, x) - \delta \nabla dF_p(\cdot, x)\left(\frac{y}{nt}, \cdot\right)^\sharp + O\left(n^{-3/2}\right). \tag{3.19}$$

To get the final result we are left to make a Taylor expansion of f up to order 2. $\square$

Define the following quantities:

$$b_n(z) = n \int_{\{|z'-z| \le 1\}} (z'-z) P^n(z, dz') \tag{3.20}$$

and

$$a_n(z) = n \int_{\{|z'-z| \le 1\}} (z'-z) \otimes (z'-z) P^n(z, dz'). \tag{3.21}$$

The following property holds:

Lemma 4.11. *Assume that either $p \ge 2$ or e does not belong to the support* $\operatorname{supp}(\mu)$

(1) For all $R > 0$ and $\varepsilon > 0$, there exists n_0 such that for all $n \ge n_0$ and $z \in [\varepsilon, T] \times B(0_e, R)$, where $B(0_e, R)$ is the open ball in T_eM centered at the origin with radius R,

$$\int 1_{\{|z'-z| > 1\}} P^n(z, dz') = 0. \tag{3.22}$$

(2) For all $R > 0$ and $\varepsilon > 0$,

$$\lim_{n\to\infty} \sup_{z\in[\varepsilon,T]\times B(0_e,R)} |b_n(z) - b(z)| = 0 \tag{3.23}$$

with

$$b(z) = \left(1, \frac{1}{t} L_\delta(y)\right) \quad \textit{and} \quad L_\delta(y) = y - \delta \nabla dH(y,\cdot)^\sharp. \tag{3.24}$$

(3) For all $R > 0$ and $\varepsilon > 0$,

$$\lim_{n\to\infty} \sup_{z\in[\varepsilon,T]\times B(0_e,R)} |a_n(z) - a(z)| = 0 \tag{3.25}$$

with

$$a(z) = \delta^2 \mathrm{diag}(0,\Gamma) \quad \textit{and} \quad \Gamma = \mathbb{E}\left[\mathrm{grad}_e F_p(\cdot,P_1) \otimes \mathrm{grad}_e F_p(\cdot,P_1)\right]. \tag{3.26}$$

Proof. (1) We use the notation $z = (t,y)$ and $z' = (t',y')$. We have

$$\begin{aligned}
&\int 1_{\{|z'-z|>1\}}\, P^n(z,dz') \\
&= \int 1_{\{\max(|t'-t|,|y'-y|)>1\}} P^n(z,dz') \\
&= \int 1_{\{\max(\frac{1}{n},|y'-y|)>1\}} P^n(z,dz') \\
&= \mathbb{P}\left[\left|\frac{nt+1}{\sqrt{n}} \exp_e^{-1}\left(\exp_{\exp_e \frac{1}{\sqrt{nt}} y}\left(-\frac{\delta}{nt+1}\,\mathrm{grad}_{\exp_e \frac{1}{\sqrt{nt}} y} F_p(\cdot,P_1)\right)\right) - y\right| > 1\right].
\end{aligned}$$

On the other hand, since $F_p(\cdot,x)$ is of class C^2 in a neighbourhood of e, we have by (3.19)

$$\left|\frac{nt+1}{\sqrt{n}} \exp_e^{-1}\left(\exp_{\exp_e \frac{1}{\sqrt{nt}} y}\left(-\frac{\delta}{nt+1}\,\mathrm{grad}_{\exp_e \frac{1}{\sqrt{nt}} y} F_p(\cdot,P_1)\right)\right) - y\right| \le \frac{C\delta}{\sqrt{n\varepsilon}} \tag{3.27}$$

for some constant $C > 0$.

(2) Equation (3.22) implies that for $n \ge n_0$

$$\begin{aligned}
&b_n(z) \\
&= n\int (z'-z)\, P^n(z,dz') \\
&= n\left(\frac{1}{n}, \mathbb{E}\left[\frac{nt+1}{\sqrt{n}} \exp_e^{-1}\left(\exp_{\exp_e \frac{y}{\sqrt{nt}}}\left(-\frac{\delta}{nt+1}\,\mathrm{grad}_{\exp_e \frac{y}{\sqrt{nt}}} F_p(\cdot,P_1)\right)\right)\right] - y\right).
\end{aligned}$$

We have by lemma 4.10

$$\begin{aligned}
&n\left(\frac{nt+1}{\sqrt{n}}\exp_e^{-1}\left(\exp_{\exp_e \frac{1}{\sqrt{nt}}y}\left(-\frac{\delta}{nt+1}\operatorname{grad}_{\exp_e \frac{1}{\sqrt{nt}}y}F_p(\cdot,P_1)\right)\right)-y\right)\\
&=\frac{1}{t}y-\delta\sqrt{n}\operatorname{grad}_e F_p(\cdot,P_1)-\delta\nabla dF_p(\cdot,P_1)\left(\frac{1}{t}y,\cdot\right)^{\sharp}+O\left(\frac{1}{n^{1/2}}\right)
\end{aligned}$$

a.s. uniformly in n, and since

$$\mathbb{E}\left[\delta\sqrt{n}\operatorname{grad}_e F_p(\cdot,P_1)\right]=0,$$

this implies that

$$n\left(\mathbb{E}\left[\frac{nt+1}{\sqrt{n}}\exp_e^{-1}\left(\exp_{\exp_e \frac{1}{\sqrt{nt}}y}\left(-\frac{\delta}{nt+1}\operatorname{grad}_{\exp_e \frac{1}{\sqrt{nt}}y}F_p(\cdot,P_1)\right)\right)\right]-y\right)$$

converges to

$$\frac{1}{t}y-\mathbb{E}\left[\delta\nabla dF_p(\cdot,P_1)\left(\frac{1}{t}y,\cdot\right)^{\sharp}\right]=\frac{1}{t}y-\delta\nabla dH_p\left(\frac{1}{t}y,\cdot\right)^{\sharp}. \tag{3.28}$$

Moreover the convergence is uniform in $z\in[\varepsilon,T]\times B(0_e,R)$, so this yields (3.23).

(3) In the same way, using lemma 4.10,

$$\begin{aligned}
&n\int(y'-y)\otimes(y'-y)\,P^n(z,dz')\\
&=\frac{1}{n}\mathbb{E}\left[\left(-\sqrt{n}\delta\operatorname{grad}_e F_p(\cdot,P_1)\right)\otimes\left(-\sqrt{n}\delta\operatorname{grad}_e F_p(\cdot,P_1)\right)\right]+o(1)\\
&=\delta^2\mathbb{E}\left[\operatorname{grad}_e F_p(\cdot,P_1)\otimes\operatorname{grad}_e F_p(\cdot,P_1)\right]+o(1)
\end{aligned}$$

uniformly in $z\in[\varepsilon,T]\times B(0_e,R)$, so this yields (3.25).

$\square$

Lemma 4.12. *Suppose that $t_n=\dfrac{\delta}{n}$ for some $\delta>0$. For all $\delta>C_{p,\mu,K}^{-1}$,*

$$\sup_{n\geq 1} n\mathbb{E}\left[\rho^2(e,X_n)\right]<\infty. \tag{3.29}$$

Proof. First consider the case $p\in[1,2)$.

We know by (2.18) that there exists some constant $C(\beta,r,p)$ such that

$$\mathbb{E}\left[\rho^2(e,X_{k+1})\right]\leq\mathbb{E}\left[\rho^2(e,X_k)\right]\exp\left(-C_{p,\mu,K}t_{k+1}\right)+C(\beta,r,p)t_{k+1}^2. \tag{3.30}$$

From this (3.29) is a consequence of Lemma 0.0.1 (case $\alpha > 1$) in [70]. We give the proof for completeness. We deduce easily by induction that for all $k \geq k_0$,

$$\begin{aligned}&\mathbb{E}\left[\rho^2(e, X_k)\right]\\&\leq \mathbb{E}\left[\rho^2(e, X_{k_0})\right] \exp\left(-C_{p,\mu,K} \sum_{j=k_0+1}^{k} t_j\right) + C(\beta, r, p) \sum_{i=k_0+1}^{k} t_i^2 \exp\left(-C_{p,\mu,K} \sum_{j=i+1}^{k} t_j\right),\end{aligned} \tag{3.31}$$

where the convention $\sum_{j=k+1}^{k} t_j = 0$ is used. With $t_n = \frac{\delta}{n}$, the following inequality holds for all $i \geq k_0$ and $k \geq i$:

$$\sum_{j=i+1}^{k} t_j = \delta \sum_{j=i+1}^{k} \frac{1}{j} \geq \delta \int_{i+1}^{k+1} \frac{dt}{t} \geq \delta \ln \frac{k+1}{i+1}. \tag{3.32}$$

Hence,

$$\begin{aligned}&\mathbb{E}\left[\rho^2(e, X_k)\right]\\&\leq \mathbb{E}\left[\rho^2(e, X_{k_0})\right] \left(\frac{k_0+1}{k+1}\right)^{\delta C_{p,\mu,K}} + \frac{\delta^2 C(\beta, r, p)}{(k+1)^{\delta C_{p,\mu,K}}} \sum_{i=k_0+1}^{k} \frac{(i+1)^{\delta C_{p,\mu,K}}}{i^2}.\end{aligned} \tag{3.33}$$

For $\delta C_{p,\mu,K} > 1$ we have as $k \to \infty$

$$\frac{\delta^2 C(\beta, r, p)}{(k+1)^{\delta C_{p,\mu,K}}} \sum_{i=k_0+1}^{k} \frac{(i+1)^{\delta C_{p,\mu,K}}}{i^2} \sim \frac{\delta^2 C(\beta, r, p)}{(k+1)^{\delta C_{p,\mu,K}}} \frac{k^{\delta C_{p,\mu,K}-1}}{\delta C_{p,\mu,K} - 1} \sim \frac{\delta^2 C(\beta, r, p)}{\delta C_{p,\mu,K} - 1} k^{-1} \tag{3.34}$$

and

$$\mathbb{E}\left[\rho^2(e, X_{k_0})\right] \left(\frac{k_0+1}{k+1}\right)^{\delta C_{p,\mu,K}} = o(k^{-1}).$$

This implies that the sequence $k\mathbb{E}\left[\rho^2(e, X_k)\right]$ is bounded.

Next consider the case $p \geq 2$.

Now we have by (2.25) that

$$\mathbb{E}\left[H_p(X_{k+1}) - H_p(e)\right] \leq \mathbb{E}\left[H_p(X_k) - H_p(e)\right] \exp\left(-C_{p,\mu,K} t_{k+1}\right) + C(\beta, r, p) t_{k+1}^2. \tag{3.35}$$

From this, arguing similarly, we obtain that the sequence $k\mathbb{E}\left[H_p(X_k) - H_p(e)\right]$ is bounded. We conclude with (2.1). □

Lemma 4.13. *Assume $\delta > C_{p,\mu,K}^{-1}$ and that H_p is C^2 in a neighbourhood of e. For all $0 < \varepsilon < T$, the sequence of processes $\left(Y_{[nt]}^n\right)_{\varepsilon \leq t \leq T}$ is tight in $\mathbb{D}([\varepsilon, T], \mathbb{R}^d)$.*

Proof. Denote by $\left(\tilde{Y}^n_\varepsilon = \left(Y^n_{[nt]}\right)_{\varepsilon\le t\le T}\right)_{n\ge 1}$, the sequence of processes. We prove that from any subsequence $\left(\tilde{Y}^{\phi(n)}_\varepsilon\right)_{n\ge 1}$, we can extract a further subsequence $\left(\tilde{Y}^{\psi(n)}_\varepsilon\right)_{n\ge 1}$ that weakly converges in $\mathbb{D}([\varepsilon,1],\mathbb{R}^d)$.

Let us first prove that $\left(\tilde{Y}^{\phi(n)}_\varepsilon(\varepsilon)\right)_{n\ge 1}$ is bounded in L^2.

$$\left\|\tilde{Y}^{\phi(n)}_\varepsilon(\varepsilon)\right\|_2^2 = \frac{[\phi(n)\varepsilon]^2}{\phi(n)}\mathbb{E}\left[\rho^2(e, X_{[\phi(n)\varepsilon]})\right] \le \varepsilon \sup_{n\ge 1}\left(n\mathbb{E}\left[\rho^2(e,X_n)\right]\right)$$

and the last term is bounded by lemma 4.12.

Consequently $\left(\tilde{Y}^{\phi(n)}_\varepsilon(\varepsilon)\right)_{n\ge 1}$ is tight. So there is a subsequence $\left(\tilde{Y}^{\psi(n)}_\varepsilon(\varepsilon)\right)_{n\ge 1}$ that weakly converges in T_eM to the distribution ν_ε. Thanks to Skorohod theorem which allows to realize it as an a.s. convergence and to lemma 4.11 we can apply Theorem 11.2.3 of [82], and we obtain that the sequence of processes $\left(\tilde{Y}^{\psi(n)}_\varepsilon\right)_{n\ge 1}$ weakly converges to a diffusion $(y_t)_{\varepsilon\le t\le T}$ with generator $G_\delta(t)$ given by (3.4) and such that y_ε has law ν_ε. This achieves the proof of lemma 4.13. □

Proof of Theorem 4.6. Let $\tilde{Y}^n = \left(Y^n_{[nt]}\right)_{0\le t\le T}$. It is sufficient to prove that any subsequence of $\left(\tilde{Y}^n\right)_{n\ge 1}$ has a further subsequence which converges in law to $(y_\delta(t))_{0\le t\le T}$. So let $\left(\tilde{Y}^{\phi(n)}\right)_{n\ge 1}$ a subsequence. By lemma 4.13 with $\varepsilon = 1/m$ there exists a subsequence which converges in law on $[1/m, T]$. Then we extract a sequence indexed by m of subsequence and take the diagonal subsequence $\tilde{Y}^{\eta(n)}$. This subsequence converges in $\mathbb{D}((0,T],\mathbb{R}^d)$ to $(y'(t))_{t\in(0,T]}$. On the other hand, as in the proof of lemma 4.13, we have

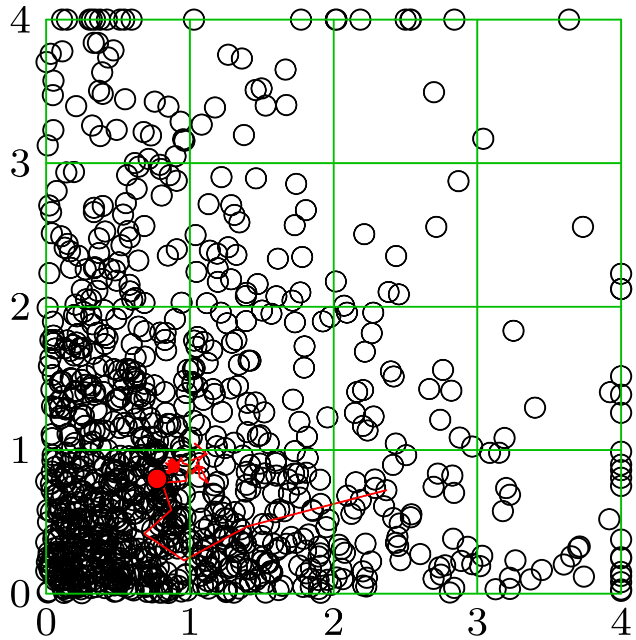
$$\|\tilde{Y}^{\eta(n)}(t)\|_2^2 \le Ct$$

for some $C > 0$. So $\|\tilde{Y}^{\eta(n)}(t)\|_2^2 \to 0$ as $t \to 0$, which in turn implies $\|y'(t)\|_2^2 \to 0$ as $t \to 0$. The unicity statement in Proposition 4.5 implies that $(y'(t))_{t\in(0,T]}$ and $(y_\delta(t))_{t\in(0,T]}$ are equal in law. This achieves the proof. □

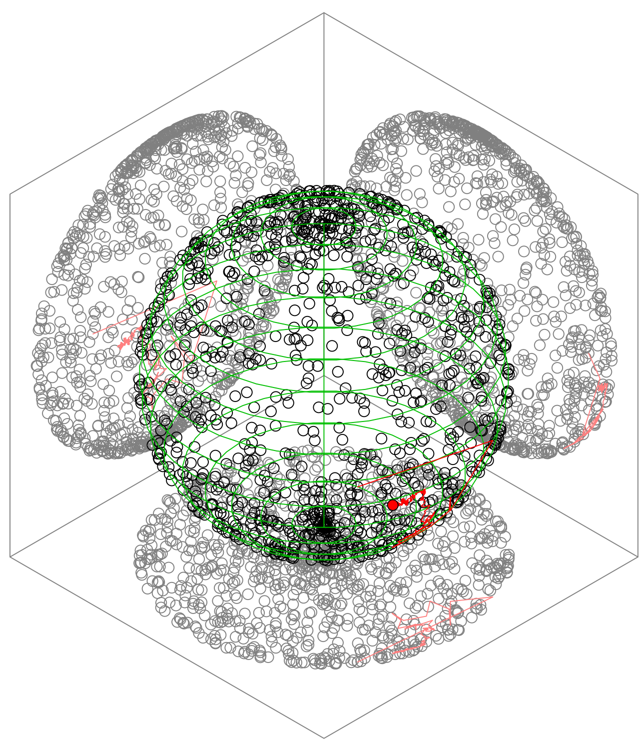
4.4 Simulations of the stochastic algorithms

4.4.1 A non uniform measure on the unit square in the plane

Here M is the Euclidean plane $\mathbb{R}^2$ and μ is the renormalized restriction to the square $[0,4]\times[0,4]$ of an exponential law on $[0,\infty)\times[0,\infty)$. The red path represents one trajectory of the inhomogeneous Markov chain $(X_k)_{k\ge 0}$ corresponding to $p = 1$, with linear interpolation between the different steps. The red point is e_1. Black circles represent the values of $(P_k)_{k\ge 1}$.

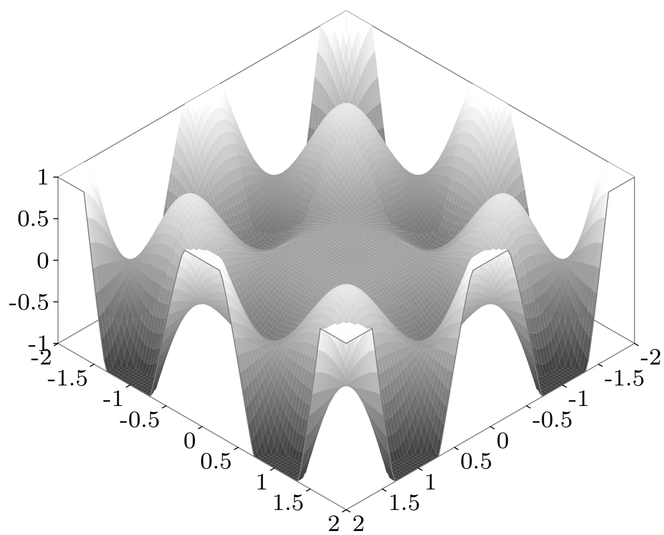


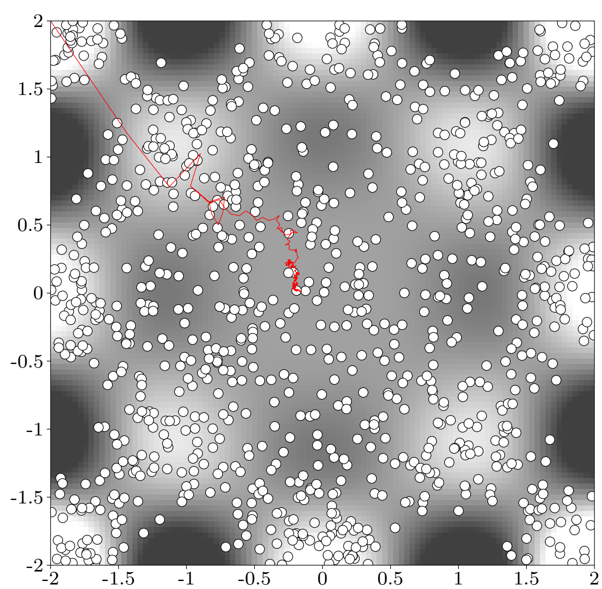
4.4.2 A non uniform measure on the sphere S^2



Here M is the embedded sphere S^2 and μ is a non uniform law. Again the red path represents one trajectory of the inhomogeneous Markov chain $(X_k)_{k\geq 0}$ corresponding to $p=1$, with linear interpolation between the different steps. The red point is e_1. Black circles represent the values of $(P_k)_{k\geq 1}$. One can observe that even if the convexity assumptions are not fulfilled, convergence still holds.

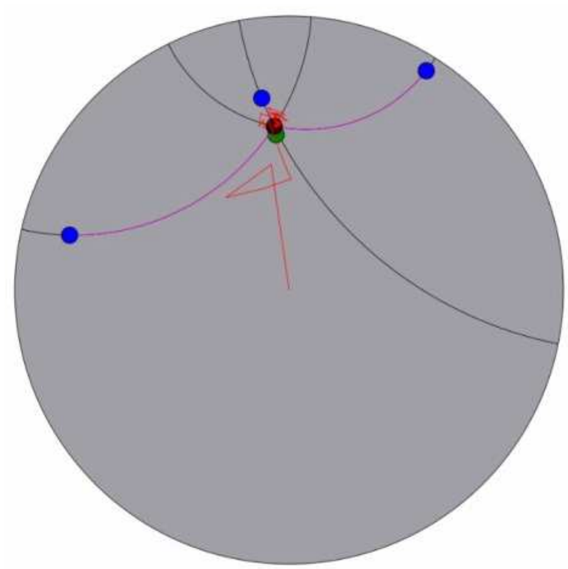
4.4.3 Another non uniform measure on the unit square in the plane

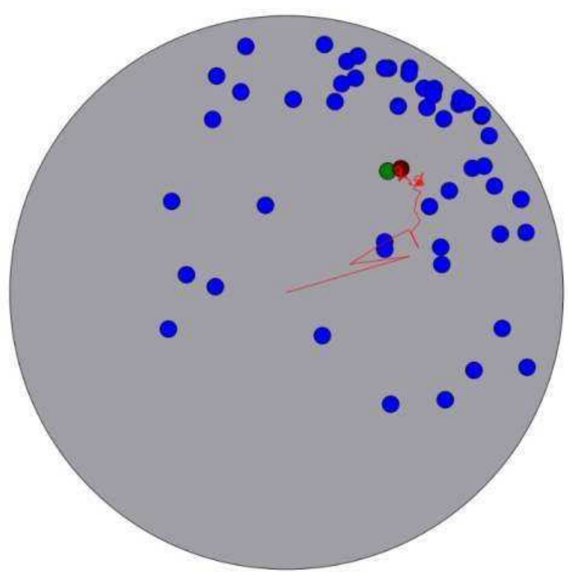




Here μ is given by the density in the top figure. In the figure at the bottom, the red path represents one trajectory of the inhomogeneous Markov chain $(X_k)_{k\geq 0}$ corresponding to $p = 1$, with linear interpolation between the different steps. The red point is e_1. Black circles represent the values of $(P_k)_{k\geq 1}$.

4.4.4 Medians in the Poincaré disc





In the above two figures, M is the Poincaré disc, the blue points are data points and the red path represents one trajectory of the inhomogeneous Markov chain $(X_k)_{k\geq 0}$ corresponding to $p = 1$, with linear interpolation between the different steps. The green points are medians computed by the subgradient method developed in Chapter 2.

4.5 Computing p-means by gradient descent

In chapter 2, a subgradient algorithm has been given to compute medians, so that assume $p > 1$ in this section. We begin with some notations.

Notation 4.14. Let $x \in \bar{B}(a, r)$, we write

$$\gamma_{x,p}(t) = \exp_x(-t \operatorname{grad}_x H_p), \quad t \geq 0;$$

and

$$t_{x,p} = \sup\{t \in [0, 2r] : \gamma_{x,p}(t) \in \bar{B}(a, r)\}.$$

The main result of this section is the following theorem.

Theorem 4.15. *Let $x_0 \in \bar{B}(a, r)$ and for $k \geq 0$ define*

$$x_{k+1} = \exp_{x_k}(-t_k \operatorname{grad}_{x_k} H_p),$$

where $t_k \in (0, t_{x_k,p}]$ such that

$$\lim_{k\to\infty} t_k = 0 \quad \text{and} \quad \sum_{k=0}^{\infty} t_k = +\infty.$$

Then $x_k \longrightarrow e_p$, when $k \longrightarrow \infty$.

Remark 4.16. One should be aware that since $t_k \in (0, t_{x_k,p}]$, the sequence $(x_k)_k$ are always contained in the convex ball $\bar{B}(a, \rho)$.

Remark 4.17. The stepsizes $(t_k)_k$ in the theorem always exist because $\inf\{t_{x,p} : x \in \bar{B}(a, r)\} > 0$ for fixed $p > 1$. To see this, it suffices to note that $t_{x,p}$ is continuous in x and $t_{x,p} > 0$ for $x \in \partial\bar{B}(a, r)$.

Proof. We firstly consider the case when $1 < p < 2$. Observe that for $x \in \bar{B}(a, \rho)$ and $t \in [0, 2r]$,

$$\|\dot{\gamma}_{x,p}(t)\| = \|\operatorname{grad}_x H_p\| = \left\| \int_M -p\rho^{p-1}(x, y) n(x, y) \mu(dy) \right\| \leq p(2r)^{p-1}.$$

It follows that, as in the proof of Theorem 4.3,

$$\frac{d^2}{dt^2}\left[\frac{1}{2}\rho^2(\gamma_{x,p}(t), e_p)\right] \leq p^2(2r)^{2p-1}\beta \coth(2\beta r) := C(\beta, r, p).$$

Hence by Taylor's formula,

$$\begin{aligned}
&\frac{1}{2}\rho^2(x_{k+1}, e_p)\\
=&\frac{1}{2}\rho^2(\gamma_{x_k,p}(0), e_p) + \frac{d}{dt}\left[\frac{1}{2}\rho^2(\gamma_{x_k,p}(t), e_p)\right]_{t=0} t_k + \frac{1}{2}\frac{d^2}{dt^2}\left[\frac{1}{2}\rho^2(\gamma_{x_k,p}(t), e_p)\right]_{t=\xi} t_k^2\\
\le&\frac{1}{2}\rho^2(x_k, e_p) + \left\langle \dot\gamma_{x_k,p}(0)\,,\ \mathrm{grad}_{x_k}\frac{1}{2}\rho^2(\cdot\,, e_p)\right\rangle t_k + \frac{C(\beta,r,p)}{2}t_k^2\\
\le&\frac{1}{2}\rho^2(x_k, e_p) + \langle\, \mathrm{grad}_{x_k} H_p\,,\ \exp_{x_k}^{-1} e_p\,\rangle t_k + \frac{C(\beta,r,p)}{2}t_k^2.
\end{aligned}$$

By the convexity of $H_p(x)$ and Proposition 4.2,

$$\langle\, \mathrm{grad}_{x_k} H_p\,,\ \exp_{x_k}^{-1} e_p\,\rangle \le H_p(e_p) - H_p(x_k) \le -\frac{C_{p,\mu,K}}{2}\rho^2(x_k, e_p).$$

Hence we get

$$\rho^2(x_{k+1}, e_p) \le (1 - C_{p,\mu,K} t_k)\rho^2(x_k, e_p) + C(\beta, r, p)t_k^2. \tag{5.36}$$

Now it suffices to use the method in the proof of Proposition 2.27 to get $\rho^2(x_k, e_p) \longrightarrow 0$.

For the case when $p \ge 2$, similarly we have

$$\frac{d^2}{dt^2}H_p(\gamma_{x,p}(t)) \le p^3(2r)^{3p-4}(2\beta r\coth(2\beta r) + p - 2) =: 2C(\beta, r, p).$$

Then by Taylor's formula and Proposition 4.2,

$$\begin{aligned}
H_p(x_{k+1}) &= H_p(\gamma_{x_k,p}(t_k))\\
&= H_p(\gamma_{x_k,p}(0)) + \frac{d}{dt}\bigg|_{t=0} H_p(\gamma_{x_k,p}(t))t_k + \frac{1}{2}\frac{d^2}{dt^2}\bigg|_{t=\xi} H_p(\gamma_{x_k,p}(t))t_k^2\\
&\le H_p(x_k) + \langle\, \dot\gamma_{x_k,p}(0)\,, \mathrm{grad}_{x_k} H_p\,\rangle t_k + C(\beta, r, p)t_k^2\\
&= H_p(x_k) - \|\,\mathrm{grad}_{x_k} H_p\|^2 t_k + C(\beta, r, p)t_k^2\\
&= H_p(x_k) - C_{p,\mu,K}(H_p(x_k) - H_p(e_p))t_k. + C(\beta, r, p)t_k^2.
\end{aligned}$$

Hence we get

$$H_p(x_{k+1}) - H_p(e_p) \le (1 - C_{p,\mu,K}t_k)(H_p(x_k) - H_p(e_p)) + C(\beta, r, p)t_k^2. \tag{5.37}$$

As in the case when $1 < p < 2$, the method of proving Proposition 2.27 yields $H_p(x_k) \longrightarrow H_p(e_p)$. This completes the proof of the theorem. $\square$

The question of the choice of stepsizes in Theorem 4.15 is answered by the following proposition.

Proposition 4.18. *For every $x \in \bar{B}(a,r)$ the following estimation holds:*

$$t_{x,p} \geq \frac{p\varepsilon^{p+1}}{\pi C(\beta,r,p) + p\varepsilon^p},$$

where $C(\beta,r,p) = p^2(2r)^{2p-1}\beta\coth(2\beta r)$.

Proof. Let $\varepsilon < \pi/(2\alpha)$ in (2.8) and $x \in \bar{B}(a,r) \setminus \bar{B}(a,r-\varepsilon)$. Observe that Lemma 2.11 still holds even if $p > 2$, hence the same method as that in the proof of Lemma 2.24 and the inequalities: $\sin x \leq x$ for $x \geq 0$, $\sin x \geq 2x/\pi$ for $x \in [0,\pi/2]$ and $\rho(x,a) \geq r - \varepsilon$, yield

$$t_{x,p} \geq \frac{p\varepsilon^p \rho(x,a)}{C(\beta,r,p)} \frac{\sin(\alpha(\rho(x,a) - (r-\varepsilon)))}{\sin(\alpha(\rho(x,a) + (r-\varepsilon)))} \geq \frac{p\varepsilon^p}{\pi C(\beta,r,p)}(\rho(x,a) - (r-\varepsilon)).$$

On the other hand, $t_{x,p} \geq r - \rho(x,a)$, it follows that

$$t_{x,p} \geq \max\left\{ r - \rho(x,a), \frac{p\varepsilon^p}{\pi C(\beta,r,p)}(\rho(x,a) - (r-\varepsilon)) \right\}.$$

Then the method in the proof of Lemma 2.25 yields the desired result. □

Now we get a practically useful version of Theorem 4.15:

Theorem 4.19. *Let $(a_k)_k$ be a sequence in $(0,1]$ such that*

$$\lim_{k\to\infty} a_k = 0 \quad \textit{and} \quad \sum_{k=0}^{\infty} a_k = +\infty.$$

Then we can choose

$$t_k = \frac{p\varepsilon^{p+1}a_k}{\pi p^2(2r)^{2p-1}\beta\coth(2\beta r) + p\varepsilon^p}$$

in Theorem 4.15 and the corresponding sequence $(x_k)_k$ converges to e_p.

The following proposition gives error estimations of the gradient descent algorithms.

Proposition 4.20. *Assume that $t_k < C_{p,\mu,K}^{-1}$ for every k in Theorem 4.15, then the following error estimations hold:*
i) if $1 < p < 2$, then for $k \geq 1$,

$$\begin{aligned}\rho^2(x_k,e_p) \leq{}& 4r^2 \prod_{i=0}^{k-1}(1 - C_{p,\mu,K}t_i) \\ &+ C(\beta,r,p)\left(\sum_{j=1}^{k-1} t_{j-1}^2 \prod_{i=j}^{k-1}(1 - C_{p,\mu,K}t_i) + t_{k-1}^2\right) := b_k;\end{aligned}$$

ii) if $p \geq 2$, then for $k \geq 1$,

$$H_p(x_k) - H_p(e_p) \leq (2r)^p \prod_{i=0}^{k-1}(1 - C_{p,\mu,K}t_i) + C(\beta, r, p)\left(\sum_{j=1}^{k-1} t_{j-1}^2 \prod_{i=j}^{k-1}(1 - C_{p,\mu,K}t_i) + t_{k-1}^2\right) := c_k,$$

where the constant

$$C(\beta, r, p) = \begin{cases} p^2(2r)^{2p-1}\beta\coth(2\beta r), & \text{if } \ 1 < p < 2; \\ p^3(2r)^{3p-4}\left(2\beta r\coth(2\beta r) + p - 2\right), & \text{if } \ p \geq 2. \end{cases}$$

Moreover, the sequences $(b_k)_k$ and $(c_k)_k$ both tend to zero.

Proof. It suffices to use (5.36), (5.37) and the method in the proof of Proposition 2.27. ∎

We end this chapter by showing that if $(t_k)_k$ is chosen to be the harmonic series, then the rate of convergence of the gradient descent algorithms are sublinear.

Proposition 4.21. *If we choose $t_k = t/(k+1)$ with some constant $t > 0$ in Theorem 4.15, then the following estimations hold:*
i) if $1 < p < 2$, then for $k \geq 1$,

$$\rho^2(x_k, e_p) \leq \begin{cases} \dfrac{1}{(k+1)^\theta}\left(4r^2 + \dfrac{2^\theta(2-\theta)C(\beta,r,p)t^2}{1-\theta}\right), & \text{if } \ 0 < \theta < 1, \\[2ex] \dfrac{1+\ln k}{k}\, C(\beta,r,p)t^2, & \text{if } \ \theta = 1, \\[2ex] \dfrac{1}{(\theta-1)(k+1)}\left(C(\beta,r,p)t^2 + \dfrac{4r^2(\theta-1) - C(\beta,r,p)t^2}{(k+1)^{\theta-1}}\right), & \text{if } \ \theta > 1, \end{cases}$$

ii) if $p \geq 2$, then for $k \geq 1$,

$$H_p(x_k) - H_p(e_p) \leq \begin{cases} \dfrac{1}{(k+1)^\theta}\left((2r)^p + \dfrac{2^\theta(2-\theta)C(\beta,r,p)t^2}{1-\theta}\right), & \text{if } \ 0 < \theta < 1, \\[2ex] \dfrac{1+\ln k}{k}\, C(\beta,r,p)t^2, & \text{if } \ \theta = 1, \\[2ex] \dfrac{1}{(\theta-1)(k+1)}\left(C(\beta,r,p)t^2 + \dfrac{(2r)^p(\theta-1) - C(\beta,r,p)t^2}{(k+1)^{\theta-1}}\right), & \text{if } \ \theta > 1, \end{cases}$$

where $\theta = C_{p,\mu,K}t$.

Proof. It suffices to use (5.36), (5.37) and Lemma 2.28. ∎

Chapter 5

Riemannian medians as solutions to fixed point problems

Abstract

We show that, under some conditions, Riemannian medians can be interpreted as solutions to fixed point problems. It is also shown that the associated iterated sequences converge to the medians. The main results of this chapter generalize those of [87], in which all the results are proved in Euclidean spaces.

5.1 Introduction

The framework of this chapter is almost the same as that of Chapter 2, except that we do not need lower bounds of sectional curvatures to obtain the main results. For the reasons of integrality and convenience, we recall this framework in the paragraph below.

Let M be a complete Riemannian manifold with Riemannian metric $\langle\, ,\, \rangle$ and Riemannian distance d. We fix an open geodesic ball

$$B(a,\rho) = \{x \in M : d(x,a) < \rho\}$$

in M centered at a with a finite radius ρ. Let Δ be an upper bound of sectional curvatures K in $\bar{B}(a,\rho)$. The injectivity radius of $\bar{B}(a,\rho)$ is denoted by $\mathrm{inj}\,(\bar{B}(a,\rho))$. Furthermore, we assume that the radius of the ball verifies

$$\rho < \min\Big\{ \frac{\pi}{4\sqrt{\Delta}}\,, \frac{\mathrm{inj}\,(\bar{B}(a,\rho))}{2} \Big\}, \tag{1.1}$$

where if $\Delta \leq 0$, then $\pi/(4\sqrt{\Delta})$ is interpreted as $+\infty$. Hence $\bar{B}(a,\rho)$ is convex. Moreover, the geodesics in $\bar{B}(a,\rho)$ vary smoothly with their endpoints.

In view of the importance of some properties of Jacobi fields in the following, we give them here.

Let $I \subset \mathbf{R}$ be an interval, $\gamma : I \longrightarrow \bar{B}(a,\rho)$ be a geodesic and $p \in \bar{B}(a,\rho)$. Consider a family of geodesics:

$$c_p(s,t) = \exp_p(s \exp_p^{-1} \gamma(t)), \quad (s,t) \in [0,1] \times I$$

The two partial derivatives of $c_p(s,t)$ are denoted by

$$c_p'(s,t) = \frac{d}{ds} c_p(s,t) \quad \text{and} \quad \dot{c}_p(s,t) = \frac{d}{dt} c_p(s,t).$$

Observe that for $t \in I$ fixed, the vector field $s \longmapsto c_p'(s,t)$ is the speed of the geodesic $s \longmapsto c_p(s,t)$ from p to $\gamma(t)$, with $c_p'(1,t) = -\exp_{\gamma(t)}^{-1} p$ and that $J_p(s) = \dot{c}_p(s,t)$ is a Jacobi field along the same geodesic with $J_p(0) = 0$, $J_p(1) = \dot{\gamma}(t)$. Moreover, it is well known that

$$\begin{aligned} \frac{d}{dt}\left(\frac{1}{2} d^2(\gamma(t),p)\right) &= \langle\, \dot{\gamma}(t),\, -\exp_{\gamma(t)}^{-1} p \,\rangle, \\ \frac{d^2}{dt^2}\left(\frac{1}{2} d^2(\gamma(t),p)\right) &= \langle\, J_p(1),\, J_p'(1) \,\rangle. \end{aligned}$$

If $\gamma(t) \neq p$, then we also have

$$\begin{aligned} \frac{d}{dt} d(\gamma(t),p) &= \frac{\langle\, \dot{\gamma}(t),\, -\exp_{\gamma(t)}^{-1} p \,\rangle}{d(\gamma(t),p)}, \\ \frac{d^2}{dt^2} d(\gamma(t),p) &= \frac{\langle\, J_p^{\mathrm{nor}}(1),\, J_p'^{\,\mathrm{nor}}(1) \,\rangle}{d(\gamma(t),p)}, \end{aligned}$$

where $J_p^{\mathrm{nor}}(1)$ and $J_p'^{\,\mathrm{nor}}(1)$ are the normal components of $J_p(1)$ and $J_p'(1)$ with respect to the geodesic $s \longrightarrow c_p(s,t)$ at $s = 1$.

With the above notations, we have the following estimations, which can be found in [53]. Remember that, as in Chapter 2, for $t, \kappa \in \mathbf{R}$,

$$S_\kappa(t) = \begin{cases} \sin(\sqrt{\kappa}\, t)/\sqrt{\kappa}, & \text{if } \kappa > 0; \\ t, & \text{if } \kappa = 0; \\ \sinh(\sqrt{-\kappa}\, t)/\sqrt{-\kappa}, & \text{if } \kappa < 0. \end{cases}$$

Proposition 5.1. *The following estimations hold:*

$$\langle\, J_p(1),\, J_p'(1) \,\rangle \geq C(\rho,\Delta) |J_p(1)|^2,$$

$$\langle\, J_p^{\mathrm{nor}}(1),\, J_p'^{\,\mathrm{nor}}(1) \,\rangle \geq C(\rho,\Delta) |J_p^{\mathrm{nor}}(1)|^2,$$

where the constant

$$C(\rho,\Delta) = \min\left(1, 2\rho \frac{S_\Delta'(2\rho)}{S_\Delta(2\rho)}\right) > 0.$$

5.2 Riemannian medians as fixed points

As in Chapter 2, let μ be a probability measure whose support is contained in the open ball $B(a,\rho)$. Define

$$f: \qquad \bar{B}(a,\rho) \longrightarrow \mathbf{R}_+\,, \qquad x \longmapsto \int_M d(x,p)\mu(dp).$$

The set of all the medians of μ, or equivalently, the set of all the minimum points of f, is denoted by $\mathfrak{M}_\mu$. Proposition 2.7 yields that $\mathfrak{M}_\mu$ is contained in the open ball $B(a,\rho)$. Moreover, for every $x \in B(a,\rho)$, we write

$$H(x) = \int_{M\setminus\{x\}} \frac{-\exp_x^{-1} p}{d(x,p)}\mu(dp).$$

In what follows, we shall always assume that the following assumption is satisfied.

Assumption 5.2. The probability measure μ verifies that for every x in the support of μ,

$$\int_{M\setminus\{x\}} \frac{1}{d(x,p)}\mu(dp) < \infty. \tag{2.2}$$

Note that, if x is not in the support of μ, then (2.2) is trivially true.

The following type of functions are important in the sequel. For every x in the open ball $B(a,\rho)$, we define

$$h_x : \bar{B}(a,\rho) \longrightarrow \mathbf{R}_+, \qquad y \longmapsto \frac{1}{2}\int_{M\setminus\{x\}} \frac{d^2(y,p)}{d(x,p)}\mu(dp) + \mu\{x\}d(y,x).$$

Observe that h_x is continuous, strictly convex and for $y \neq x$,

$$\mathrm{grad} h_x(y) = \int_{M\setminus\{x\}} \frac{-\exp_y^{-1} p}{d(x,p)}\mu(dp) + \mu\{x\}\frac{-\exp_y^{-1} x}{d(y,x)}.$$

Proposition 5.3. *For every $x \in B(a,\rho)$, h_x has a unique minimum point, which is denoted by $T(x)$.*

Proof. Let $T(x)$ be a minimum point of h_x. If $T(x) = x$, then one trivially has $T(x) \in B(a,\rho)$. If $T(x) \neq x$, then $\mathrm{grad} h_x(y)$ is pointing outside of the ball $B(a,\rho)$ for $y \in \partial B(a,\rho)$, hence we have necessarily $T(x) \in B(a,\rho)$. Uniqueness is trivial since h_x is strictly convex.

□

The following lemma gives the right derivatives of h_x along geodesics.

Lemma 5.4. *Let $x \in B(a,\rho)$ and $\gamma : [0,1] \to B(a,\rho)$ a geodesic such that $\gamma(0) = x$, then we have*

$$\frac{d}{dt}h_x(\gamma(t))\Big|_{t=0+} = \langle\, \dot\gamma(0),\, H(x)\,\rangle + \mu\{x\}|\dot\gamma|.$$

Proof. It suffices to note that,

$$\operatorname{grad}_y\left(\frac{1}{2}\int_{M\setminus\{x\}}\frac{d^2(y,p)}{d(x,p)}\mu(dp)\right)\bigg|_{y=x}=H(x).$$

□

The following theorem is the main result of this section, which says that the medians of μ coincide with the fixed points of the mapping T. This generalizes a similar result in [87] from Euclidean spaces to Riemannian manifolds.

Theorem 5.5. *The medians of μ are characterized by*

$$\mathfrak{M}_\mu=\{x\in B(a,\rho):\ T(x)=x\}.$$

Proof. ($\subset$) Let $x\in\mathfrak{M}_\mu$, $y\in\bar{B}(a,\rho)$ and $\gamma(t):[0,1]\to\bar{B}(a,\rho)$ be a geodesic such that $\gamma(0)=x$ and $\gamma(1)=y$. Then by Theorem 2.5, Lemma 5.4 and Cauchy-Schwartz inequality,

$$\frac{d}{dt}h_x(\gamma(t))\bigg|_{t=0+}=\langle\,\dot\gamma(0),\,H(x)\,\rangle+\mu\{x\}|\dot\gamma|\geq|\dot\gamma|(-|H(x)|+\mu\{x\})\geq 0.$$

Since $h_x\circ\gamma$ is convex, we have

$$h_x(y)-h_x(x)\geq\frac{d}{dt}h_x(\gamma(t))\bigg|_{t=0+}\geq 0,$$

this means that $x=T(x)$.
($\supset$) If $H(x)=0$, then Theorem 2.5 yields that $x\in\mathfrak{M}_\mu$. Now assume that $H(x)\neq 0$. Choose a geodesic

$$\gamma(t)=\exp_x\left(-t\frac{H(x)}{|H(x)|}\right),\quad t\in[0,\varepsilon].$$

By the definition of $T(x)$, $t=0$ is the minimum point of $h_x\circ\gamma$, hence

$$\frac{d}{dt}h_x(\gamma(t))\bigg|_{t=0+}\geq 0.$$

On the other hand, by Lemma 5.4,

$$\frac{d}{dt}h_x(\gamma(t))\bigg|_{t=0+}=\left\langle-\frac{H(x)}{|H(x)|},\,H(x)\right\rangle+\mu\{x\}=-|H(x)|+\mu\{x\}.$$

Hence we get $|H(x)|\leq\mu\{x\}$, that is to say, $x\in\mathfrak{M}_\mu$.

□

5.3 Approximating Riemannian medians by iteration

The following proposition says that T diminishes the value of f. This is the reason why we add the penalty term $\mu\{x\}d(y,x)$ in the definition of h_x.

Proposition 5.6. *Let $x \in B(a,\rho)$, then*

$$h_x(T(x)) \geq f(T(x)) - \frac{1}{2}f(x).$$

Particularly, if $T(x) \neq x$, then

$$f(T(x)) < f(x).$$

Proof. To show the first inequality,

$$\begin{aligned}
h_x(y) &= \frac{1}{2}\int_{M\setminus\{x\}} \frac{d^2(y,p)}{d(x,p)}\mu(dp) + \mu\{x\}d(y,x) \\
&= \frac{1}{2}\int_{M\setminus\{x\}} \frac{(d(y,p)-d(x,p)+d(x,p))^2}{d(x,p)}\mu(dp) + \mu\{x\}d(y,x) \\
&= \frac{1}{2}\int_{M\setminus\{x\}} \frac{(d(y,p)-d(x,p))^2}{d(x,p)}\mu(dp) + f(y) - f(x) + \frac{1}{2}f(x) \\
&= \frac{1}{2}\int_{M\setminus\{x\}} \frac{(d(y,p)-d(x,p))^2}{d(x,p)}\mu(dp) + f(y) - \frac{1}{2}f(x) \\
&\geq f(y) - \frac{1}{2}f(x).
\end{aligned}$$

Taking $y = T(x)$, we get the first inequality. Now assume that $x \neq T(x)$. To show the second one, it suffices to note that Proposition 5.3 yields $h_x(T(x)) < h_x(x) = \frac{1}{2}f(x)$. The proof is complete. □

Here is a useful observation: if $T(x) \neq x$, then

$$\mathrm{grad} h_x(T(x)) = \int_{M\setminus\{x\}} \frac{-\exp_{T(x)}^{-1} p}{d(x,p)}\mu(dp) + \mu\{x\}\frac{-\exp_{T(x)}^{-1} x}{d(T(x),x)} = 0$$

We need the following estimation on $d(y,T(x))$ for x and y in $B(a,\rho)$.

Proposition 5.7. *Let $x, y \in B(a,\rho)$ with $T(x), y \neq x$. Then the following estimation holds:*

$$d(y,T(x)) \leq \frac{1}{C(\rho,\Delta)}\frac{|\,\mathrm{grad}\ h_x(y)|}{\int_{M\setminus\{x\}}\frac{1}{d(x,p)}\mu(dp)}.$$

Proof. Let $\gamma : [0,1] \to B(a,\rho)$ be a geodesic such that$\gamma(0) = T(x)$ and $\gamma(1) = y$. If $x \notin \mathrm{Im}\gamma$, consider $\mathrm{grad} h_x$ along γ:

$$V(t) = \int_{M\backslash\{x\}} \frac{-\exp_{\gamma(t)}^{-1} p}{d(x,p)} \mu(dp) + \mu\{x\} \frac{-\exp_{\gamma(t)}^{-1} x}{d(\gamma(t),x)}.$$

Since $V(0) = \mathrm{grad} h_x(T(x)) = 0$,

$$\langle\, V(1), \dot\gamma(1)\,\rangle = \int_0^1 \frac{d}{dt} \langle\, V(t), \dot\gamma(t)\,\rangle dt.$$

By Proposition 5.1,

$$\begin{aligned}
\int_0^1 \frac{d}{dt} \langle\, V(t), \dot\gamma(t)\,\rangle dt &= \int_0^1 \left\langle \frac{DV}{dt}, \dot\gamma(t) \right\rangle dt \\
&= \int_{M\backslash\{x\}} \frac{\langle\, J_p'(1), J_p(1)\,\rangle}{d(x,p)} \mu(dp) + \mu\{x\} \frac{\langle\, J_x'^{\mathrm{nor}}(1), J_x^{\mathrm{nor}}(1)\,\rangle}{d(\gamma(t),x)} \\
&\geq C(\rho,\Delta) d^2(y, T(x)) \int_{M\backslash\{x\}} \frac{1}{d(x,p)} \mu(dp).
\end{aligned}$$

Note that $V(1) = \mathrm{grad} h_x(y)$, then

$$\langle\, V(1), \dot\gamma(1)\,\rangle \leq |V(1)|\, |\dot\gamma(1)| = |\mathrm{grad} h_x(y)| d(y, T(x)).$$

These two inequalities give the result.

If $x \in \mathrm{Im}\gamma$, assume that $x = \gamma(t_0)$ with $t_0 \in (0,1)$, consider the following two vector fields:

$$V_1(t) = \int_{M\backslash\{x\}} \frac{-\exp_{\gamma(t)}^{-1} p}{d(x,p)} \mu(dp) - \mu\{x\} \frac{\dot\gamma(t)}{d(y, T(x))} \qquad t \in [0, t_0],$$

$$V_2(t) = \int_{M\backslash\{x\}} \frac{-\exp_{\gamma(t)}^{-1} p}{d(x,p)} \mu(dp) + \mu\{x\} \frac{\dot\gamma(t)}{d(y, T(x))} \qquad t \in [t_0, 1].$$

Since $V_1(0) = \mathrm{grad} h_x(T(x)) = 0$,

$$\begin{aligned}
\langle\, V_1(t_0), \dot\gamma(t_0)\,\rangle &= \int_0^{t_0} \frac{d}{dt} \langle\, V_1(t), \dot\gamma(t)\,\rangle dt \\
&= \int_0^{t_0} \left\langle \frac{DV_1}{dt}, \dot\gamma(t) \right\rangle dt \\
&= \int_0^{t_0} dt \int_{M\backslash\{x\}} \frac{\langle\, J_p'(1), J_p(1)\,\rangle}{d(x,p)} \mu(dp).
\end{aligned}$$

Moreover, one also has

$$\begin{aligned}\langle V_2(1), \dot\gamma(1) \rangle &= \langle V_2(t_0), \dot\gamma(t_0) \rangle + \int_{t_0}^1 \frac{d}{dt}\langle V_2(t), \dot\gamma(t) \rangle dt \\ &= \langle V_2(t_0), \dot\gamma(t_0) \rangle + \int_{t_0}^1 dt \int_{M\setminus\{x\}} \frac{\langle J_p'(1), J_p(1) \rangle}{d(x,p)} \mu(dp).\end{aligned}$$

Summing up the above two identities we get

$$\langle V_2(1), \dot\gamma(1) \rangle = 2\mu\{x\} d(y, T(x)) + \int_0^1 dt \int_{M\setminus\{x\}} \frac{\langle J_p'(1), J_p(1) \rangle}{d(x,p)} \mu(dp).$$

By Cauchy-Schwartz inequality and Proposition 5.1,

$$\langle V_2(1), \dot\gamma(1) \rangle = \langle \mathrm{grad} h_x(y), \dot\gamma(1) \rangle \le |\mathrm{grad} h_x(y)| d(y, T(x)),$$
$$\int_{M\setminus\{x\}} \frac{\langle J_p'(1), J_p(1) \rangle}{d(x,p)} \mu(dp) \ge C(\rho,\Delta) d^2(y, T(x)) \int_{M\setminus\{x\}} \frac{1}{d(x,p)} \mu(dp).$$

Since $2\mu\{x\} d(y, T(x)) \ge 0$, the proof is complete. □

We also need the following estimation.

Proposition 5.8. *Let $x \in B(a,\rho)$, then*

$$\frac{1}{2} d^2(x, T(x)) \le \frac{1}{C(\rho,\Delta)} \frac{f(x) - f(T(x))}{\int_{M\setminus\{x\}} \frac{1}{d(x,p)} \mu(dp)}.$$

Proof. Let $\gamma : [0,1] \to M$ be a geodesic such that $\gamma(0) = x$ and $\gamma(1) = T(x)$, then one has

$$h_x(\gamma(t)) = \frac{1}{2} \int_{M\setminus\{x\}} \frac{d^2(\gamma(t), p)}{d(x,p)} \mu(dp) + \mu\{x\} d(x, T(x)) t.$$

By Proposition 5.1,

$$\begin{aligned}\frac{d^2}{dt^2} h_x(\gamma(t)) &= \int_{M\setminus\{x\}} \frac{\langle J_p(1),\ J_p'(1) \rangle}{d(x,p)} \mu(dp) \\ &\ge C(\rho,\Delta) d^2(x, T(x)) \int_{M\setminus\{x\}} \frac{1}{d(x,p)} \mu(dp).\end{aligned}$$

On the other hand, since $T(x)$ is the minimum point of h_x, the second order Taylor's formula gives

$$h_x(\gamma(0)) = h_x(\gamma(1)) + \frac{1}{2} \frac{d^2}{dt^2} h_x(\gamma(t))\Big|_{t=\xi}, \quad \xi \in (0,1).$$

Hence

$$\begin{aligned}\frac{1}{2}f(x) - h_x(T(x)) &= \frac{1}{2}\frac{d^2}{dt^2}h_x(\gamma(t))\big|_{t=\xi}\\ &\geq \frac{1}{2}C(\rho,\Delta)d^2(x,T(x))\int_{M\setminus\{x\}}\frac{1}{d(x,p)}\mu(dp).\end{aligned}$$

Now it suffices to use Proposition 5.6 to get

$$\frac{1}{2}f(x) - h_x(T(x)) \leq f(x) - f(T(x)).$$

The proof is complete. □

In what follows, we assume further that the following assumption is fulfilled.

Assumption 5.9. The probability measure μ verifies:

1) μ is not a Dirac measure;

2) μ has a unique median m;

3) for every convergent sequence $(y_n)_n$ in $B(a,\rho)$,

$$\lim_{\mu A\to 0}\limsup_{n\to\infty}\int_{A\setminus\{y_n\}}\frac{1}{d(y_n,p)}\mu(dp) = 0;$$

4) the atoms of μ are isolated.

Remark 5.10. It is easily seen that if $N \geq 3$, $\sum_{k=1}^N \omega_k = 1$, $\omega_k > 0$ and $p_1,\ldots,p_N$ are distinct points in $B(a,\rho)$ which are not contained in a single geodesic, then the probability measure

$$\mu = \sum_{k=1}^N \omega_k\delta_{p_k},$$

satisfies Assumption 5.2 and Assumption 5.9.

We fix a point $x_0 \in B(a,\rho)$ and define a sequence $(x_n)_n$ by $x_{n+1} = T(x_n)$, $n \geq 0$. By Lemma 5.3, this sequence is contained in $B(a,\rho)$, thus bounded. We also assume that $x_n \neq m$ for every n. Then by Theorem 5.5 and Proposition 5.6, $(x_n)_n$ are distinct since $(f(x_n))_n$ is strictly decreasing. Observe that for $x \in B(a,\rho)$,

$$\int_{M\setminus\{x\}}\frac{1}{d(x,p)}\mu(dp) \geq L,$$

where $L = (1 - \sup\{\mu\{y\} : y \in B(a,\rho)\})/(2\rho) > 0$.

Lemma 5.11. *For every subsequence $(x_{n_k})_k$ of $(x_n)_n$, one has*

$$d(x_{n_k}, T(x_{n_k})) \longrightarrow 0, \quad \textit{when} \quad k \longrightarrow \infty.$$

Particularly, if $x_{n_k} \longrightarrow x_*$, then $T(x_{n_k}) \longrightarrow x_*$.

Proof. By Proposition 5.8,

$$d^2(x_{n_k}, T(x_{n_k})) \leq \frac{2(f(x_{n_k}) - f(T(x_{n_k})))}{LC(\rho, \Delta)} = \frac{2(f(x_{n_k}) - f(x_{n_k+1})}{LC(\rho, \Delta)} \longrightarrow 0.$$

□

The following property of the sequence $(x_n)_n$ is of fundamental importance.

Proposition 5.12. *Let $(x_{n_k})_k$ be a subsequence of $(x_n)_n$ such that $x_{n_k} \longrightarrow x_*$, then one has*

$$\lim_{k\to\infty} \frac{d(T(x_{n_k}), x_*)}{d(x_{n_k}, x_*)} \mu\{x_*\} = |H(x_*)|.$$

Proof. Firstly, since $f(x_n)$ is strictly decreasing, $x_{n_k} \neq x_*$ for every k. Hence the expression on the left is well defined. Note that, by Lemma 5.11, $T(x_{n_k}) \longrightarrow x_*$. Since $T(x_{n_k}) \neq x_{n_k}$, we get

$$\mathrm{grad} h_{x_{n_k}}(T(x_{n_k})) = \int_{M\setminus\{x_{n_k}\}} \frac{-\exp_{T(x_{n_k})}^{-1} p}{d(x_{n_k}, p)} \mu(dp) + \mu\{x_{n_k}\} \frac{-\exp_{T(x_{n_k})}^{-1} x_{n_k}}{d(T(x_{n_k}), x_{n_k})} = 0,$$

hence

$$\frac{\exp_{T(x_{n_k})}^{-1} x_*}{d(x_{n_k}, x_*)} \mu\{x_*\} = \int_{M\setminus\{x_{n_k}, x_*\}} \frac{-\exp_{T(x_{n_k})}^{-1} p}{d(x_{n_k}, p)} \mu(dp) + \mu\{x_{n_k}\} \frac{-\exp_{T(x_{n_k})}^{-1} x_{n_k}}{d(T(x_{n_k}), x_{n_k})}.$$

Taking the norm, one gets

$$\frac{d(T(x_{n_k}), x_*)}{d(x_{n_k}, x_*)} \mu\{x_*\} = \left| \int_{M\setminus\{x_{n_k}, x_*\}} \frac{-\exp_{T(x_{n_k})}^{-1} p}{d(x_{n_k}, p)} \mu(dp) + \mu\{x_{n_k}\} \frac{-\exp_{T(x_{n_k})}^{-1} x_{n_k}}{d(T(x_{n_k}), x_{n_k})} \right|.$$

We shall show that

$$\left| \int_{M\setminus\{x_{n_k}, x_*\}} \frac{-\exp_{T(x_{n_k})}^{-1} p}{d(x_{n_k}, p)} \mu(dp) + \mu\{x_{n_k}\} \frac{-\exp_{T(x_{n_k})}^{-1} x_{n_k}}{d(T(x_{n_k}), x_{n_k})} \right| \longrightarrow |H(x_*)|.$$

Note that

$$\left| \mu\{x_{n_k}\} \frac{-\exp_{T(x_{n_k})}^{-1} x_{n_k}}{d(T(x_{n_k}), x_{n_k})} \right| = \mu\{x_{n_k}\} \longrightarrow 0,$$

since the $(x_{n_k})_k$ are distinct and the series $\sum_k \mu\{x_{n_k}\}$ converges. Thus it suffices to show that

$$\left| \int_{M\backslash\{x_{n_k}, x_*\}} \frac{-\exp^{-1}_{T(x_{n_k})} p}{d(x_{n_k}, p)} \mu(dp) \right| \longrightarrow |H(x_*)|$$

To this end, we will show that the limit can be taken into the integral of

$$\int_{M\backslash\{x_{n_k}, x_*\}} \frac{d(T(x_{n_k}), p)}{d(x_{n_k}, p)} \mu(dp).$$

Then it suffices to show that the sequence of functions

$$\left(p \longmapsto \frac{d(T(x_{n_k}), p)}{d(x_{n_k}, p)} \mathbf{1}_{\{p \neq x_{n_k}\}} \right)_k$$

is uniformly integrable. In fact, by Proposition 5.7

$$d(T(x_{n_k}), p) \leq \frac{\left|\mathrm{grad} h_{x_{n_k}}(p)\right|}{C(\rho, \Delta) \int_{M\backslash\{x_{n_k}\}} \frac{1}{d(x_{n_k}, p)} \mu(dp)}.$$

However,

$$\begin{aligned} |\mathrm{grad} h_{x_{n_k}}(p)| &= \left| \int_{M\backslash\{x_{n_k}\}} \frac{-\exp^{-1}_p q}{d(x_{n_k}, q)} \mu(dq) + \mu\{x_{n_k}\} \frac{-\exp^{-1}_p x_{n_k}}{d(x_{n_k}, p)} \right| \\ &\leq \int_{M\backslash\{x_{n_k}\}} \frac{d(p, q)}{d(x_{n_k}, q)} \mu(dq) + \mu\{x_{n_k}\} \\ &\leq \int_{M\backslash\{x_{n_k}\}} \frac{d(p, x_{n_k}) + d(x_{n_k}, q)}{d(x_{n_k}, q)} \mu(dq) + \mu\{x_{n_k}\} \\ &= d(x_{n_k}, p) \int_{M\backslash\{x_{n_k}\}} \frac{1}{d(x_{n_k}, q)} \mu(dq) + 1 \end{aligned}$$

Thus for $p \neq x_{n_k}$,

$$\frac{d(T(x_{n_k}), p)}{d(x_{n_k}, p)} \leq \frac{1}{C(\rho, \Delta)} \left(\frac{1}{d(x_{n_k}, p)} \Big/ \int_{M\backslash\{x_{n_k}\}} \frac{1}{d(x_{n_k}, p)} \mu(dp) + 1 \right).$$

It follows that

$$\int_{M\backslash\{x_{n_k}\}} \frac{d(T(x_{n_k}), p)}{d(x_{n_k}, p)} \mu(dp) \leq \frac{2}{C(\rho, \Delta)}.$$

Hence the sequence is bounded in $L^1(\mu)$. On the other hand,

$$\frac{d(T(x_{n_k}), p)}{d(x_{n_k}, p)} \mathbf{1}_{\{p \neq x_{n_k}\}} \leq \frac{1}{C(\rho, \Delta)} \left(\frac{L^{-1}}{d(x_{n_k}, p)} \mathbf{1}_{\{p \neq x_{n_k}\}} + 1 \right).$$

The condition 3) in Assumption 5.9 implies that the right hand side has equi-absolutely continuous integrals, hence the same holds for the left hand side. The uniform integrability is proved and this completes the proof.

$\square$

The lemma below is a final preparation for the main result of this section.

Lemma 5.13. *Let x_* be an accumulating point of $(x_n)_n$ such that $\mu\{x_*\} > 0$. Then it is the unique accumulating point of $(x_n)_n$. As a result, $x_n \longrightarrow x_*$, when $n \longrightarrow \infty$.*

Proof. By Proposition 5.12 and Theorem 2.5, the accumulating points of $(x_n)_n$ are contained in the set $\{m\} \cup \{x \in B(a,r) : \mu\{x\} > 0\}$. Since the atoms of μ are isolated, there exists $\delta > 0$ such that x_* is the unique accumulating point of $(x_n)_n$ in $B(x_*, \delta)$. If x_* is not the unique accumulating point of $(x_n)_n$, we can choose a subsequence $(x_{n_k})_k$ of $(x_n)_n$ such that $d(x_{n_k}, x_*) < \delta$ and $d(T(x_{n_k}), x_*) > \delta$. Since x_* is the unique accumulating point of $(x_n)_n$ in $B(x_*, \delta)$, one has necessarily $x_{n_k} \to x_*$. It follows that

$$\lim_{k\to\infty} \frac{d(T(x_{n_k}), x_*)}{d(x_{n_k}, x_*)} \geq \lim_{k\to\infty} \frac{\delta}{d(x_{n_k}, x_*)} = \infty.$$

This contradicts Proposition 5.12, the proof is complete. □

Now we are ready to give the main result of this section, which generalizes the convergence result of [87] from Euclidean spaces to Riemannian manifolds. Note that for a bounded sequence $(a_n)_n$ of positive numbers, one always has

$$\liminf_{n\to\infty} \frac{a_{n+1}}{a_n} \leq 1.$$

Theorem 5.14. *Let $x_0 \in B(a, \rho)$, define a sequence $(x_n)_n$ by*

$$x_{n+1} = T(x_n), \quad n \geq 0.$$

Then $x_n \longrightarrow m$.

Proof. If there exists some N such that $x_N = m$, then Theorem 5.5 yields that $x_n = m$ for every $n \geq N$, hence the assertion is true. Now assume that $x_n \neq m$ for all n. Let x_* be an accumulating point of $(x_n)_n$, then there exists a subsequence $(x_{n_k})_k$ of $(x_n)_n$ which converges to x_*. If $\mu\{x_*\} = 0$, by Proposition 5.12, $x_* = m$. If $\mu\{x_*\} > 0$, then by Lemma 5.13, the sequence $(x_n)_n$ converges to x_*. By Proposition 5.12,

$$\frac{|H(x_*)|}{\mu\{x_*\}} = \lim_{n\to\infty} \frac{d(x_{n+1}, x_*)}{d(x_n, x_*)} \leq 1.$$

Then Theorem 2.5 yields that $x_* = m$. Now we have proved that $(x_n)_n$ has a unique accumulating point m, since it is also bounded, it must converges to m. The proof is complete. □

Remark 5.15. Let $N \geq 3$, $\sum_{k=1}^N \omega_k = 1$, $\omega_k > 0$ and $p_1, \ldots, p_N$ are distinct points in $B(a, \rho)$ which are not contained in a single geodesic, then Theorem 5.14 holds for the probability measure

$$\mu = \sum_{k=1}^N \omega_k \delta_{p_k}.$$

Moreover, we can choose the initial point x_0 such that $(x_n)_n$ is in fact a sequence of barycenters. To this end, firstly determine i such that $f(p_i) = \min\{f(p_k) : k = 1, \dots, N\}$ and then compare $|H(p_i)|$ and ω_i. If $|H(p_i)| \leq \omega_i$, then we know that p_i is the median. In the opposite case, one has necessarily $|H(p_i)| > \omega_i$. Choose a geodesic

$$\gamma(t) = \exp_{p_i}\Big(-t\frac{H(p_i)}{|H(p_i)|}\Big), \quad t \geq 0,$$

then by Lemma 2.4,

$$\frac{d}{dt} f(\gamma(t))\Big|_{t=0+} = -|H(p_i)| + \omega_i < 0,$$

hence there exists $\varepsilon > 0$ such that $f(\gamma(\varepsilon)) < f(p_i)$, then let $x_0 = \gamma(\varepsilon)$. Since $(f(x_n))_n$ is strictly decreasing, we have $\mu\{x_n\} = 0$ for all n. As a result,

$$h_{x_n}(y) = \frac{1}{2}\sum_{k=1}^{N} \omega_k \frac{d^2(y, p_k)}{d(x_n, p_k)},$$

thus $T(x_n)$ is the barycenter of the measure

$$\mu_n = \sum_{k=1}^{N} \frac{\omega_k}{d(x_n, p_k)} \delta_{p_k}.$$

For the particular case when M is a Euclidean space, the barycenter of weighted sample points can be expressed explicitly, so that

$$x_{n+1} = \sum_{k=1}^{N} \frac{\omega_k p_k}{\|x_n - p_k\|} \Big/ \sum_{k=1}^{N} \frac{\omega_k}{\|x_n - p_k\|},$$

which is exactly the Weiszfeld algorithm.

5.4 Appendix

In this appendix, we give some supplementary estimates. Let δ be a lower bound of sectional curvatures K in $\bar{B}(a, \rho)$. Firstly, by [53] one has following estimations:

Proposition 5.16.

$$|J_p'(1)| \leq D(\rho, \delta, \Delta)|J_p(1)|,$$

$$\langle\, J_p(1),\, J_p'(1)\, \rangle \leq D(\rho, \delta, \Delta)|J_p(1)|^2,$$

where the constant

$$D(\rho, \delta, \Delta) = \max\Big(1,\, 2\rho\big(\frac{S_\delta'(2\rho)}{S_\delta(2\rho)} - \frac{S_\Delta'(2\rho)}{S_\Delta(2\rho)}\big)\Big) \geq 1.$$

We give some estimations of $d(x,T(x))$.

Proposition 5.17. *Let $x \in B(a,\rho)$ and $T(x) \neq x$, then the following estimations hold:*

$$d(x,T(x)) \leq \frac{1}{C(\rho,\Delta)} \frac{|H(x)| - \mu\{x\}}{\int_{M\setminus\{x\}} \frac{1}{d(x,p)}\mu(dp)},$$

$$d(x,T(x)) \geq \frac{1}{D(\rho,\delta,\Delta)} \frac{|H(x)| - \mu\{x\}}{\int_{M\setminus\{x\}} \frac{1}{d(x,p)}\mu(dp)}.$$

Proof. Let $\gamma : [0,1] \to B(a,\rho)$ be a geodesic such that $\gamma(0) = T(x)$ and $\gamma(1) = x$. Consider the following vector field along γ:

$$V(t) = \int_{M\setminus\{x\}} \frac{-\exp_{\gamma(t)}^{-1} p}{d(x,p)}\mu(dp) - \mu\{x\}\frac{\dot\gamma(t)}{d(x,T(x))},$$

then $V(0) = \mathrm{grad} h_x(T(x)) = 0$, as a result,

$$\langle V(1), \dot\gamma(1) \rangle = \langle V(0), \dot\gamma(0) \rangle + \int_0^1 \frac{d}{dt}\langle V(t), \dot\gamma(t) \rangle dt = \int_0^1 \frac{d}{dt}\langle V(t), \dot\gamma(t) \rangle.$$

On the one had, Cauchy-Schwartz inequality gives

$$\begin{aligned}
\langle V(1), \dot\gamma(1) \rangle &= \langle H(x), \dot\gamma(1) \rangle - \mu\{x\} d(x,T(x)) \\
&\leq |H(x)||\dot\gamma(1)| - \mu\{x\} d(x,T(x)) \\
&= d(x,T(x))(|H(x)| - \mu\{x\}).
\end{aligned}$$

On the other hand, by Proposition 5.1,

$$\begin{aligned}
\int_0^1 \frac{d}{dt}\langle V(t), \dot\gamma(t) \rangle dt &= \int_0^1 \left\langle \frac{DV}{dt}, \dot\gamma(t) \right\rangle dt \\
&= \int_0^1 \left\langle \int_{M\setminus\{x\}} \frac{J_p'(1)}{d(x,p)}\mu(dp), \dot\gamma(t) \right\rangle dt \\
&= \int_0^1 dt \int_{M\setminus\{x\}} \frac{\langle J_p'(1), J_p(1) \rangle}{d(x,p)}\mu(dp) \\
&\geq C(\rho,\Delta) d^2(x,T(x)) \int_{M\setminus\{x\}} \frac{1}{d(x,p)}\mu(dp).
\end{aligned}$$

These two inequalities give the first estimation.

In order to prove the second one, note that for $t \neq 0$, one has $V(t) \neq 0$, hence

$$\frac{d}{dt}|V(t)| \leq \left|\frac{DV}{dt}\right| \quad \text{for} \quad t \in (0,1].$$

By Proposition 5.16,

$$\begin{aligned}
|V(1)| &= \int_0^1 \frac{d}{dt}|V(t)|dt \leq \int_0^1 \left|\frac{DV}{dt}\right| dt \\
&= \int_0^1 \left| \int_{M\setminus\{x\}} \frac{J_p'(1)}{d(x,p)}\mu(dp)\right| dt \\
&\leq \int_0^1 dt \int_{M\setminus\{x\}} \frac{|J_p'(1)|}{d(x,p)}\mu(dp) \\
&\leq \int_0^1 dt \int_{M\setminus\{x\}} \frac{D(\rho,\delta,\Delta)|J_p(1)|}{d(x,p)}\mu(dp) \\
&= D(\rho,\delta,\Delta)\int_{M\setminus\{x\}} \frac{1}{d(x,p)}\mu(dp).
\end{aligned}$$

However,

$$|V(1)| = \left| H(x) + \mu\{x\}\frac{\exp_x^{-1} T(x)}{d(x,T(x))}\right| \geq |H(x)| - \mu\{x\}.$$

These two inequalities complete the proof. ☐

As a complement to Proposition 5.7 we also have:

Proposition 5.18. *Let $x, y \in B(a,\rho)$ with $T(x), y \neq x$. Then the following estimation holds:*

$$d(y,T(x)) \geq \frac{|\operatorname{grad} h_x(y)|}{D(\rho,\delta,\Delta)\int_{M\setminus\{x\}} \frac{1}{d(x,p)}\mu(dp) + (1 + D(\rho,\delta,\Delta))\frac{\mu\{x\}}{d(y,x)}}.$$

Proof. Let $\gamma : [0,1] \to B(a,\rho)$ be a geodesic such that$\gamma(0) = T(x)$ and $\gamma(1) = y$. Consider the following vector field along the geodesic γ:

$$V(t) = \int_{M\setminus\{x\}} \frac{-\exp_{\gamma(t)}^{-1} p}{d(x,p)}\mu(dp) + \mu\{x\}\frac{-\exp_{\gamma(t)}^{-1} x}{d(y,x)}.$$

Observe that the function

$$G(z) = \frac{1}{2}\int_{M\setminus\{x\}} \frac{d^2(z,p)}{d(x,p)}\mu(dp) + \frac{1}{2}\mu\{x\}\frac{d^2(z,x)}{d(y,x)}$$

is strictly convex, there exists a unique point $z_0 \in B(a,\rho)$ such that $\operatorname{grad} G(z_0) = 0$. Note that $V(t) = \operatorname{grad} G(\gamma(t))$, then one gets

$$\frac{d}{dt}|V(t)| \leq \left|\frac{DV}{dt}\right| \quad \text{for a.e.} \quad t \in [0,1].$$

Hence by Proposition 5.16,

$$\begin{aligned}\int_0^1 \frac{d}{dt}|V(t)|dt &\le \int_0^1 \left|\frac{DV}{dt}\right| dt = \int_0^1 \left| \int_{M\setminus\{x\}} \frac{J_p'(1)}{d(x,p)}\mu(dp) + \mu\{x\}\frac{J_x'(1)}{d(y,x)}\right| dt \\ &\le \int_0^1 \left(\int_{M\setminus\{x\}} \frac{|J_p'(1)|}{d(x,p)}\mu(dp) + \mu\{x\}\frac{|J_x'(1)|}{d(y,x)}\right) dt \\ &\le D(\rho,\delta,\Delta)d(y,T(x))\left(\int_{M\setminus\{x\}} \frac{1}{d(x,p)}\mu(dp) + \frac{\mu\{x\}}{d(y,x)}\right).\end{aligned}$$

On the other hand, since $\mathrm{grad} h_x(T(x)) = 0$, one has

$$\begin{aligned}|V(0)| &= \left| \int_{M\setminus\{x\}} \frac{-\exp_{T(x)}^{-1} p}{d(x,p)}\mu(dp) + \mu\{x\}\frac{-\exp_{T(x)}^{-1} x}{d(y,x)}\right| \\ &= \left| \mu\{x\}\frac{\exp_{T(x)}^{-1} x}{d(T(x),x)} + \mu\{x\}\frac{-\exp_{T(x)}^{-1} x}{d(y,x)}\right| \\ &\le \mu\{x\}d(x,T(x))\left| \frac{1}{d(T(x),x)} - \frac{1}{d(y,x)}\right| \\ &\le d(y,T(x))\frac{\mu\{x\}}{d(y,x)}.\end{aligned}$$

The proof will be finished by noting that

$$|\mathrm{grad} h_x(y)| = |V(1)| = |V(0)| + \int_0^1 \frac{d}{dt}|V(t)|dt.$$

□

Chapter 6

Riemannian geometry of Toeplitz covariance matrices and applications to radar target detection

Abstract

In this chapter, we consider the manifold of Toeplitz covariance matrices of order n parameterized by the reflection coefficients which are derived from Levinson's recursion of autoregressive models. The explicit expression of the reparametrization and its inverse are obtained. With the Riemannian metric given by the Hessian of a Kähler potential, we show that the manifold is in fact a Cartan-Hadamard manifold with lower sectional curvature bound -4. After that, we compute the geodesics and use the subgradient algorithm introduced in Chapter 2 to find the median of Toeplitz covariance matrices. Finally, we give some simulated examples to illustrate the application of the median method to radar target detection.

6.1 Reflection coefficients parametrization

Let $n \geq 1$ be a fixed integer and $\mathcal{T}_n$ be the set of Toeplitz Hermitan positive definite matrices of order n, then $\mathcal{T}_n$ is an open submanifold of $\mathbf{R}^{2n-1}$. We fix an element $R_n \in \mathcal{T}_n$ which can be written as

$$R_n = \begin{bmatrix} r_0 & \overline{r}_1 & \dots & \overline{r}_{n-1} \\ r_1 & r_0 & \dots & \overline{r}_{n-2} \\ \vdots & \ddots & \ddots & \vdots \\ r_{n-1} & \dots & r_1 & r_0 \end{bmatrix},$$

then there exists a complex valued second order stationary process X indexed by $\mathbf{Z}_+$ such that for every $0 \le k \le n-1$ we have $r_k = \mathbf{E}[\,X_0 \bar{X}_k\,]$. An estimate of X_l $(l \ge k)$ by linear combination of k most recent pasts is given by

$$\hat{X}_l = -\sum_{j=1}^{k} a_j^{(k)} X_{l-j}\,.$$

The mean squared error of this estimate is denoted by

$$P_k = \mathbf{E}|X_l - \hat{X}_l|^2.$$

The optimal $\hat{X}_l$ is the one that minimizes P_k. In this case the optimal coefficients $a_1^{(k)}, \dots, a_k^{(k)}$ and the mean squared error P_k verify the following normal equation

$$\begin{bmatrix} r_0 & \overline{r}_1 & \dots & \overline{r}_k \\ r_1 & r_0 & \dots & \overline{r}_{k-1} \\ \vdots & \ddots & \ddots & \vdots \\ r_k & \dots & r_1 & r_0 \end{bmatrix} \begin{bmatrix} 1 \\ a_1^{(k)} \\ \vdots \\ a_k^{(k)} \end{bmatrix} = \begin{bmatrix} P_k \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

It is well known that the mean squared error is given by

$$P_k = \frac{\det R_{k+1}}{\det R_k}.$$

where

$$R_k = \begin{bmatrix} r_0 & \overline{r}_1 & \dots & \overline{r}_k \\ r_1 & r_0 & \dots & \overline{r}_{k-1} \\ \vdots & \ddots & \ddots & \vdots \\ r_k & \dots & r_1 & r_0 \end{bmatrix}.$$

The positive definiteness of R_n yields that $\det R_k > 0$ and thus the expression of P_k is well defined and $P_k > 0$. We recall the following definition of reflection coefficients.

Definition 6.1. For every $1 \le k \le n-1$, the last optimal coefficient $a_k^{(k)}$ is called the k-th reflection coefficient and is denoted by μ_k.

By the normal equation and the expression of the mean squared error we know that $\mu_1, \dots, \mu_{n-1}$ are uniquely determined by the matrix R_n. Moreover, the classical Levinson's recursion gives that

$$P_0 = r_0, \quad P_k = (1 - |\mu_n|^2) P_{k-1}, \quad 1 \le k \le n-1. \tag{1.1}$$

Hence

$$|\mu_k| < 1, \quad 1 \le k \le n-1.$$

Now we obtain a map between two open submanifolds of $\mathbf{R}^{2n-1}$:

$$\varphi: \quad \mathcal{T}_n \longrightarrow \mathbf{R}_+^* \times \mathbf{D}^{n-1}, \quad R_n \longmapsto (P_0, \mu_1, \ldots, \mu_{n-1}),$$

where $\mathbf{D} = \{z \in \mathbf{C} : |z| < 1\}$ is the unit disc of the complex plane.

The map φ is of fundamental importance in the following since it gives us another coordinate on $\mathcal{T}_n$ which simplifies lots of calculations. It is necessary to give the explicit expression of it.

Proposition 6.2. *The reflection coefficients are given by*

$$\mu_k = (-1)^k \frac{\det S_k}{\det R_k}, \quad k = 1, \ldots, n-1$$

where

$$S_k = R_{k+1}\begin{pmatrix} 2, \ldots, k+1 \\ 1, \ldots, k \end{pmatrix}.$$

Proof. By Cramer's rule and the definition of μ_k, we obtain that

$$\mu_k = (-1)^k \frac{P_k \det S_k}{\det R_{k+1}} = (-1)^k \frac{\det R_{k+1}}{\det R_k} \frac{\det S_k}{\det R_{k+1}} = (-1)^k \frac{\det S_k}{\det R_k}.$$

Hence it is clear that φ is differentiable. To see that φ is surjective, it suffices to note that $r_0 = P_0$ and that for every $1 \leq k \leq n-1$, if we already know $r_0, \ldots, r_{k-1}$, then the above identity is in fact a linear equation with respect to r_k, hence r_k can be calculated explicitly and the surjectivity holds by induction. The same method also yields the injectivity of φ. $\square$

The expression of φ^{-1} has also to be calculated for future use.

Proposition 6.3. *Let $(P_0, \mu_1, \ldots, \mu_{n-1}) \in \mathbf{R}_+^* \times \mathbf{D}^{n-1}$, then its inverse image R_n under φ can be calculated by the following algorithm:*

$$r_0 = P_0, \qquad r_1 = -P_0\mu_1,$$

$$r_k = -\mu_k P_{k-1} + \alpha_{k-1}^T J_{k-1} R_{k-1}^{-1} \alpha_{k-1}, \quad 2 \leq k \leq n-1$$

where

$$\alpha_{k-1} = \begin{bmatrix} r_1 \\ \vdots \\ r_{k-1} \end{bmatrix}, \quad J_{k-1} = \begin{bmatrix} 0 & \ldots & 0 & 1 \\ 0 & \ldots & 1 & 0 \\ & \ldots & & \\ 1 & \ldots & 0 & 0 \end{bmatrix},$$

$$P_{k-1} = P_0 \prod_{i=1}^{k-1} (1 - |\mu_i|^2).$$

Proof. The first two identities follow from direct calculation and for $2 \leq k \leq n-1$, it is easy to see that

$$S_k = \begin{bmatrix} \alpha_{k-1} & R_{k-1} \\ r_k & \alpha_{k-1}^T J_{k-1} \end{bmatrix}.$$

Using the method of Schur complement (see [95]) we get

$$\begin{aligned} \det S_k &= \det \begin{bmatrix} \alpha_{k-1} & R_{k-1} \\ r_k - \alpha_{k-1}^T J_{k-1} R_{k-1}^{-1} \alpha_{k-1} & 0 \end{bmatrix} \\ &= (-1)^{k-1} \det \begin{bmatrix} R_{k-1} & \alpha_{k-1} \\ 0 & r_k - \alpha_{k-1}^T J_{k-1} R_{k-1}^{-1} \alpha_{k-1} \end{bmatrix} \\ &= (-1)^{k-1} \det R_{k-1} (r_k - \alpha_{k-1}^T J_{k-1} R_{k-1}^{-1} \alpha_{k-1}). \end{aligned}$$

hence we obtain that

$$\begin{aligned} r_k &= (-1)^{k-1} \frac{\det S_k}{\det R_{k-1}} + \alpha_{k-1}^T J_{k-1} R_{k-1}^{-1} \alpha_{k-1} \\ &= (-1)^{k-1} \frac{\det S_k}{\det R_k} \frac{\det R_k}{\det R_{k-1}} + \alpha_{k-1}^T J_{k-1} R_{k-1}^{-1} \alpha_{k-1} \\ &= -\mu_k P_{k-1} + \alpha_{k-1}^T J_{k-1} R_{k-1}^{-1} \alpha_{k-1}. \end{aligned}$$

□

Corollary 6.4. *φ is a diffeomorphism.*

Proof. By Proposition 6.2, φ is injective. Hence it suffices to show that φ and φ^{-1} are differentiable, which are direct corollaries to the previous two propositions. □

6.2 Riemannian geometry of $\mathcal{T}_n$

From now on, we regard $\mathcal{T}_n$ as a Riemannian manifold whose metric, which is introduced in [14] by the Hessian of the Kähler potential (see e.g. [12] for the definition of a Kähler potential)

$$\Phi(R_n) = -\ln(\det R_n) - n \ln(\pi e),$$

is given by

$$ds^2 = n \frac{dP_0^2}{P_0^2} + \sum_{k=1}^{n-1} (n-k) \frac{|d\mu_k|^2}{(1-|\mu_k|^2)^2}, \tag{2.2}$$

where $(P_0, \mu_1, \ldots, \mu_{n-1}) = \varphi(R_n)$.

The metric (2.2) is a Bergman type metric and it will be shown in the appendix of Chapter 6 that this metric is not equal to the Fisher information metric of $\mathcal{T}_n$. But J. Burbea and C. R. Rao have proved in [28, Theorem 2] that the Bergman metric and the Fisher information metric do coincide for some probability density functions of particular forms. A similar potential function was used by S. Amari in [5] to derive the Riemannian metric of multi-variate Gaussian distributions by means of divergence functions. We refer to [78] for more account on the geometry of hessian structures.

With the metric given by (2.2) the space $\mathbf{R}_+^* \times \mathbf{D}^{n-1}$ is just the product of the Riemannian manifolds $(\mathbf{R}_+^*, ds_0^2)$ and $(\mathbf{D}, ds_k^2)_{1\le k\le n-1}$, where

$$ds_0^2 = n\frac{dP_0^2}{P_0^2} \quad \text{and} \quad ds_k^2 = (n-k)\frac{|d\mu_k|^2}{(1-|\mu_k|^2)^2}.$$

Also observe that $(\mathbf{R}_+^*, ds_0^2)$ has zero sectional curvature and that $(\mathbf{D}, ds_k^2)$ has constant sectional curvature $-4/(n-k)$.

We proceed to show a simple geometric lemma which is needed to obtain the lower sectional curvature bound of the product space.

Lemma 6.5. *Let M, N be two Riemannian manifolds such that the sectional curvatures verify*

$$-C \le K_M, K_N \le 0,$$

where $C \ge 0$ is a constant. Then the sectional curvatures of the product $M \times N$ also verify

$$-C \le K_{M\times N} \le 0.$$

Proof. Let $(x,y) \in M\times N$ and u, v be independent tangent vectors in $T_{(x,y)}M\times N$. Assume that $u = (u_1, u_2)$, $v = (v_1, v_2)$ with u_1, $v_1 \in T_xM$, u_2, $v_2 \in T_yN$, then we have

$$\begin{aligned} K_{M\times N} &= \frac{R_{M\times N}(u,v,u,v)}{|u|^2|v|^2 - \langle u,v\rangle^2} = \frac{R_M(u_1,v_1,u_1,v_1) + R_N(u_2,v_2,u_2,v_2)}{|u|^2|v|^2 - \langle u,v\rangle^2} \\ &= \frac{K_M(\,|u_1|_M^2|v_1|_M^2 - \langle u_1,v_1\rangle_M^2) + K_N(\,|u_2|_N^2|v_2|_N^2 - \langle u_2,v_2\rangle_N^2)}{|u|^2|v|^2 - \langle u,v\rangle^2} \\ &\le 0. \end{aligned}$$

To prove the lower bound, if u_i and v_i are linearly dependent for $i = 1, 2$, then we have trivially $K_{M\times N} = 0$, thus without loss of generality, we may assume that u_1 and v_1 are linearly independent. In this case, we have

$$\begin{aligned} K_{M\times N} &= \frac{K_M(\,|u_1|_M^2|v_1|_M^2 - \langle u_1,v_1\rangle_M^2) + K_N(\,|u_2|_N^2|v_2|_N^2 - \langle u_2,v_2\rangle_N^2)}{(|u_1|_M^2 + |u_2|_N^2)(|v_1|_M^2 + |v_2|_N^2) - (\langle u_1,v_1\rangle_M + \langle u_2,v_2\rangle_N)^2} \\ &\ge \frac{K_M(\,|u_1|_M^2|v_1|_M^2 - \langle u_1,v_1\rangle_M^2) + K_N(\,|u_2|_N^2|v_2|_N^2 - \langle u_2,v_2\rangle_N^2)}{(\,|u_1|_M^2|v_1|_M^2 - \langle u_1,v_1\rangle_M^2) + (\,|u_2|_N^2|v_2|_N^2 - \langle u_2,v_2\rangle_N^2)} \\ &\ge -C. \end{aligned}$$

$\square$

Proposition 6.6. *$(\mathbf{R}_+^* \times \mathbf{D}^{n-1}, ds^2)$ is a Cartan-Hadamard manifold whose sectional curvatures K verify $-4 \leq K \leq 0$.*

Proof. Since the product of finitely many Cartan-Hadamard manifolds is again Cartan-Hadamard, the first assertion follows. To show the curvature bounds, it suffices to use the preceding lemma by noting that every manifold in the product has sectional curvatures in $[-4, 0]$. □

6.3 Geodesics in $\mathbf{R}_+^* \times \mathbf{D}^{n-1}$

In practice, the geodesics in $\mathbf{R}_+^* \times \mathbf{D}^{n-1}$ are necessary to calculate the median of covariance matrices by using a subgradient algorithm and now we take the task of calculating these geodesics.

6.3.1 Geodesics in the Poincaré disc

Let $\mathbf{D} = \{z \in \mathbf{C} : |z| < 1\}$ be the Poicaré disc with its Riemannian metric

$$ds^2 = \frac{|dz|^2}{(1 - |z|^2)^2}.$$

The Riemannian distance between $z_1, z_2 \in \mathbf{D}$ is given by

$$\sigma(z_1, z_2) = \frac{1}{2} \ln \frac{1 + |\frac{z_2 - z_1}{1 - \bar{z}_1 z_2}|}{1 - |\frac{z_2 - z_1}{1 - \bar{z}_1 z_2}|}.$$

Lemma 6.7. *Let $0 < |z| < 1$, the geodesic from 0 to z parameterized by arc length is given by*

$$\gamma(s, 0, z) = e^{i\theta} \frac{e^{2s} - 1}{e^{2s} + 1}, \quad s \in [0, \ln \frac{1 + |z|}{1 - |z|}],$$

where $\theta = \arg z$.

Proof. The image of this geodesic is

$$\gamma(t, 0, z) = tz, \quad t \in [0, 1].$$

Then it suffices to reparametrize it by arc length

$$s = s(t) = \sigma(0, \gamma(t, 0, z)) = \frac{1}{2} \ln \frac{1 + |z|t}{1 - |z|t} \in [0, \ln \frac{1 + |z|}{1 - |z|}],$$

thus we obtain

$$t = \frac{1}{|z|} \frac{e^{2s} - 1}{e^{2s} + 1},$$

then

$$\gamma(s, 0, z) = \frac{z}{|z|} \frac{e^{2s} - 1}{e^{2s} + 1} = e^{i\theta} \frac{e^{2s} - 1}{e^{2s} + 1}.$$

□

Proposition 6.8. *Let $z_1, z_2 \in \mathbf{D}$, $z_1 \neq z_2$. The geodesic from z_1 to z_2 parameterized by arc length is given by*

$$\gamma(s, z_1, z_2) = \frac{(z_1 + e^{i\theta})e^{2s} + (z_1 - e^{i\theta})}{(1 + \bar{z}_1 e^{i\theta})e^{2s} + (1 - \bar{z}_1 e^{i\theta})}, \quad s \in [0, \frac{1}{2} \ln \frac{1 + |\phi_{z_1}(z_2)|}{1 - |\phi_{z_1}(z_2)|}],$$

where

$$\theta = \arg \phi_{z_1}(z_2), \quad \phi_{z_1}(z) = \frac{z - z_1}{1 + \bar{z}_1 z}.$$

Proof. Observe that

$$\gamma(s, 0, \phi_{z_1}(z_2)) = \gamma(s, \phi_{z_1}(z_1), \phi_{z_1}(z_2)),$$

since ϕ_{z_1} is an isometry of $\mathbf{D}$, we get

$$\gamma(s, z_1, z_2) = \phi_{z_1}^{-1} \circ \gamma(s, 0, \phi_{z_1}(z_2)).$$

Then the preceding lemma gives

$$\gamma(s, z_1, z_2) = \frac{e^{i\theta} \frac{e^{2s}-1}{e^{2s}+1} + z_1}{1 + \bar{z}_1 e^{i\theta} \frac{e^{2s}-1}{e^{2s}+1}} = \frac{(z_1 + e^{i\theta})e^{2s} + (z_1 - e^{i\theta})}{(1 + \bar{z}_1 e^{i\theta})e^{2s} + (1 - \bar{z}_1 e^{i\theta})}.$$

with

$$s \in [0, \frac{1}{2} \ln \frac{1 + |\phi_{z_1}(z_2)|}{1 - |\phi_{z_1}(z_2)|}].$$

□

Corollary 6.9.

$$\gamma'(s, z_1, z_2) = \frac{4e^{i\theta}(1 - |z_1|^2)e^{2s}}{((1 + \bar{z}_1 e^{i\theta})e^{2s} + (1 - \bar{z}_1 e^{i\theta}))^2}.$$

Particularly,

$$\gamma'(0, z_1, z_2) = e^{i\theta}(1 - |z_1|^2).$$

Proof. Direct calculation. □

Proposition 6.10. *Let $z_1 \in \mathbf{D}$ and $v \in T_{z_1}\mathbf{D}$, then the geodesic starting from z_1 with velocity v is given by*

$$\zeta(t, z_1, v) = \frac{(z_1 + e^{i\theta})e^{2||v||t} + (z_1 - e^{i\theta})}{(1 + \bar{z}_1 e^{i\theta})e^{2||v||t} + (1 - \bar{z}_1 e^{i\theta})}, \quad t \in \mathbf{R}$$

where

$$\theta = \arg v, \quad ||v|| = \frac{|v|}{1 - |z_1|^2}.$$

Proof. Let $z_2 \neq z_1$ be another point in the geodesic, then

$$\gamma'(0, z_1, z_2) = \frac{v}{||v||}.$$

By the last corollary,

$$e^{i\theta}(1 - |z_1|^2) = \frac{v}{||v||},$$

hence $\theta = \arg v$ and

$$\zeta(t, z_1, v) = \zeta(||v||t, z_1, \frac{v}{||v||}) = \gamma(||v||t, z_1, z_2).$$

it remains to use proposition 6.8.

□

6.3.2 Geodesics in $\mathbf{R}_+^*$

Let $\mathbf{R}_+^*$ be with the Riemannian metric

$$ds^2 = \frac{dP^2}{P^2}.$$

By [31], the Riemannian distance between $P, Q \in \mathbf{R}_+^*$ is given by

$$\tau(P, Q) = |\ln(\frac{Q}{P})|.$$

Proposition 6.11. *Let $P, Q \in \mathbf{R}_+^*$ and $P \neq Q$. Then the geodesic from P to Q parameterized by arc length is given by*

$$\gamma(s, P, Q) = Pe^{(\operatorname{sign}(Q-P))s}, \quad s \in [0, |\ln(\frac{Q}{P})|].$$

Proof. By [31], the geodesic from P to Q is given by

$$\gamma(t, P, Q) = P(\frac{Q}{P})^t, \quad t \in [0, 1]$$

it suffices to reparametrize it by letting $t = |\ln(Q/P)|/s$. □

Corollary 6.12.

$$\gamma'(s, P, Q) = (\operatorname{sign}(Q - P))Pe^{(\operatorname{sign}(Q-P))s}.$$

Particularly,

$$\gamma'(0, P, Q) = (\operatorname{sign}(Q - P))P.$$

Proof. Direct calculation. □

Proposition 6.13. *Let $P > 0$ and $v \in T_P\mathbf{R}_+^*$, the geodesic starting from P with velocity v is given by*

$$\zeta(t, P, v) = Pe^{\frac{v}{P}t}, \quad t \in \mathbf{R}.$$

Proof. Let $Q \neq P$ be another point in this geodesic, by the preceding corollary we have

$$\gamma'(0, P, Q) = (\mathrm{sign}(Q - P))P = \frac{v}{||v||}.$$

with $||v|| = |v|/P$ and hence

$$\mathrm{sign}(Q - P) = \frac{v}{||v||P}.$$

Then Proposition 6.11 yields that

$$\zeta(s, P, \frac{v}{||v||}) = Pe^{\frac{vs}{||v||P}},$$

thus we get

$$\zeta(t, P, v) = \zeta(t||v||, P, \frac{v}{||v||}) = Pe^{\frac{v}{P}t}.$$

□

6.3.3 Geodesics in $\mathbf{R}_+^* \times \mathbf{D}^{\mathbf{n-1}}$

As is shown in [31], the Riemannian distance between two points x and y in $\mathbf{R}_+^* \times \mathbf{D}^{n-1}$ is given by

$$d(x, y) = \left(n\sigma(P, Q)^2 + \sum_{k=1}^{n-1}(n - k)\tau(\mu_k, \nu_k)^2\right)^{1/2},$$

where $x = (P, \mu_1, \ldots, \mu_{n-1})$ and $y = (Q, \nu_1, \ldots, \nu_{n-1})$.

Proposition 6.14. *Let $x = (P, \mu_1, \ldots, \mu_{n-1})$ and $y = (Q, \nu_1, \ldots, \nu_{n-1})$ be two different points in $\mathbf{R}_+^* \times \mathbf{D}^{n-1}$. Then the geodesic from x to y parameterized by arc length is given by*

$$\gamma(s, x, y) = \left(\gamma_0(\frac{\sigma(P, Q)}{d(x, y)}s), \gamma_1(\frac{\tau(\mu_1, \nu_1)}{d(x, y)}s), \ldots, \gamma_1(\frac{\tau(\mu_{n-1}, \nu_{n-1})}{d(x, y)}s)\right).$$

where γ_0 is the geodesic in $(\mathbf{R}_+^*, ds_0^2)$ from P to Q parameterized by arc length and for $1 \leq k \leq n-1$, γ_k is the geodesic in $(\mathbf{D}, ds_k^2)$ from μ_k to ν_k parameterized by arc length. More precisely,

$$\gamma_0(\frac{\sigma(P, Q)}{d(x, y)}s) = Pe^{\frac{(\mathrm{sign}(Q-P))\sigma(P,Q)}{d(x,y)}s},$$

and for $1 \leq k \leq n-1$,

$$\gamma_k(\frac{\tau(\mu_k,\nu_k)}{d(x,y)}s) = \frac{(\mu_k + e^{i\theta_k})e^{\frac{2\tau(\mu_k,\nu_k)}{d(x,y)}s} + (\mu_k - e^{i\theta_k})}{(1+\bar{\mu}_k e^{i\theta_k})e^{\frac{2\tau(\mu_k,\nu_k)}{d(x,y)}s} + (1-\bar{\mu}_k e^{i\theta_k})},$$

with

$$\theta_k = \arg\frac{\nu_k - \mu_k}{1-\bar{\mu}_k\nu_k}.$$

Particularly,

$$\gamma'(0,x,y) = \left(\gamma_0'(0)\frac{\sigma(P,Q)}{d(x,y)}, \gamma_1'(0)\frac{\tau(\mu_1,\nu_1)}{d(x,y)}, \ldots, \gamma_{n-1}'(0)\frac{\tau(\mu_{n-1},\nu_{n-1})}{d(x,y)}\right).$$

Proof. It suffices to note that [31] the geodesic in $\mathbf{R}_+^* \times \mathbf{D}^{n-1}$ is the product of the geodesics in each manifold in the product and then use Proposition 6.8 as well as Proposition 6.11. The last identity follows from direct calculation. □

Proposition 6.15. *Let* $x = (P, \mu_1, \ldots, \mu_{n-1}) \in \mathbf{R}_+^* \times \mathbf{D}^{n-1}$ *and a tangent vector* $v = (v_0, v_1, \ldots, v_{n-1}) \in T_x(\mathbf{R}_+^* \times \mathbf{D}^{n-1})$, *then the geodesic starting from* x *with velocity* v *is given by*

$$\zeta(t,x,v) = (\zeta_0(t), \zeta_1(t), \ldots, \zeta_{n-1}(t)),$$

where ζ_0 *is the geodesic in* $(\mathbf{R}_+^*, ds_0^2)$ *starting from* P *with velocity* v_0 *and for* $1 \leq k \leq n-1$, ζ_k *is the geodesic in* $(\mathbf{D}, ds_k^2)$ *starting from* μ_k *with velocity* v_k. *More precisely,*

$$\zeta_0(t) = Pe^{\frac{v_0}{P}t},$$

and for $1 \leq k \leq n-1$,

$$\zeta_k(t) = \frac{(\mu_k + e^{i\theta_k})e^{\frac{2|v_k|t}{1-|\mu_k|^2}} + (\mu_k - e^{i\theta_k})}{(1+\bar{\mu}_k e^{i\theta_k})e^{\frac{2|v_k|t}{1-|\mu_k|^2}} + (1-\bar{\mu}_k e^{i\theta_k})},$$

with

$$\theta_k = \arg\frac{\nu_k - \mu_k}{1-\bar{\mu}_k\nu_k}.$$

Proof. The same method as the preceding proof but using Proposition 6.10 and Proposition 6.13. □

6.4 Simulations

With all of the above preparative calculations, now we use the subgradient algorithm introduced in [91] to calculate the median of covariant matrices. By the change of coordinates φ, it suffices to do this in the product space $(\mathbf{R}_+^* \times \mathbf{D}^{n-1}, ds^2)$.

By Proposition 6.6, the upper and lower curvature bounds are given by 0 and -4 respectively, hence the algorithm here is simpler and more explicit than the general one. Let $p_1, \dots, p_N$ be different points contained in an open ball $B(a,\rho)$ of $\mathbf{R}_+^* \times \mathbf{D}^{n-1}$ and assume that they are not totally contained in any geodesic. Then [91] the median m of $p_1, \dots, p_N$, or equivalently, the minimum point of the function

$$f: \qquad \bar{B}(a,\rho) \longrightarrow \mathbf{R}_+\,, \qquad x \longmapsto \frac{1}{N}\sum_{i=1}^{N} d(x,p_i).$$

exists and is unique. In order to introduce the subgradient algorithm, we need the following notations.

Notation 6.16. For $x \in \bar{B}(a,\rho)$, let

$$H(x) = \frac{1}{N}\sum_{\substack{1\le i\le N\\ p_i\neq x}} \frac{-\exp_x^{-1} p_i}{d(x,p_i)}.$$

If $H(x) \neq 0$, then let

$$\gamma_x(t) = \exp_x\Big(-t\frac{H(x)}{|H(x)|}\Big)\,, \quad t \geq 0.$$

Moreover,

$$\beta = \frac{\rho-\sigma}{4\rho\coth(4\rho)+1}, \qquad \sigma = \max_{1\le i\le N} d(p_i,a).$$

Now we specialize the subgradient algorithm in [91] to the space of reflection coefficients.

Algorithm 6.17. Subgradient algorithm in $(\mathbf{R}_+^* \times \mathbf{D}^{n-1}, ds^2)$:
Step 1:
Choose a point $x_1 \in \bar{B}(a,\rho)$ and let $k=1$.
Step 2:
If $H(x_k)=0$, then stop and let $m = x_k$.
If not, go to step 3.
Step 3:
Let $x_{k+1} = \gamma_{x_k}(\beta/\sqrt{k})$ and go back to step 2 with $k = k+1$.

6.4.1 A Numerical example

Firstly, in order to illustrate the above method, we calculate the geometric median of 4 elements in $\mathcal{T}_4$, whose first lines are given by

$$[\,1.0000,\ -0.5000+0.3000i,\ 0.4240-0.7620i,\ -0.0903+0.4528i\,]$$

$$[\,2.5000,\ -1.5000-1.0000i,\ 0.3800+1.4400i,\ -0.2088-1.3544i\,]$$

$$[\,3.7000,\ -0.7400 - 0.7400i,\ -2.3828 + 0.6364i,\ 2.0608 - 0.6855i\,]$$
$$[\,0.3000,\ -0.0300 + 0.2400i,\ -0.2625 - 0.1005i,\ 0.0861 - 0.2123i\,]$$

Then Lemma 6.2 gives their reflection coefficients parametrization:

$$[\,1,\ 0.5 + 0.3i,\ -0.4 - 0.7i,\ 0.5 - 0.5i\,]$$
$$[\,2.5,\ 0.6 - 0.4i,\ 0.1 + 0.2i,\ 0.6 - 0.2i\,]$$
$$[\,3.7,\ 0.2 - 0.2i,\ 0.7 + 0.1i,\ -0.4 - 0.6i\,]$$
$$[\,0.3,\ 0.1 + 0.8i,\ 0.7 - 0.5i,\ 0.5 + 0.5i\,].$$

By using the subgradient algorithm, Proposition 6.14 and Proposition 6.15, we get the median in terms of reflection coefficients:

$$[\,1.6611,\ 0.3543 - 0.0379i,\ 0.1663 - 0.1495i,\ 0.2749 - 0.2371i\,]$$

Then Lemma 6.3 gives the first line of the corresponding Toeplitz Hermitan positive definite matrix:

$$[\,1.6611,\ -0.5885 - 0.0630i,\ -0.0350 - 0.1722i,\ -0.2901 - 0.1531i\,]$$

6.4.2 Radar simulations

Next we give some simulating examples of the median method applied to radar target detection. I would like to thank Guillaume Bouyt and Nicolas Charon for their generosity of providing basic simulation programs and many helpful discussions.

Since the autoregressive spectra are closely related to the speed of targets, we shall first investigate the spectral performance of the median method. In order to illustrate the basic idea, we only consider the detection of one fixed direction. The range along this direction is subdivided into 200 lattices in which we add two targets, the echo of each lattice is modeled by an autoregressive process. The following Figure 6.1 obtained in [27] gives the initial spectra of the simulation, where x axis represents the lattices and y axis represents frequencies. Every lattice is identified with a 1×8 vector of reflection coefficients which is calculated by using the regularized Burg algorithm [17] to the original simulating data. The spectra are represented by different colors whose corresponding values are indicated in the colorimetric on the right.

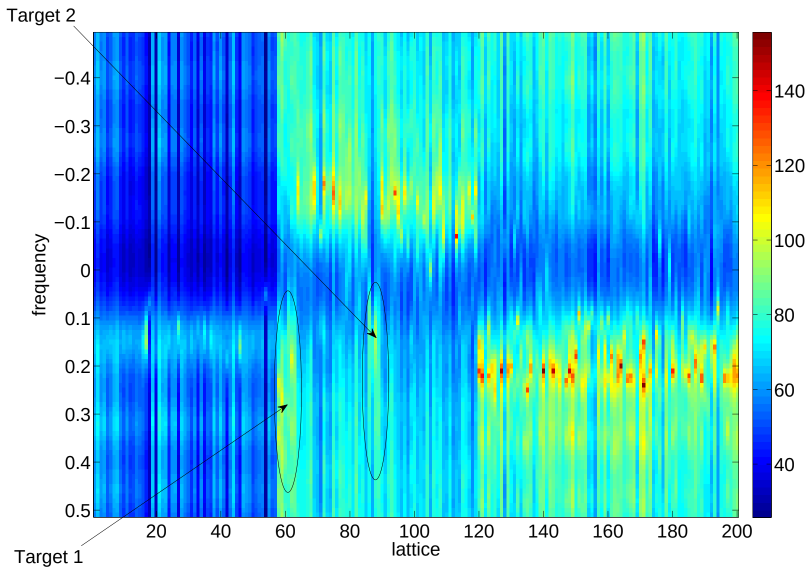


Figure 6.1: Initial spectra with two added targets

For every lattice, by using the subgradient algorithm, we calculate the median of the window centered on it and consisting of 15 lattices and then we get the spectra of medians shown in Figure 6.2. Furthermore, by comparing it with Figure 6.3 which are spectra of barycenters, we see that in the middle of the barycenter spectra, this is just the place where the second target appears, there is an obvious distortion. This explains that median is much more robust than barycenter when outliers come.

The principle of target detection is that a target appears in a lattice if the distance between this lattice and the median of the window around it is much bigger than that of the ambient lattices. The following Figure 6.4 shows that the two added targets are well detected by the median method, where x axis represents lattice and y axis represents the distance in $\mathcal{T}_8$ between each lattice and the median of the window around it.

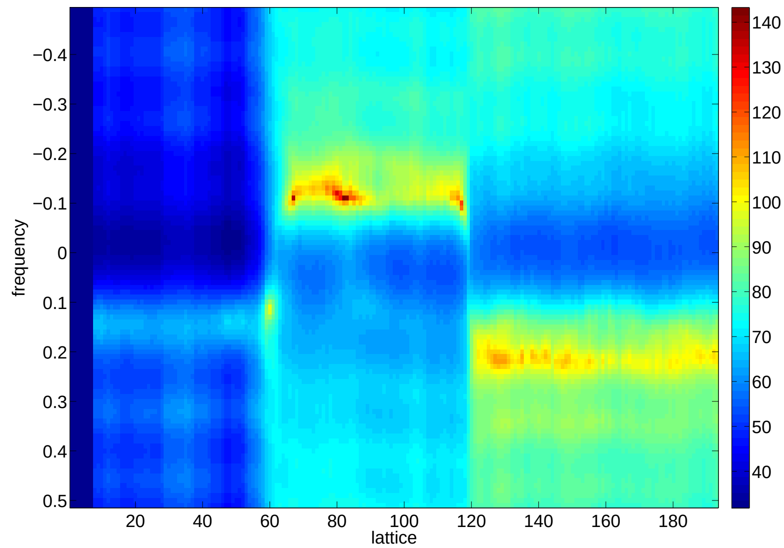


Figure 6.2: Median spectra

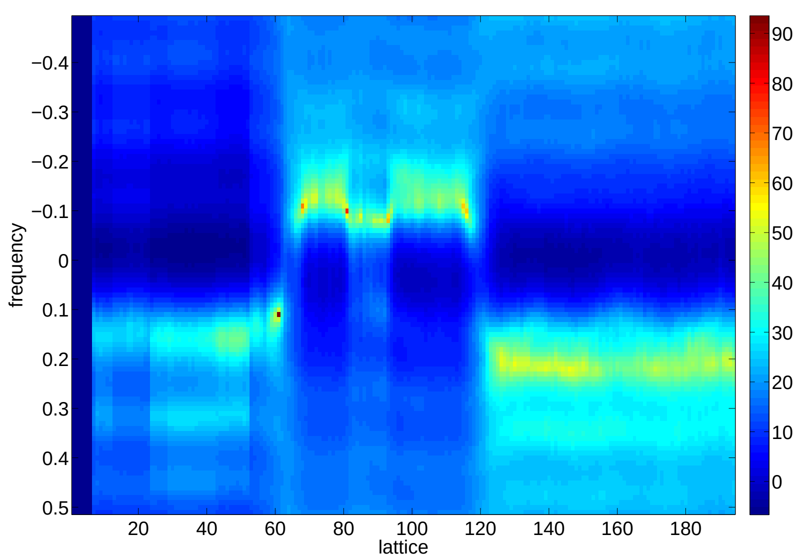


Figure 6.3: Barycenter spectra

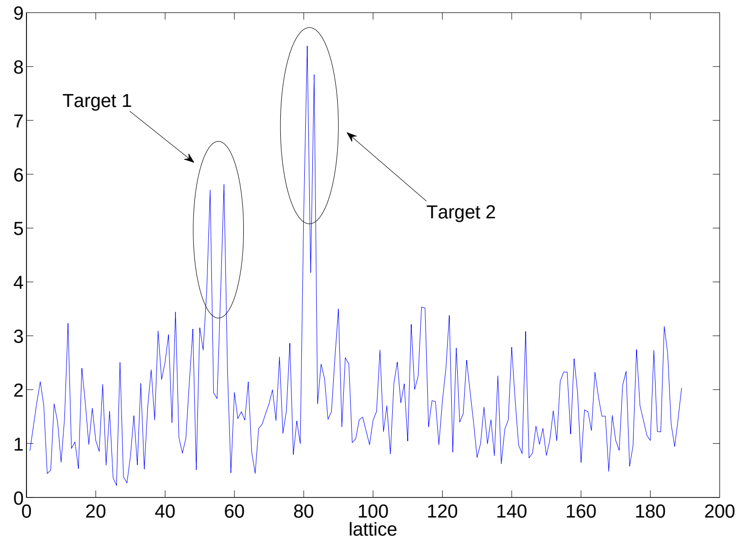


Figure 6.4: Detection by median

We conclude our discussion by showing the performance of the median method in real target detection. As above, we give the images of autoregressive spectra and the figure of target detection obtained by using real data which are records of a radar located on a coast. These records consist of about 5000 lattices of a range of about 10km-140km as well as 109 azimuth values corresponding to approximately 30 scanning degrees of the radar. For simplicity we consider the data of all the lattices but in a fixed direction, hence each lattice corresponds to a 1×8 vector of reflection coefficients computed by applying the regularized Burg algorithm to the original real data. Figure 6.5 gives the initial autoregressive spectra whose values are represented by different color according to the colorimetric on the right. For each lattice, by using the subgradient algorithm, we calculate the median of the window centered on it and consisting of 17 lattices and then we get the spectra of medians shown in Figure 6.6.

In order to know in which lattice target appears, we compare the distance between each lattice and the median of the window around it. The following Figure 6.7 shows that the four targets are well detected by our method, where x axis represents distance and y axis represents the distance in $\mathcal{T}_8$ between each lattice and the median of the window around it.

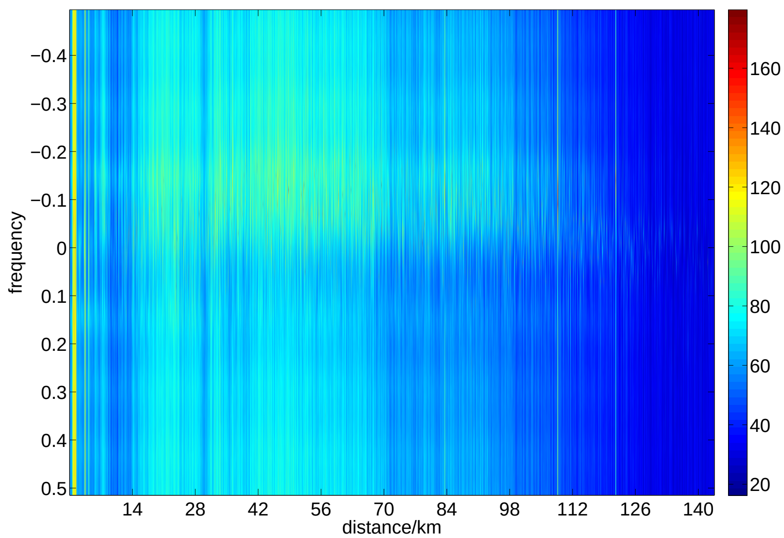


Figure 6.5: Initial spectra of real radar data

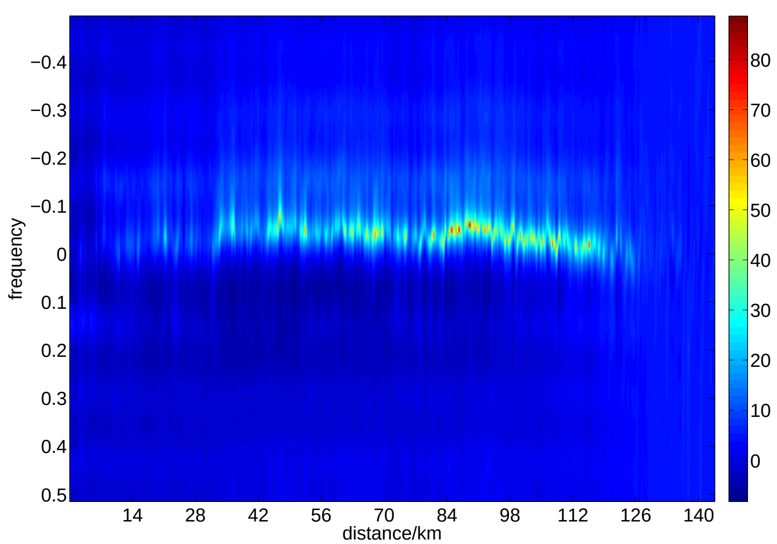


Figure 6.6: Median spectra of real radar data

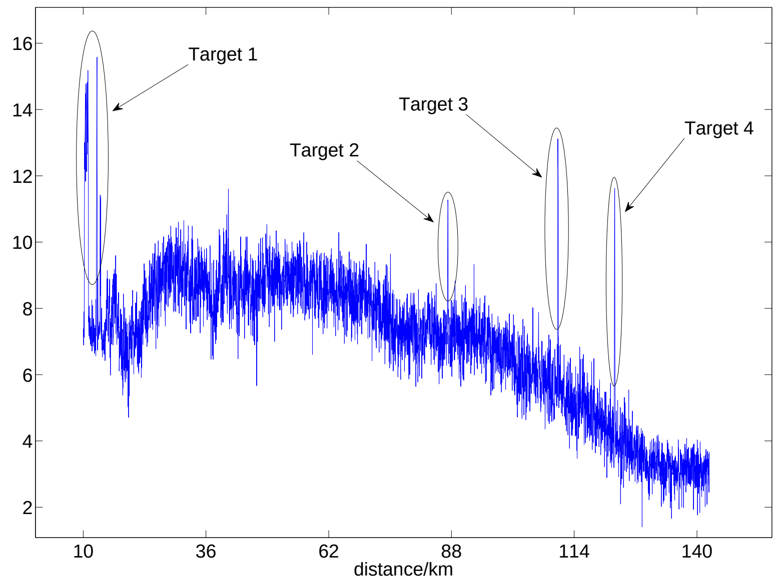


Figure 6.7: Real detection by median

6.5 Appendix

In this appendix, we firstly show that the metric given by (2.2) is not equal to the Fisher information metric of Toeplitz covariance matrices. To this end, we shall compute the elements g_{00} and g_{01} of the Fisher information metric in terms of reflection coefficients. We will see that the term g_{00} is equal to that of the metric (2.2), but $g_{10} \neq 0$, hence the Fisher information metric is not diagonal under reflection coefficients parametrization. As a result, these two metrics are not equal.

Let $R_n \in \mathcal{T}_n$, then the probability density function of the centered complex normal random variable of covariance matrix R_n is given by

$$p(Z|R_n) = \pi^{-n}(\det R_n)^{-1} \exp(-Z^* R_n^{-1} Z).$$

Hence we get,

$$\begin{aligned} -\ln p(Z|R_n) &= \ln \det R_n + Z^* R_n^{-1} Z + n \ln \pi \\ &= n \ln P_0 + \sum_{k=1}^{n-1} (n-k) \ln(1 - |\mu_k|^2) + Z^* R_n^{-1} Z + n \ln \pi. \end{aligned}$$

Let us compute the first coordinate of the Fisher information metric:

$$g_{00} = -\mathbf{E}[\frac{\partial^2 \ln p(Z|R_n)}{\partial^2 P_0}].$$

Observe that

$$-\frac{\partial^2 \ln p(Z|R_n)}{\partial^2 P_0} = -\frac{n}{P_0^2} + Z^* \frac{\partial^2 R_n^{-1}}{\partial^2 P_0} Z.$$

For simplicity, we will show that R_n is linear on P_0, that is,

$$\frac{\partial R_n}{\partial P_0} = \frac{R_n}{P_0}. \tag{5.3}$$

To this end, it suffices to show that

$$\frac{\partial r_k}{\partial P_0} = \frac{r_k}{P_0}, \quad k = 0, \dots n-1.$$

This is trivial for $k = 0, 1$. Now assume that this is true for $0, \dots k-1$, then for k one has, by Proposition 6.3,

$$\begin{aligned}
\frac{\partial r_k}{\partial P_0} =& -\mu_k \frac{\partial P_{k-1}}{\partial P_0} + \frac{\partial \alpha_{k-1}^T}{\partial P_0} J_{k-1} R_{k-1}^{-1} \alpha_{k-1} + \alpha_{k-1}^T J_{k-1} \frac{\partial R_{k-1}^{-1}}{\partial P_0} \alpha_{k-1} + \alpha_{k-1}^T J_{k-1} R_{k-1}^{-1} \frac{\partial \alpha_{k-1}}{\partial P_0} \\
=& -\mu_k \frac{P_{k-1}}{P_0} + \frac{\alpha_{k-1}^T}{P_0} J_{k-1} R_{k-1}^{-1} \alpha_{k-1} + \alpha_{k-1}^T J_{k-1} (-R_{k-1}^{-1} \frac{R_{k-1}}{P_0} R_{k-1}^{-1}) \alpha_{k-1} \\
& + \alpha_{k-1}^T J_{k-1} R_{k-1}^{-1} \frac{\alpha_{k-1}}{P_0} \\
=& \frac{1}{P_0} (-\mu_k P_{k-1} + \alpha_{k-1}^T J_{k-1} R_{k-1}^{-1} \alpha_{k-1}) = \frac{r_k}{P_0}.
\end{aligned}$$

Thus we have

$$\frac{\partial R_n^{-1}}{\partial P_0} = -R_n^{-1} \frac{\partial R_n}{\partial P_0} R_n^{-1} = -R_n^{-1} \frac{R_n}{P_0} R_n^{-1} = -\frac{R_n^{-1}}{P_0}.$$

$$\frac{\partial^2 R_n^{-1}}{\partial^2 P_0} = -\frac{P_0 \frac{\partial R_n^{-1}}{\partial P_0} - R_n^{-1}}{P_0^2} = \frac{2R_n^{-1}}{P_0^2}.$$

It follows that

$$\mathbf{E}[Z^* \frac{\partial^2 R_n^{-1}}{\partial^2 P_0} Z] = \mathrm{tr}[\frac{\partial^2 R_n^{-1}}{\partial^2 P_0} R_n] = \mathrm{tr}[\frac{2R_n^{-1}}{P_0^2} R_n] = 2\frac{n}{P_0^2}.$$

So that

$$g_{00} = -\frac{n}{P_0^2} + 2\frac{n}{P_0^2} = \frac{n}{P_0^2}$$

That is to say, the first coordinates of the two metrics coincide. But generally speaking, this is not true for the other coordinates. For example, we can show that $g_{10} \neq 0$ for the Fisher information metric. In fact, by (5.3) we have

$$\begin{aligned}
g_{10} =& \mathbf{E}[-\frac{\partial^2 \ln p(Z|R_n)}{\partial \mu_1 \partial P_0}] = \mathbf{E}[Z^* \frac{\partial^2 R_n^{-1}}{\partial \mu_1 \partial P_0} Z] = \mathrm{tr}[R_n \frac{\partial^2 R_n^{-1}}{\partial \mu_1 \partial P_0}] = \mathrm{tr}[R_n \frac{\partial}{\partial \mu_1} \frac{\partial R_n^{-1}}{\partial P_0}] \\
=& \mathrm{tr}[R_n \frac{\partial}{\partial \mu_1} \frac{-R_n^{-1}}{P_0}] = -\frac{1}{P_0} \mathrm{tr}[R_n \frac{\partial R_n^{-1}}{\partial \mu_1}] = \frac{1}{P_0} \mathrm{tr}[\frac{\partial R_n}{\partial \mu_1} R_n^{-1}].
\end{aligned}$$

This does not necessarily vanish. For instance, for the case when $n = 2$,

$$R_2 = \begin{bmatrix} P_0 & -P_0\bar{\mu}_1 \\ -P_0\mu_1 & P_0 \end{bmatrix} = P_0 \begin{bmatrix} 1 & -\bar{\mu}_1 \\ -\mu_1 & 1 \end{bmatrix}.$$

It follows that

$$\frac{\partial R_2}{\partial \mu_1} = P_0 \begin{bmatrix} 0 & 0 \\ -1 & 0 \end{bmatrix}, \quad R_2^{-1} = \frac{1}{P_0(1-|\mu_1|^2)} \begin{bmatrix} 1 & \bar{\mu}_1 \\ \mu_1 & 1 \end{bmatrix}.$$

Hence we get

$$g_{10} = \frac{1}{P_0} \mathrm{tr}[\frac{\partial R_2}{\partial \mu_1} R_2^{-1}] = \frac{-\bar{\mu}_1}{P_0(1-|\mu_1|^2)} \neq 0.$$

The second aim of this appendix is to give a direct algebraic proof of the identity $P_k = (1-|\mu_k|^2)P_{k-1}$ in (1.1). This proof is independent of the classical Levinson's recursion.

Proposition 6.18. *Let $P_0 = r_0$ and for $1 \leq k \leq n-1$, we define*

$$\mu_k = (-1)^k \frac{\det S_k}{\det R_k}, \quad P_k = \frac{\det R_{k+1}}{\det R_k},$$

then

$$P_k = (1-|\mu_k|^2)P_{k-1}.$$

Proof. According to Sylvester's identity,

$$\det R_{k+1} \det R_{k-1} = \det R_k \det R_k - \det R_{k+1}\begin{pmatrix} 2,\dots,k+1 \\ 1,\dots,k \end{pmatrix} \det R_{k+1}\begin{pmatrix} 1,\dots,k \\ 2,\dots,k+1 \end{pmatrix}.$$

Since R_{k+1} is hermitian, we get

$$\det R_{k+1}\begin{pmatrix} 1,\dots,k \\ 2,\dots,k+1 \end{pmatrix} = \overline{\det R_{k+1}\begin{pmatrix} 2,\dots,k+1 \\ 1,\dots,k \end{pmatrix}}.$$

So that

$$\begin{aligned} P_k = \frac{\det R_{k+1}}{\det R_k} &= \left(1 - \left|\frac{\det R_{k+1}\binom{2,\dots,k+1}{1,\dots,k}}{\det R_k}\right|^2\right) \frac{\det R_k}{\det R_{k-1}} \\ &= \left(1 - \left|\frac{\det S_k}{\det R_k}\right|^2\right) P_{k-1} \\ &= (1-|\mu_k|^2)P_{k-1}. \end{aligned}$$

The proof is complete. $\square$

We finish this appendix by showing some supplementary detection results of the median method (OS-HDR-CFAR) applied to real radar data of land clutter with synthetic targets injected. It can be seen that our new method has a better performance than the classical ones. I would like to thank Alexis Decurninge for his generosity of providing me his simulation results and helpful discussions.

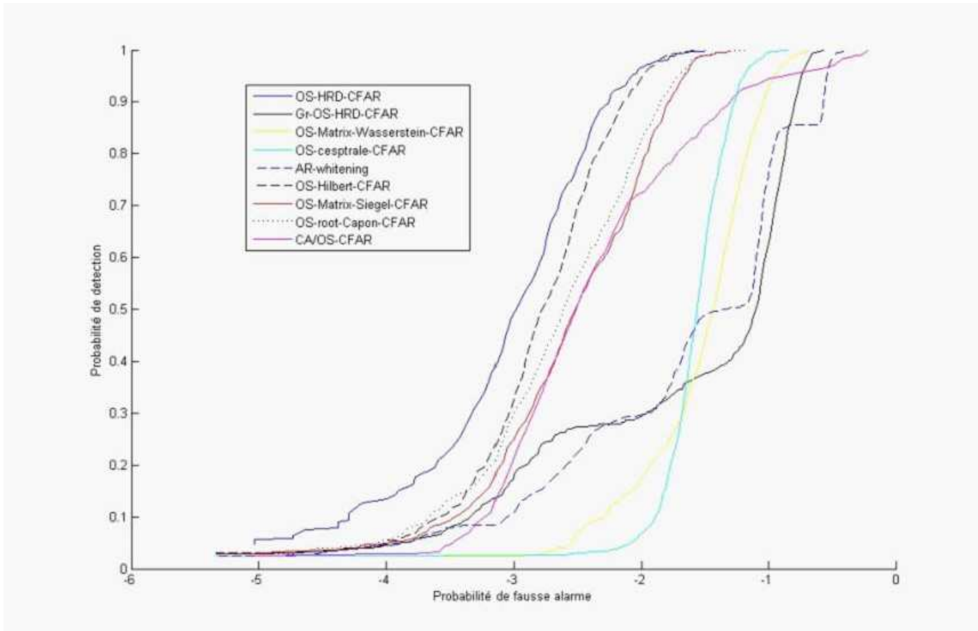


Figure 6.8: COR curves (sample with 160 azimuths), $\alpha = 0.25$, $\mathrm{SNR}_{\mathrm{moy}} = 17\mathrm{dB}$

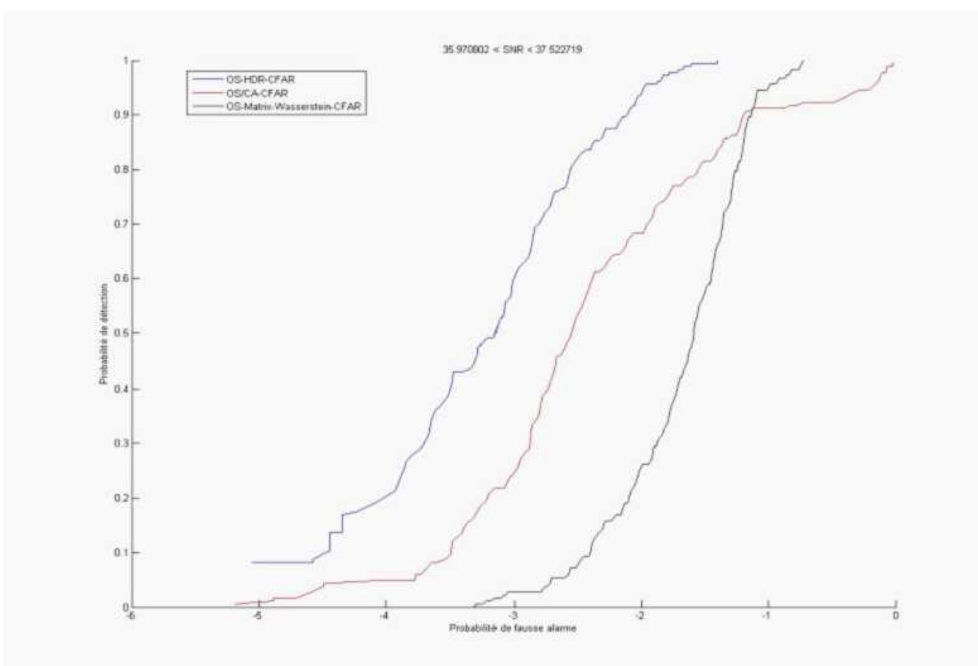


Figure 6.9: COR curves (SNR fixed), $\alpha = 0.25$, SNR=26dB

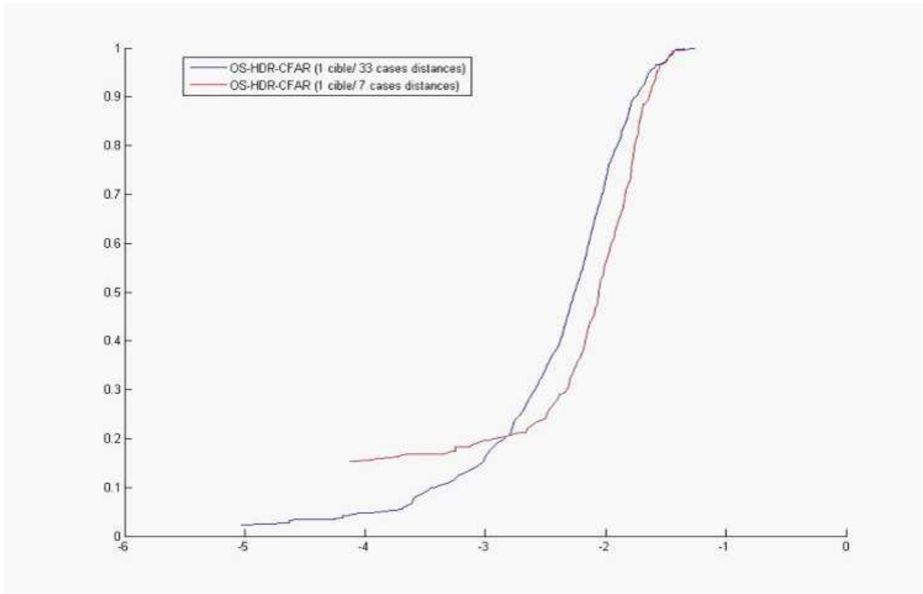


Figure 6.10: Robustness when many targets appear: OS-HDR-CFAR

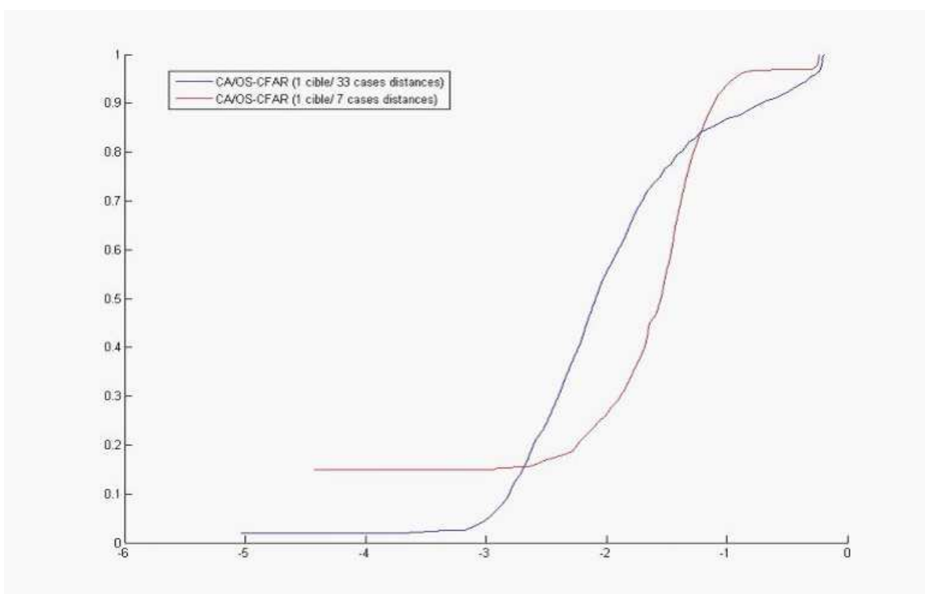


Figure 6.11: Robustness when many targets appear: CA/OS-CFAR

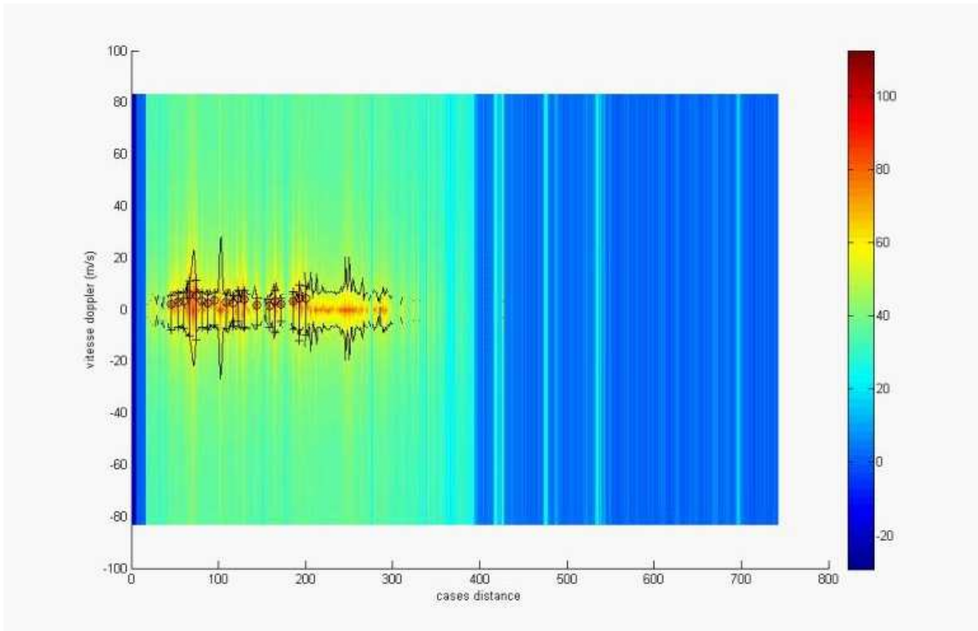


Figure 6.12: Robustness when many targets appear: Capon spectra

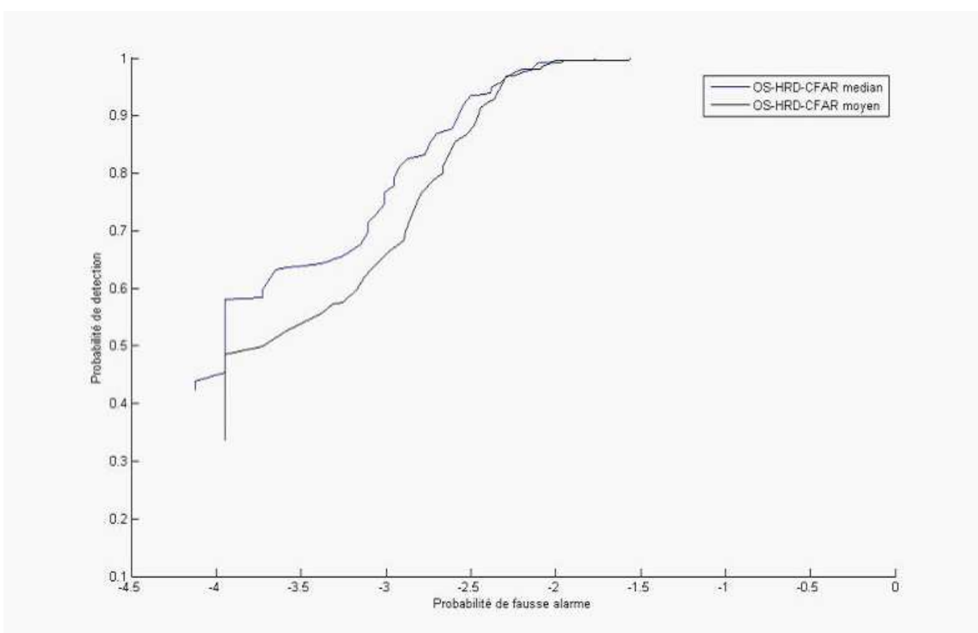


Figure 6.13: Comparison of median and mean: $\alpha = 10$, $\mathrm{SNR}_{\mathrm{moy}}$=17dB

Bibliography

[1] B. Afsari, *Riemannian L^p center of mass : existence, uniqueness, and convexity*, Proceedings of the American Mathematical Society, S 0002-9939(2010)10541-5, Article electronically published on August 27, 2010.

[2] A. D. Alexandrov, *Über eine Verallgemeinerung der Riemannschen Geometrie*, Schr. Forschungsinst. Math. 1 (1957) 33-84.

[3] A. D. Aleksandrov et al, *Generalized Riemannian spaces*, Russian Math. Surveys. 41 (1986), no. 3, 1-54.

[4] A. A. Aly et al., *Location dominance on spherical surfaces*, Operation Research, 27 (1979), no. 5.

[5] S. Amari and A. Cichocki, *Information geometry of divergence functions*, Bulletin of the Polish Academy of Sciences, Technical Sciences, Vol. 58, No. 1, 2010.

[6] M. Armatte, *Fréchet et la médiane: un moment dans une histoire de la robustesse*, Journal de la Société Française de Statistique, tome 147, n°2, 2006.

[7] M. Arnaudon, *Espérances conditionnelles et C-martingales dans les variétés*, Séminaire de Probabilités-XXVIII, Lecture Notes in Mathematics 1583. (Springer, Berlin, 1994), pp. 300-311.

[8] M. Arnaudon, *Barycentres convexes et approximations des martingales dans les variétés*, Séminaire de Probabilités-XXIX, Lecture Notes in Mathematics 1613. (Springer, Berlin, 1995), pp. 70-85.

[9] M. Arnaudon and X. M. Li, *Barycenters of measures transported by stochastic flows*, The Annals of probability, 33 (2005), no. 4, 1509-1543.

[10] M. Arnaudon, C. Dombry, A. Phan and L. Yang, *Stochastic algorithms for computing means of probability measures*, preprint hal-00540623, version 2, (2011). to appear in Stochastic Processes and their Applications.

[11] C. Bajaj, *The algebraic degree of geometric optimization problems*, Discret and Computational Geometry, Vol 3, No. 1, pp. 171-191 (1988).

[12] W. Ballmann, *Lectures on Kähler manifolds*, ESI Lectures in Mathematics and Physics, European Math. Soc., Zürich. 2006

[13] F. Barbaresco, *Innovative Tools for Radar Signal Processing Based on Cartan's Geometry of SPD Matrices and Information Geometry*, IEEE International Radar Conference (2008).

[14] F. Barbaresco, *Interactions between Symmetric Cone and Information Geometries*, ETVC'08, Springer Lecture Notes in Computer Science 5416 (2009), pp. 124-163.

[15] F. Barbaresco and G. Bouyt, *Espace Riemannien symétrique et géométrie des espaces de matrices de covariance : équations de diffusion et calculs de médianes*, GRETSI'09 conference, Dijon, September 2009

[16] F. Barbaresco, *New Foundation of Radar Doppler Signal Processing based on Advanced Differential Geometry of Symmetric Spaces: Doppler Matrix CFAR and Radar Application*, Radar'09 Conference, Bordeaux, October 2009

[17] F. Barbaresco, *Annalyse Doppler: régularisation d'un problème inverse mal posé*, Support de cours

[18] F. Barbaresco, *Science géométrique de lInformation : Géométrie des matrices de covariance, espace métrique de Fréchet et domaines bornés homogénes de Siegel*, Conférence GRETSI'11, Bordeaux, Sept. 2011

[19] F. Barbaresco, *Robust Statistical Radar Processing in Fréchet Metric Space: OS-HDR-CFAR and OS-STAP Processing in Siegel Homogeneous Bounded Domains*, Proceedings of IRS'11, International Radar Conference, Leipzig, Sept. 2011

[20] F. Barbaresco, *Geometric Radar Processing based on Fréchet Distance : Information Geometry versus Optimal Transport Theory*, Proceedings of IRS'11, International Radar Conference, Leipzig, Sept. 2011

[21] R. Bhatia, *Positive definite matrices*, Princeton university press. 2007

[22] M. Becker and E. L. Stark, *On a hierarchy of quolynomial inequalities for* $\tan x$, Univ. Beograd. Publ. Elektrotehn. Fak. Ser. Mat. Fiz. No. 602-No. 633 (1978), 133-138

[23] A. Benveniste, M. Goursat and G. Ruget, *Analysis of stochastic approximation schemes with discontinuous and dependent forcing terms with applications to data communication algorithm*, IEEE Transactions on Automatic Control, Vol. AC-25, no. 6, December 1980.

[24] E. Berger, *An almost sure invariance principle for stochastic approximation procedures in linear filtering theory*, the Annals of applied probability, vol. 7, no. 2 (May 1997), pp. 444–459.

[25] R. Bhattacharya and V. Patrangenaru, *Large sample theory of intrinsic and extrinsic sample means on manifolds. I*, The Annals of Statistics, 2003, Vol 31, No. 1, 1-29

[26] M. Bridson and A. Haefliger, *Metric spaces of non-positive curvature*, (Springer, Berlin, 1999).

[27] G. Bouyt, *Algorithmes innovants pour la détection de cible radar*, Rapport de travail d'option de Thalès, 2009

[28] J. Burbea and C. R. Rao, *Differntial metrics in probability spaces*, Probability and Mathematical Statistics, Vol. 3, Fasc. 2, pp. 241-258, 1984.

[29] S. R. Buss and J. P. Fillmore, *Spherical averages and applications to spherical splines and interpolation*, ACM Transactions on Graphics vol. 20(2001), pp. 95-126.

[30] E. Cartan, *Groupes simples clos et ouverts et géométrie riemannienne*, J. Math. pures et appl., t. 8, 1929, pp. 1-33.

[31] N. Charon, *Une nouvelle approche pour la détection de cibles dans les images radar*, Rapport du stage de Thalès, 2008

[32] I. Chavel, *Riemannian Geometry: A Modern Introduction*, second edition. Cambridge University Press, New York, 2006.

[33] J. Cheeger and D. G. Ebin, *Comparison theorems in Riemannian geometry*, (North Holland, Amsterdam, 1975).

[34] R. Correa and C. Lemaréchal, *Convergence of some algorithms for convex minimization*, Mathematical Programming, 62 (1993), North-Holland, 261-275.

[35] Z. Drezner and G. O. Wesolowsky, *Facility location on a sphere*, J. Opl Res Soc. Vol. 29, 10, pp. 997-1004.

[36] Z. Drezner, *On location dominance on spherical surfaces*, Operation Research, Vol. 29, No. 6, November-December 1981, pp. 1218-1219.

[37] F. Y. Edgworth, *The choice of means*, Philosophical Magazine, vol. 24, 1887.

[38] M. Emery and G. Mokobodzki, *Sur le barycentre d'une probabilité dans une variété*, Séminaire de Probabilités-XXV, Lecture Notes in Mathematics 1485. (Springer, Berlin, 1991), pp. 220-233.

[39] P. Fermat, *Essai sur les maximas et les minimas, in Œuvres de Fermat (1629)*, Paris: Gauthier-Villars et fils, 1891-1912.

[40] O. P. Ferreira and P. R. Oliveira, *Subgradient Algorithm on Riemannian Manifolds*, Journal of Optimization Theory and Applications, 97 (1998), no. 1, 93-104.

[41] S. Fiori, *Learning the Fréchet mea over the manifold of symmetric positive-definite matrices*, Cogn Comput(2009)1:279-291

[42] S. Fiori and T. Tanaka, *An algorithm to compute averages on matrix Lie groups*, IEEE Transactions on signal processing, vol. 57, No. 12, Decembre 2009.

[43] N. I. Fisher, *Spherical medians*, J. R. Statist. Soc. B (1985), 47, No. 2, pp. 342-348.

[44] R. A. Fisher, *On the mathematical foundations of theoretical statistics*, Philosophical Transactions of the Royal Society of London (A), vol. 222, pp. 309-368 (1922).

[45] P. T. Fletcher and S. Joshi, *Principle geodesic analysis on symmetric spaces: statistics of diffusion tensors*, Proceedings of ECCV Workshop on Computer Vision Approaches to Medical Image Analysis (2004), 87-98.

[46] P. T. Fletcher et al., *Statistics of shape via principle geodesic analysis on Lie groups*, Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. (2003).

[47] P. T. Fletcher et al., *The geometric median on Riemannian manifolds with application to robust atlas estimation*, NeuroImage, 45 (2009), S143-S152.

[48] M. Fréchet, *Les éléments aléatoires de natures quelconque dans un espace distancié*, Annales de l'I.H.P., tome 10, n^{o}4 (1948), p. 215-310.

[49] S. Gouëzel, *Almost sure invariance principle for dynamical systems by spectral methods*, The Annals of Probability, Vol 38, no. 4 (2010), pp. 1639–1671

[50] V. Guillemin and A. Pollack, *Differential Topology*, Prentice-Hall, Englewood Cliffs, NJ, 1974.

[51] I. S. Iohvidov, *Hankel and Toeplitz matrices and forms - Algebraic theory*, Birkhäuser 1982

[52] J. Jost, *Riemannian Geometry and Geometric Analysis*, (Springer, Berlin, 2005).

[53] H. Karcher, *Riemannian center of mass and mollifier smoothing*, Communications on Pure and Applied Mathematics, vol xxx (1977), 509-541.

[54] I. N. Katz and L. Cooper, *Optimal location on a sphere*, Comp. & Maths. with Appls., 6 (1980), 175-196.

[55] W. S. Kendall, *Probability, convexity, and harmonic maps with small image I: uniqueness and fine existence*, Proc. London Math. Soc., (3) 61 (1990), no. 2, 371-406.

[56] W. S. Kendall, *Convexity and the hemisphere*, J. London Math. Soc., (2) 43 (1991), no. 3, 567-576.

[57] R.Z. Khas'Minskii, *On stochastic processes defined by differential equations with a small parameter*, Theory Prob. Appl., vol. XI-2 (1968), pp. 211–228

[58] H. W. Kuhn, *A note on Fermat's problem*, Mathematical Programming, 4 (1973), 98-107, North-Holland Publishing Company.

[59] R. Koenker, *The median is the message: toward the Fréchet median*, Journal de la Société Française de Statistique, tome 147, n°2, 2006.

[60] S. G. Krantz, *Complex analysis: The geometric view point*, The mathematical association of America (1990)

[61] P. S. Laplace, *Mémoire sur la probabilité des causes par les événements (1774)*, Œuvres Compl., 1891, VIII, 27-65 et 141-153.

[62] H. Le, *Locating Fréchet means with application to shape spaces*, Advances in Applied Proability, Vol. 33, No. 2 (Jun., 2001), pp. 324-338.

[63] H. Le, *Estimation of Riemannian barycentres*, LMS J. Comput. Math., 7 (2004), 193-200.

[64] J. M. Lee, *Introduction to smooth manifolds*, 2003 Springer Science+Business Media, Inc.

[65] L. Ljung, *Analysis of recursive stochastic algorithms*, IEEE Transactions on Automatic Control, Vol. AC-22, no. 4, August 1977.

[66] H. P. Lopuhaä and P. J. Rousseeuw, *Breakdown points of affine equivalent estimators of multivariate location and covariance matrices*, The Annals of Statistics, Volume 19, Number 1 (1991), 229-248.

[67] D. L. MCLeish, *Dependent central limit theorems and invariance principles*, The Annals of Probability, vol. 2, no. 4 (1974), pp. 620–628

[68] B. Monjardet, *"Mathématique Sociale" and Mathematics. A case study: Condorcet's effect and medians*, Electronic Journ@l for History of Probability and Statistics, vol 4, n°1; juin/june 2008.

[69] B. N. Mukherjee and S. S. MaitiOn, *some properties of positive definite Toeplitz Matrices and their possible applications*, Linear algebra and its applications, 102: 211-240, 1988

[70] A. Nedic and D. P. Bertsekas, *Convergence Rate of Incremental Subgradient Algorithms*, Stochastic Optimization: Algorithms and Applications (S. Uryasev and P. M. Pardalos, Editors), pp. 263-304. Kluwer Academic Publishers (2000).

[71] A. Nedic and D. P. Bertsekas, *Incremental Subgradient Methods for Nondifferentiable Optimization*, SIAM J. Optim., 12 (2001), no. 1, 109-138.

[72] R. Noda, T. Sakai and M. Morimoto, *Generalized Fermat's problem*, Canad. Math. Bull. Vol. 34(1), 1991, pp. 96-104.

[73] L. M. JR. Ostresh, *On the convergence of a class of iterative methods for solving Weber location problem*, Operation Research, Vol. 26, No. 4, July-August (1978).

[74] X. Pennec, *Intrinsic statistics on Riemannian manifolds: Basic tools for geometric measurements*, Journal of Mathematical Imaging and Vision, 25 (2006), 127-154.

[75] J. Picard, *Barycentres et martingales sur une variété*, Ann. Inst. H. Poincaré Probab. Statist, 30 (1994), no. 4, 647-702.

[76] A. Sahib, *Espérance d'une variable aléatoire à valeur dans un espace métrique*, Thèse de l'Université de Rouen (1998).

[77] T. Sakai, *Riemannian geometry*, Translations of Mathematical Monographs, vol 149. American Mathematical Society (1996).

[78] H. Shima, *The geometry of hessian structures*, World Scientific Publishing (2007).

[79] L. T. SkovgaardA *riemannian geometry of the multivariate normal model*, Scand J Statist, 11: 211-223. 1984

[80] C. G. Small, *Multidimensional medians arising from geodesics on graphs*, The Annals of Statistics, 25 (1997), no. 2, 478-194.

[81] J. Steiner, *Von den Krümmungs-Schwerpuncte ebenen Curven*, Journal fûr die reine und angewante Mathematik, vol 21 (1838), pp. 33-101.

[82] D. W. Stroock and S. R. S. Varadhan, *Multidimensional diffusion processes*, Grundlehren der mathematischen Wissenschaften 233, Springer, 1979.

[83] R. Sturm, *Ueber den Punkt Keinster Entfernugssumme von gegebenen Punkten*, Journal fûr die reine und angewante Mathematik, vol 97 (1884), pp. 49-61.

[84] E. Torricelli, *Opere di Evangelista Torricelli*, G. Loria and G. Vassura (Eds.), Vol. I, 2ème partie, pp. 90-97, Vol. III, pp. 426-431, Faënza, 1919

[85] C. Udriste, *Convex functions and optimization methods on Riemannian manifolds*, Kluwer Academic Publishers (1994).

[86] P. P. Vaidyanathan *The Theory of Linear Prediction*, Morgan and Claypool. 2008

[87] Y. Vardi and C. H. Zhang, *A modified Weiszfeld algorithm for the Fermat-Weber location problem*, Math. Program., Ser. A 90: 559-566 (2001)

[88] C. Villani, *Optimal Transport: Old and New*. Springer-Verlag, 2009.

[89] A. Weber, *Uber den standort der Industrien, Teil I: Reine Theorie des Standorts*, Mohr, Tübingen, 1909

[90] E. Weiszfeld, *Sur le point pour lequel la somme des distances de n points donnés est minimum*, Tohoku Math. J., 43 (1937), 355-386.

[91] L. Yang, *Riemannian median and its estimation*, LMS J. Comput. Math. vol 13 (2010), pp. 461-479.

[92] L. Yang, M. Arnaudon and F. Barbaresco, *Riemannian median, geometry of covariance matrices and radar target detection*, European Radar Conference 2010, pp. 415-418.

[93] L. Yang, M. Arnaudon and F. Barbaresco, *Geometry of covariance matrices and computation of median*, AIP Conf. Proc. Volume 1305 (2011), pp. 479-486.

[94] L. Yang, *Some properties of Fréchet medians in Riemannian manifolds*, preprint (2011), submitted.

[95] F. Z. Zhang, *The Schur complements and its applications*, Springer (2005).