Convergence Rates for the Generalized Fréchet Mean via the Quadruple Inequality

Christof Schötz*

June 13, 2023

Abstract

For sets $\mathcal{Q}$ and $\mathcal{Y}$, the *generalized Fréchet mean* $m \in \mathcal{Q}$ of a random variable Y, which has values in $\mathcal{Y}$, is any minimizer of $q \mapsto \mathbb{E}[\mathfrak{c}(q, Y)]$, where $\mathfrak{c}\colon \mathcal{Q} \times \mathcal{Y} \to \mathbb{R}$ is a cost function. There are little restrictions to $\mathcal{Q}$ and $\mathcal{Y}$. In particular, $\mathcal{Q}$ can be a non-Euclidean metric space. We provide convergence rates for the empirical generalized Fréchet mean. Conditions for rates in probability and rates in expectation are given. In contrast to previous results on Fréchet means, we do not require a finite diameter of the $\mathcal{Q}$ or $\mathcal{Y}$. Instead, we assume an inequality, which we call *quadruple inequality*. It generalizes an otherwise common Lipschitz condition on the cost function. This quadruple inequality is known to hold in Hadamard spaces. We show that it also holds in a suitable way for certain powers of a Hadamard-metric.

Contents

1. Introduction **3**
1.1. Our Contribution 4
1.2. Outline 6

2. Abstract Results **6**
2.1. Setting 7
2.2. Rate of Convergence in Probability 7
2.3. Rate of Convergence in Expectation 10
2.4. Further Extensions 12

3. Quadruple Inequalities **12**
3.1. Bounded Spaces and Smooth Cost Function 13

*Institute of Applied Mathematics, Heidelberg University, Im Neuenheimer Feld 205, 69120 Heidelberg, Germany, schoetz@math.uni-heidelberg.de

3.2. Relation to Inner Product and Cauchy–Schwartz Inequality 14
3.2.1. Inner Product Space 14
3.2.2. Bregman Divergence 14
3.2.3. Hadamard Spaces and Quasi-Inner Product 15
3.3. Power Inequality 16
3.4. Weak Implies Strong 17
3.4.1. Projection Metric 18
3.4.2. Power Metric 18

4. Application of the Abstract Results **19**
4.1. Euclidean Spaces 19
4.2. Hilbert Spaces 20
4.3. Non-Convex Subsets 21
4.4. Hadamard Spaces 23
4.4.1. Fréchet Mean 25
4.4.2. Power Fréchet Mean 26

5. Further Research **28**

Appendix A. Proofs of Theorem 1, 2, and 4 **28**
A.1. Proof of Theorem 1 and Corollary 2 28
A.2. Proof of Theorem 2 31
A.3. Proof of Theorem 4 33

Appendix B. Stability of Quadruple Inequalities **36**

Appendix C. Proof of Lemma 1 **38**

Appendix D. Projection Metric Counter Example **39**

Appendix E. Optimality of Power Inequality **40**

Appendix F. Chaining **41**

Appendix G. Proof of the Power Inequality, Theorem 3 **44**
G.1. Arithmetic Form 44
G.2. First Proof Steps and Outline of the Remaining Proof 46
G.2.1. The Case $|a-c| \leq ra - sc$ 46
G.2.2. The Case $|a-c| \geq ra - sc$ 47
G.2.3. Outline 49
G.3. Tight Power Bound 50
G.4. Merging Lemma 51
G.4.1. Simple Merging Lemma 51
G.4.2. $ra - sc$–Merging Lemma 52
G.4.3. $a - sc$–Merging Lemma 54

G.5. Application of Tight Power Bound and Merging Lemma 56
G.6. The Case $ra - sc \geq |a - c|$ 56
G.6.1. The Case $a \geq c$ 56
G.6.2. The Case $a \leq c$ 58
G.7. The Case $|a - c| \geq ra - sc$ 65

Index of Notations **70**

Index of Assumptions **71**

References **71**

1. Introduction

Let $\mathcal{Q}, \mathcal{Y}$ be sets, Y a $\mathcal{Y}$-valued random variable, and $\mathfrak{c}\colon \mathcal{Y} \times \mathcal{Q} \to \mathbb{R}$ a *cost function*. Every element m of the set $\arg\min_{q \in \mathcal{Q}} \mathbb{E}[\mathfrak{c}(Y, q)]$ is called *generalized Fréchet mean* or $\mathfrak{c}$-*Fréchet mean*. Given independent copies $Y_1, \ldots, Y_n$ of Y, natural estimators of the generalized Fréchet mean are elements m_n of the set $\arg\min_{q \in \mathcal{Q}} \frac{1}{n} \sum_{i=1}^{n} \mathfrak{c}(Y_i, q)$. Our goal is to find suitable conditions for establishing convergence rates for such plug-in estimators.

The described setting generalizes the usual setting for *Fréchet means*, where $\mathcal{Q} = \mathcal{Y}$ is a metric space with metric $d : \mathcal{Q} \times \mathcal{Q} \to [0, \infty)$ and $\mathfrak{c} = d^2$, which has been introduced in [Fré48].

The Fréchet mean has been investigated in many specific settings, often under a different name, e.g., center of mass or barycenter. In the context of Riemannian manifolds, it has been studied – among others – by [BP03]. An asymptotic normality result for generalized Fréchet means on finite dimensional manifolds is shown in [**huckemann18**]. For complete metric spaces of nonpositive curvature, called Hadamard spaces, [Stu03] shows how some classical results of probability theory in Euclidean spaces (e.g., strong law of large numbers, Jensen's inequality) can be transferred to the Fréchet mean setting. An algorithm for calculating Fréchet means in Hadamard spaces is described in [Bač14a].

One important application of statistics in Hadamard spaces is the space phylogenetic trees. A phylogenetic tree represents the genetic relatedness of biological species. The geometry of the space of phylogenetic trees T_m with m leaves is studied in [BHV01]. In particular, it is shown that T_m is a Hadamard space. There has been a lot of recent interest in statistics on T_m. E.g., [BLO18] show a central limit theorem for the Fréchet mean in T_m and [Nye11] apply principal component analysis in that space.

For general metric spaces [Zie77] shows consistency of the Fréchet mean estimator. This is extended to generalized Fréchet means in [Huc11].

The Fréchet mean estimator is a *M-estimator*. Thus, we can build upon many classical and deep results from the M-estimation literature, see, e.g., [VW96; Gee00; Tal14]. Using such M-estimation techniques, rates of convergence in probability for Fréchet means in general bounded metric spaces are obtained in [PM19]; in fact the authors consider a

more complex regression setting. In [DM17] results on the analysis of variance in metric spaces are shown.

Results on convergence rates in expectation, i.e, bounds on $\mathbb{E}[d(m, m_n)^2]$, seem to be rare in the literature on the Fréchet mean. Common are convergence rates in probability or exponential concentration. The latter also implies rates in expectation, but under rather strong assumptions. One publication that establishes rates in expectation more directly, for general cost functions in Euclidean spaces is [BFW17].

The recent article [AGP18] provides nonasymptotic concentration rates in general bounded metric spaces. Its relation to our results will be discussed in the next subsection.

1.1. Our Contribution

Our contribution consists of three parts:

(a) We introduce a condition, which we call *quadruple inequality*, that is used to establish convergence rates in probability and expectation for spaces with infinite diameter, see Theorem 1, Theorem 2, and Theorem 4.

(b) We formulate our results in the setting of the generalized Fréchet mean with a cost-function $\mathfrak{c}$ that is not restricted to being the square of a metric.

(c) We prove a quadruple inequality for exponentiated metrics of Hadamard spaces, Theorem 3. We apply it to obtain rates of convergence for estimators of the Fréchet mean of an exponentiated metric.

[PM19] and [AGP18] show rates of convergence for metric spaces which have a finite diameter (or at least the support of the distribution of observations must be bounded). The proofs in both papers rely on *empirical process theory*. In particular, they make use of *symmetrization* and the *generic chaining* to bound the supremum of an empirical process. But where [AGP18] use that bound to be able to apply *Talagrand's inequality* [Bou02], [PM19] employ a *peeling device* (also called *slicing*; see, e.g., [Gee00]) to obtain rates. As a consequence, [AGP18] achieve stronger results (nonasymptotic exponential concentration instead of $\mathbf{O}_{\mathbb{P}}$-statements), but they rely more heavily on the boundedness of the metric. As our goal is to obtain results for spaces with infinite diameter, our proof technique is closer to [PM19], i.e., we also apply a peeling device.

A law of large numbers, such that the estimator of the Fréchet mean converges in probability to the true value, implies that the estimator eventually is in a subset with finite diameter. Thus, for asymptotic rates in probability as in [PM19], it is not very restrictive to assume a finite diameter. Our motivation to directly deal with infinite diameter comes from our interest in nonasymptotic results and in rates in expectation (asymptotic or nonasymptotic).

As [PM19] and [AGP18], we use the *generic chaining*. Therefore we have entropy bounds as conditions of our theorems. These entropy bounds can be stated by requiring

a bound on the *covering numbers*

$$N(Q,d,r) := \min\left\{ k \in \mathbb{N} \,\middle|\, \exists q_1,\dots,q_k \in \mathcal{Q}\colon Q \subseteq \bigcup_{j=1}^{k} \mathrm{B}_r(q_j) \right\},$$

where $(\mathcal{Q},d)$ is a metric space, $Q \subseteq \mathcal{Q}$, and $r > 0$. To be more precise, in a metric space $(\mathcal{Q},d)$, we require $\log N(\mathrm{B}_\delta(m),d,r) \leq \left(\frac{C\delta}{r}\right)^D$ for some constants $C, D > 0$ and all $0 < r < \delta$, which is the same assumption as in [AGP18]. We note, that this requirement could be weakened by using the optimal bound on Rademacher (or Bernoulli) processes [BL14] at the cost of a more complicated and less comprehensible condition.

In the classical Fréchet mean case, where $(\mathcal{Q},d)$ is a metric space and the cost function is $\mathfrak{c} = d^2$, the empirical process that has to be bounded consists of functions of the form $y \mapsto d(y,q)^2$ for $q \in \mathcal{Q}$. To apply some classical empirical process results, one requires a Lipschitz condition on these functions. In [PM19] and [AGP18] this *Lipschitz condition* is fulfilled by

$$d(y,q)^2 - d(y,p)^2 \leq 2\,\mathsf{diam}(\mathcal{Q})d(q,p) \tag{1}$$

for all $y,q,p \in \mathcal{Q}$. Thus, a finite diameter is required. We show, that one can instead require that

$$d(y,q)^2 - d(y,p)^2 - d(z,q)^2 + d(z,p)^2 \leq 2d(y,z)d(q,p) \tag{2}$$

holds for all $y,z,q,p \in \mathcal{Q}$ and then bound the supremum of the empirical process even if $\mathsf{diam}(\mathcal{Q}) = \infty$. Equation (2) is a special instance of what we call *quadruple inequality*.

Roughly speaking, the transition from Lipschitz to quadruple condition removes certain squared terms and the right hand side by adding and subtracting further squared terms on the left hand side. This is related to the idea of defining the Fréchet mean as minimizer of $q \mapsto \mathbb{E}[d(Y,q)^2 - d(Y,o)^2]$ for an arbitrary fixed point $o \in \mathcal{Q}$ instead of $q \mapsto \mathbb{E}[d(Y,q)^2]$. Then, for existence of the Fréchet mean, only a first moment condition on Y is required instead of a second moment condition, see [Stu03, Acknowledgement to Lutz Mattner].

The inequality (2) does not hold in every metric space. But it characterizes Hadamard spaces among geodesic metric spaces, see [BN08]. In Hadamard spaces, (2) is known as *Reshetnyak's quadruple inequality* [Stu03] or *quadrilateral inequality* [BN08] and can be interpreted as generalization of the Cauchy–Schwartz inequality to metric spaces [BN08]. Note that our results are not restricted to geodesic metric spaces.

In (subsets of) Hadamard spaces $(\mathcal{Q},d)$, we can not only utilize the quadruple inequality with the squared metric d^2 (2). But we show that for d^a with $a \in [1,2]$, we also obtain a version of the quadruple inequality, namely

$$d(y,q)^a - d(y,p)^a - d(z,q)^a + d(z,p)^a \leq 4a2^{-a}d(y,z)^{a-1}d(q,p)\,, \tag{3}$$

for all $y,z,q,p \in \mathcal{Q}$, see Theorem 3. We show that the constant $4a2^{-a}$ is optimal. Similar to equation (1), one can easily show — using the mean value theorem — that

$$d(y,q)^a - d(y,p)^a \leq a\,\mathsf{diam}(\mathcal{Q})^{a-1}d(q,p)$$

for $a > 0$, $q, p, y \in \mathcal{Q}$, where $(\mathcal{Q}, d)$ is an arbitrary metric space. The proof of equation (3) is much more complicated, see appendix Appendix G.

We state our convergence rate results in a general way, where observations live in a space $\mathcal{Y}$ and a cost function $\mathfrak{c}\colon \mathcal{Y} \times \mathcal{Q} \to \mathbb{R}$ is minimized over $\mathcal{Q}$. The quadruple inequality then reads

$$\mathfrak{c}(y, q) - \mathfrak{c}(y, p) - \mathfrak{c}(z, q) + \mathfrak{c}(z, p) \leq \mathfrak{a}(y, z)\mathfrak{b}(q, p)$$

for all $y, z \in \mathcal{Y}$ and $q, p \in \mathcal{Q}$ and an arbitrary function $\mathfrak{a}\colon \mathcal{Y} \times \mathcal{Y} \to [0, \infty)$ and a pseudo-metric $\mathfrak{b}\colon \mathcal{Q} \times \mathcal{Q} \to [0, \infty)$. This general formulation includes, among others, arbitrary bounded metric spaces, Hadamard spaces (including Euclidean and non-Euclidean spaces) with an exponentiated metric d^a, $a \in [1, 2]$, and regression settings with $\mathcal{Q} \neq \mathcal{Y}$, where observations $(x, y) \in \mathcal{Y}$ are described by regression functions $(x \mapsto q(x)) \in \mathcal{Q}$.

Furthermore, some trivial statements in appendix Appendix B show that the quadruple inequality is stable under many operations such as taking subsets, limits, or product spaces.

We prove – via a peeling device – nonasymptotic rates of convergence in probability, Theorem 1. We do not achieve exponential concentration as [AGP18], but our results can be applied in cases where the cost function is not bounded by a finite constant, i.e., in metric spaces with infinite diameter. Furthermore, we show two ways of obtaining rates in expectation: One – nonasymptotic – under the assumption of a stronger version quadruple inequality, Theorem 2; the other – asymptotic – with a stricter entropy condition, Theorem 4.

Aside from the application in Hadamard spaces (including the use of the power inequality, Theorem 3), we illustrate our results in different toy examples: Euclidean spaces and infinite dimensional Hilbert spaces. In (convex subsets of) Hilbert spaces the Fréchet mean is equal to the expectation. Thus, these examples are interesting as a benchmark, because we can compare results from our general Fréchet mean approach to exact results. In two additional examples, we apply our results to nonconvex subsets of Hilbert spaces and to Hadamard spaces.

1.2. Outline

We start by presenting the convergence rates results of Theorem 1 (rates in probability) and Theorem 2 (rates in expectation) in the abstract setting in section 2. The different versions of the quadruple inequality are discussed in section 3, including the power inequality, Theorem 3. This discussion concludes with the statement of Theorem 4 (alternative route to rates in expectation). In section 4, we apply the abstract results in different settings: Euclidean spaces, infinite dimensional Hilbert spaces, nonconvex sets, and Hadamard spaces.

2. Abstract Results

In this section, we prove rates of convergence for the Fréchet mean in a very general setting, see section 2.1. For rates in probability Theorem 1 is stated in section 2.2 and

for rates in expectation Theorem 2 is stated in section 2.3. The proofs can be found in appendix Appendix A. Some remarks on further extensions are given in section 2.4.

2.1. Setting

Here we define an *Abstract Setting* in which we will state our most general results. This setting of the generalized Fréchet mean is similar to what is used in [**huckemann18**; Huc11].

Let $\mathcal{Q}$ be a set, which is called *descriptor space*. Let $(\mathcal{Y}, \Sigma_{\mathcal{Y}})$ be a measurable space, which is called *data space*. Let Y be a $\mathcal{Y}$-valued random variable. Let $\mathfrak{c}\colon \mathcal{Y} \times \mathcal{Q} \to \mathbb{R}$ be a function such that $y \to \mathfrak{c}(y,q)$ is measurable for every $q \in \mathcal{Q}$. We call $\mathfrak{c}$ *cost function*. Define $F\colon \mathcal{Q} \to \mathbb{R}\,, q \mapsto \mathbb{E}[\mathfrak{c}(Y,q)]$, assuming that $\mathbb{E}[|\mathfrak{c}(Y,q)|] < \infty$ for all $q \in \mathcal{Q}$. The function F is called *objective function*. Let $n \in \mathbb{N}$. Let $Y_1, \ldots, Y_n$ be independent copies of Y. Define $F_n\colon \mathcal{Q} \to \mathbb{R}\,, q \mapsto \frac{1}{n}\sum_{i=1}^n \mathfrak{c}(Y_i,q)$. We call F_n *empirical objective function*. Let $\mathfrak{l}\colon \mathcal{Q} \times \mathcal{Q} \to [0,\infty)$ be a function such that $\mathfrak{l}(m,q)$ measures the *loss* of choosing q given that the true value is m.

We want to bound $\mathfrak{l}(m, m_n)$ for $m \in \arg\min_{q\in\mathcal{Q}} F(q)$ and $m_n \in \arg\min_{q\in\mathcal{Q}} F_n(q)$.

2.2. Rate of Convergence in Probability

For our result on convergence rates in probability, we make some assumptions, which are listed in the following. We denote the "closed" ball with center $o \in \mathcal{Q}$ of radius $r > 0$ in the set $\mathcal{Q}$ with respect to an arbitrary distance function $d\colon \mathcal{Q} \times \mathcal{Q} \to [0,\infty)$ as

$$\mathrm{B}_r(o,d) := \{q \in \mathcal{Q}\colon d(o,q) \le r\}\ .$$

Assumptions.

`Existence`:
It holds $\mathbb{E}[|\mathfrak{c}(Y,q)|] < \infty$ for all $q \in \mathcal{Q}$. There are $m_n \in \arg\min_{q\in\mathcal{Q}} F_n(q)$ measurable and $m \in \arg\min_{q\in\mathcal{Q}} F(q)$.

`Growth`:
There are constants $\gamma > 0$ and $c_{\mathsf{g}} > 0$ such that $F(q) - F(m) \ge c_{\mathsf{g}}\mathfrak{l}(m,q)^\gamma$ for all $q \in \mathcal{Q}$.

`Weak Quadruple`:
There are a function $\mathfrak{a}\colon \mathcal{Y} \times \mathcal{Y} \to [0,\infty)$ measurable and a pseudo-metric $\mathfrak{b}\colon \mathcal{Q} \times \mathcal{Q} \to [0,\infty)$, such that, for all $p, q \in \mathcal{Q}$, $y, z \in \mathcal{Y}$, it holds

$$^{\mathfrak{c}}\overline{yq} - {}^{\mathfrak{c}}\overline{zq} - {}^{\mathfrak{c}}\overline{yp} + {}^{\mathfrak{c}}\overline{zp} \ \le\ \mathfrak{a}(y,z)\,\mathfrak{b}(q,p)\,,$$

where we use the notation $^{\mathfrak{c}}\overline{yq} := \mathfrak{c}(y,q)$. We call $\mathfrak{a}$ the *data distance* and $\mathfrak{b}$ the *descriptor metric*.

`Moment`:

Let $\zeta \geq 1$. Set

$$\mathfrak{M}(\zeta) := \begin{cases} \mathbb{E}\Big[\mathfrak{a}(Y', Y)^{\zeta}\Big], & \text{if } \zeta \geq 2\,, \\ \mathbb{E}\big[\mathfrak{a}(Y', Y)^{2}\big]^{\frac{\zeta}{2}}, & \text{if } \zeta \leq 2\,, \end{cases}$$

where Y' is an independent copy of Y. It holds $\mathfrak{M}(\zeta) < \infty$.

`Entropy`:

There are $\alpha, \beta > 0$ with $\frac{\alpha}{\beta} < \gamma$ such that

$$\sqrt{\log N(\mathrm{B}_{\delta}(m, \mathfrak{l}), \mathfrak{b}, r)} \leq c_{\mathsf{e}} \frac{\delta^{\alpha}}{r^{\beta}}$$

for a constant $c_{\mathsf{e}} > 0$ and all $\delta, r > 0$.

Here

$$N(A, \mathfrak{b}, r) = \min\left\{ k \in \mathbb{N} \,\middle|\, \exists q_1, \ldots, q_k \in \mathcal{Q} \colon A \subseteq \bigcup_{j=1}^{k} \mathrm{B}_r(q_j, \mathfrak{b}) \right\},$$

is the *covering number* of $A \subseteq \mathcal{Q}$ with respect to $\mathfrak{b}$-balls $\mathrm{B}_r(\cdot, \mathfrak{b})$ of radius r. `Entropy` is essentially the same condition as in [AGP18], but written down for the setting of the generalized Fréchet mean instead of the classical Fréchet mean in metric spaces.

We shortly discuss other assumptions before stating the theorem for rates of convergence in probability.

The measurability assumptions can be weakened by using the *outer expectation*, see [VW96].

In [BDG07], the `Growth` condition is called *margin condition*. It is called *low noise assumption* in [AGP18]. If `Growth` holds for every distribution of Y and we are in the traditional setting of the (not generalized) Fréchet mean, it implies that the metric space $\mathcal{Q}$ has nonpositive curvature: Assume that $(\mathcal{Q}, d)$ is a complete *geodesic space* [Stu03, Definition 1.1], i.e., every pair of points y_1, y_2 has a *mid-point* m, i.e., $\overline{y_1 m} = \overline{y_2 m} = \frac{1}{2}\overline{y_1 y_2}$, where we use the notation $\overline{qp} := d(q, p)$. Set $\mathcal{Y} = \mathcal{Q}$, $\mathfrak{c} = d^2$, and $\mathfrak{l} = d$. If $\mathbb{P}(Y = y_1) = \mathbb{P}(Y = y_2) = \frac{1}{2}$ with $y_1, y_2 \in \mathcal{Q}$, the Fréchet mean $m \in \mathcal{Q}$ of Y is the mid-point between y_1 and y_2. If we assume that the growth condition holds for every distribution of Y, in particular, for every uniform 2-point distribution, with $c_{\mathsf{g}} = 1$ and $\gamma = 2$, then

$$\frac{1}{2}\overline{y_1 q}^2 + \frac{1}{2}\overline{y_2 q}^2 - \frac{1}{2}\overline{y_1 m}^2 - \frac{1}{2}\overline{y_2 m}^2 \geq \overline{mq}^2\,.$$

As m is the mid-point between y_1 and y_2, we obtain

$$\overline{mq}^2 \leq \frac{1}{2}\overline{y_1 q}^2 + \frac{1}{2}\overline{y_2 q}^2 - \frac{1}{4}\overline{y_1 y_2}^2\,.$$

This inequality implies that the space $(\mathcal{Q}, d)$ has nonpositive curvature [Stu03, Definition 2.1]. Such spaces are called *Hadamard spaces*. Aside from the `Growth` condition they also fulfill the quadruple inequality, which we discuss in section 3.2.3.

The `Weak Quadruple`-condition will be discussed in detail in section 3. Among other things, we will show that it holds in a nice way in all Hadamard spaces, which include the Euclidean spaces.

The following theorem states rates of convergence for the estimator m_n to the true value m measured with respect to the loss function $\mathfrak{l}$.

Theorem 1 (Convergence rate in probability)**.** In the *Abstract Setting* of section 2.1, assume that following conditions hold: `Existence`, `Growth`, `Weak Quadruple`, `Moment`, `Entropy`. Define

$$\eta_{\beta,n} := \begin{cases} n^{-\frac{1}{2}} & \text{for } \beta < 1\,, \\ n^{-\frac{1}{2}} \log(n+1) & \text{for } \beta = 1\,, \\ n^{-\frac{1}{2\beta}} & \text{for } \beta > 1\,. \end{cases}$$

Then, for all $t > 0$, it holds

$$\mathbb{P}\left(\eta_{\beta,n}^{-\frac{1}{\gamma-\frac{\alpha}{\beta}}} \mathfrak{l}(m, m_n) \geq t \right) \leq c\, \mathfrak{M}(\zeta)\, t^{-\zeta(\gamma-\frac{\alpha}{\beta})}$$

where $c > 0$ depends on $\alpha, \beta, \gamma, c_{\mathtt{e}}, c_{\mathtt{g}}, \zeta$.

The proof can be found in appendix Appendix A.

Without loss of generality, one can choose $\gamma = 1$ by using the loss $\mathfrak{l}' = \mathfrak{l}^\gamma$. This is consistent with the result: If `Growth` and `Entropy` are fulfilled with $\mathfrak{l}, \alpha, \beta, \gamma$, then they are also fulfilled with $\mathfrak{l}' = l^\gamma, \alpha' = \frac{\alpha}{\gamma}, \beta' = \beta, \gamma' = 1$, which gives the same result. We keep this redundancy in the parameters of the theorem for convenience.

A more common way of stating rates of convergence in probability is the $\mathbf{O}_{\mathbb{P}}$-notation, as in the following corollary. Note that the $\mathbf{O}_{\mathbb{P}}$-result is asymptotic and, thus, weaker than the non-asymptotic Theorem 1.

Corollary 1. In the *Abstract Setting* of section 2.1, assume that following conditions hold: `Existence`, `Weak Quadruple`, `Growth`, `Moment` with $\zeta = 1$, `Entropy`. Then

$$\mathfrak{l}(m, m_n) = \mathbf{O}_{\mathbb{P}}\left(\eta_{\beta,n}^{\frac{1}{\gamma-\frac{\alpha}{\beta}}} \right)$$

with $\eta_{\beta,n}$ as in Theorem 1.

It is possible to weaken the assumptions in Corollary 1. In particular, we can restrict the `Growth` and `Entropy` conditions to hold only in a neighborhood of m if we already know that $\mathfrak{l}(m_n, m) \in \mathbf{o}_{\mathbb{P}}(1)$.

In Theorem 1, the probability of large losses decays polynomially. If the exponent $\zeta(\gamma - \frac{\alpha}{\beta})$ is strictly greater than 1, we can integrate the tail probabilities to obtain a bound on the expectation of the loss.

Corollary 2. Let $\kappa \geq 1$. In the *Abstract Setting* of section 2.1, assume that following conditions hold: `Existence`, `Weak Quadruple`, `Growth`, `Moment` with $\zeta > \kappa(\gamma - \frac{\alpha}{\beta})^{-1}$, `Entropy`. Set $\xi := \zeta(\gamma - \frac{\alpha}{\beta})\kappa^{-1}$. Then

$$\eta_{\beta,n}^{-\frac{\kappa}{\gamma-\frac{\alpha}{\beta}}} \mathbb{E}[\mathfrak{l}(m, m_n)^\kappa] \leq c' \frac{\xi}{\xi - 1} \mathfrak{M}(\zeta)^{\frac{1}{\xi}} .$$

The proof can be found in appendix Appendix A.

Corollary 2 may require unnecessarily high moments as ξ needs to be *strictly* larger than 1. In the next section, we present a more direct approach to rates in expectation, that requires weaker moment conditions, at least in some settings.

2.3. Rate of Convergence in Expectation

For obtaining rates in expectation directly, we need slightly modified, stronger assumptions.

Assumptions.

`Strong Quadruple`:
Define $\dot{\mathcal{Q}} := \mathcal{Q} \setminus \mathrm{B}_0(m, \mathfrak{l}) = \{q \in \mathcal{Q} \colon \mathfrak{l}(m, q) > 0\}$. There are functions $\mathfrak{b}_m \colon \dot{\mathcal{Q}} \times \dot{\mathcal{Q}} \to [0, \infty)$ (possibly depending on m) and $\mathfrak{a} \colon \mathcal{Y} \times \mathcal{Y} \to [0, \infty)$ with $\mathfrak{a}$ measurable and $\xi \in (0, \gamma)$, such that, for all $p, q \in \dot{\mathcal{Q}}$, $y, z \in \mathcal{Y}$, it holds

$$\frac{{}^{\mathfrak{c}}\overline{yq} - {}^{\mathfrak{c}}\overline{ym} - {}^{\mathfrak{c}}\overline{zq} + {}^{\mathfrak{c}}\overline{zm}}{\mathfrak{l}(m, q)^\xi} - \frac{{}^{\mathfrak{c}}\overline{yp} - {}^{\mathfrak{c}}\overline{ym} - {}^{\mathfrak{c}}\overline{zp} + {}^{\mathfrak{c}}\overline{zm}}{\mathfrak{l}(m, p)^\xi} \leq \mathfrak{a}(y, z)\, \mathfrak{b}_m(q, p) ,$$

Assume that $\mathfrak{b}_m$ is a pseudo-metric on $\dot{\mathcal{Q}}$. We call $\mathfrak{a}$ the *data distance* and $\mathfrak{b}_m$ the *strong quadruple metric* at m.

`Strong Moment`:
For $\zeta > 0$, set

$$\mathfrak{M}(\zeta) := \begin{cases} \mathbb{E}\Big[\mathfrak{a}(Y', Y)^\zeta\Big] , & \text{if } \zeta \geq 2 , \\ \mathbb{E}\big[\mathfrak{a}(Y', Y)^2\big]^{\frac{\zeta}{2}} , & \text{if } \zeta \leq 2 , \end{cases}$$

where Y' is an independent copy of Y. Let $\kappa \geq \gamma - \xi$ and assume $\mathfrak{M}\Big(\frac{\kappa}{\gamma - \xi}\Big) < \infty$.

`Strong Entropy`:

It holds $D := \mathsf{diam}(\dot{\mathcal{Q}}, \mathfrak{b}_m) < \infty$ and there is $\beta > 0$ such that

$$\sqrt{\log N(\dot{\mathcal{Q}}, \mathfrak{b}_m, r)} \leq c_{\mathsf{e}} \left(\frac{D}{r}\right)^{\beta}$$

for all $r \in (0, D)$.

For later use in the application to Hilbert spaces, section 4.2, and for Theorem 2, we state the entropy part of Theorem 2 in a more general way than in Theorem 1. To this end, we need to introduce different measures of entropy.

Definition 1 (Measures of Entropy)**.**

(i) Given a set $\mathcal{Q}$ an *admissible sequence* is an increasing sequence $(\mathcal{A}_k)_{k\in\mathbb{N}_0}$ of partitions of $\mathcal{Q}$ such that $\mathcal{A}_0 = \mathcal{Q}$ and $\mathsf{card}(\mathcal{A}_k) \leq 2^{2^k}$ for $k \geq 1$.

By an increasing sequence of partitions we mean that every set of $\mathcal{A}_{k+1}$ is contained in a set of $\mathcal{A}_k$. We denote by $A_k(q)$ the unique element of $\mathcal{A}_k$ which contains $q \in \mathcal{Q}$.

(ii) Let $(\mathcal{Q}, \mathfrak{b})$ be a pseudo-metric space. Define

$$\gamma_2(\mathcal{Q}, \mathfrak{b}) := \inf \sup_{q\in\mathcal{Q}} \sum_{k=0}^{\infty} 2^{\frac{k}{2}} \,\mathsf{diam}(A_k(q), \mathfrak{b})\,,$$

where the infimum is taken over all admissible sequences in $\mathcal{Q}$ and

$$\mathsf{diam}(A, \mathfrak{b}) := \sup_{q,p\in A} \mathfrak{b}(q, p)$$

for $A \subseteq \mathcal{Q}$.

(iii) Let $(Q, \mathfrak{b})$ be a pseudo-metric space and $n \in \mathbb{N}$. Define

$$\mathsf{entr}_n(Q, \mathfrak{b}) := \inf_{\epsilon>0} \left(\epsilon\sqrt{n} + \int_{\epsilon}^{\infty} \sqrt{\log N(Q, \mathfrak{b}, r)}\mathrm{d}r\right)\,.$$

Items (i) and (ii) are basic definitions form [Tal14]. Item (iii) is just a convenient notation.

Theorem 2 (Convergence rate in expectation)**.** In the *Abstract Setting* of section 2.1, assume that following conditions hold: `Existence`, `Growth`, `Strong Quadruple`,

`Strong Moment`. Then, it holds

$$\mathbb{E}[\mathfrak{l}(m, m_n)^\kappa] \;\leq\; c n^{-\frac{\kappa}{2(\gamma-\xi)}} \mathfrak{M}\left(\frac{\kappa}{\gamma-\xi}\right) \min(\mathsf{entr}_n(\mathcal{Q}, \mathfrak{b}_m), \gamma_2(\mathcal{Q}, \mathfrak{b}_m))^{\frac{\kappa}{\gamma-\xi}} \;,$$

where $c > 0$ depends only on $\kappa, \gamma, \xi, c_{\mathtt{g}}$.

If additionally `Strong Entropy` holds, then

$$\mathbb{E}[\mathfrak{l}(m, m_n)^\kappa] \;\leq\; C\, \mathfrak{M}\left(\frac{\kappa}{\gamma-\xi}\right) D^{\frac{\kappa}{\gamma-\xi}}\, \eta_{\beta,n}^{\frac{\kappa}{\gamma-\xi}} \;,$$

where

$$\eta_{\beta,n} := \begin{cases} n^{-\frac{1}{2}} & \text{for } \beta < 1\,, \\ n^{-\frac{1}{2}} \log(n+1) & \text{for } \beta = 1\,, \\ n^{-\frac{1}{2\beta}} & \text{for } \beta > 1\,, \end{cases}$$

and $C > 0$ depends only on $\kappa, \beta, \gamma, \xi, c_{\mathtt{g}}$.

The proof can be found in appendix Appendix A.

As in Theorem 1 the statement contains some redundancy. E.g., by using the loss $\tilde{\mathfrak{l}} = \mathfrak{l}^\xi$ we set $\xi = 1$ without loss of generality. Then the growth exponent and the resulting rate of convergence will scale accordingly.

2.4. Further Extensions

In general $M := \arg\min_{q\in\mathcal{Q}} \mathbb{E}[\mathfrak{c}(Y, q)]$ is some subset of $\mathcal{Q}$. One can also extend the main theorems of this paper to deal with a the whole set of Fréchet means and Fréchet mean estimators. To do that, the `Growth` condition has to be stated as growth of the minimal distance to M. Furthermore, some of the statements and assumptions made in the theorems and proofs have to be modified so that the hold uniformly over all $m \in M$. Additionally, one has to think about the right notion of convergence for sets. We found that those results hard to read without significantly increasing insight into the problem, which is why we chose to stick with unique Fréchet means and only remark that an extension to Fréchet mean sets is possible.

One can also consider ε-$\arg\min$-sets, i.e., the sets of elements which minimize a function up to an $\varepsilon > 0$. If one chooses $m_n \in \varepsilon_n$-$\arg\min_{q\in\mathcal{Q}} F_n(q)$ with $\varepsilon_n \to 0$ fast enough, the convergence rate is of the same as for the absolute minimizer.

3. Quadruple Inequalities

Recall the definition of the weak and strong quadruple inequalities. Let $(\mathcal{Q}, \mathfrak{b})$ be a pseudo-metric space (*descriptor space* with *descriptor metric*), $(\mathcal{Y}, \Sigma_\mathcal{Y})$ a measurable space (*data space*), $\mathfrak{c}\colon \mathcal{Y} \times \mathcal{Q} \to \mathbb{R}$ such that $y \mapsto \mathfrak{c}(y, q)$ is measurable for every $q \in \mathcal{Q}$ (*cost function*), $\mathfrak{a}\colon \mathcal{Y} \times \mathcal{Y} \to [0, \infty)$ measurable (*data distance*), $m \in \mathcal{Q}$ (reference point, usually the Fréchet mean), $\mathfrak{l}\colon \mathcal{Q} \times \mathcal{Q} \to [0, \infty)$ (*loss*), $\xi > 0$ (*rate parameter*),

$\dot{\mathcal{Q}} = \{q \in \mathcal{Q} \colon \mathfrak{l}(m,q) > 0\}$ $\mathfrak{b}_m \colon \dot{\mathcal{Q}} \times \dot{\mathcal{Q}} \to [0,\infty)$ a pseudo-metric on $\dot{\mathcal{Q}}$ (*strong quadruple metric* at m). We write ${}^{\mathfrak{c}}\overline{yq} := \mathfrak{c}(y,q)$.

(a) The tuple $(\mathcal{Q}, \mathcal{Y}, \mathfrak{c}, \mathfrak{a}, \mathfrak{b})$ fulfills the *(weak) quadruple inequality* if and only if for all $p, q \in \mathcal{Q}$, $y, z \in \mathcal{Y}$ it holds

$$ {}^{\mathfrak{c}}\overline{yq} - {}^{\mathfrak{c}}\overline{zq} - {}^{\mathfrak{c}}\overline{yp} + {}^{\mathfrak{c}}\overline{zp} \leq \mathfrak{a}(y,z)\mathfrak{b}(q,p)\,. $$

(b) The tuple $(\mathcal{Q}, \mathcal{Y}, \mathfrak{c}, \mathfrak{l}^{\xi}, \mathfrak{a}, \mathfrak{b}_m)$ fulfills the *strong quadruple inequality* at $m \in \mathcal{Q}$ if and only if for all $p, q \in \dot{\mathcal{Q}}$, $y, z \in \mathcal{Y}$ it holds

$$ \frac{{}^{\mathfrak{c}}\overline{yq} - {}^{\mathfrak{c}}\overline{ym} - {}^{\mathfrak{c}}\overline{zq} + {}^{\mathfrak{c}}\overline{zm}}{\mathfrak{l}(m,q)^{\xi}} - \frac{{}^{\mathfrak{c}}\overline{yp} - {}^{\mathfrak{c}}\overline{ym} - {}^{\mathfrak{c}}\overline{zp} + {}^{\mathfrak{c}}\overline{zm}}{\mathfrak{l}(m,p)^{\xi}} \leq \mathfrak{a}(y,z)\mathfrak{b}_m(q,p)\,. $$

There are a couple of trivial stability results for quadruple inequalities, see appendix Appendix B.

In section 3.1 we compare the quadruple inequality with a more common Lipschitz property. The simplest advantageous applications of the quadruple inequality are in inner product spaces and quasi-inner product spaces, as is discussed in section 3.2. In section 3.3 we state the power inequality, Theorem 3. It allows to establish quadruple inequalities for exponentiated metrics. We conclude with Theorem 4 in section 3.4, which yields rates of convergence in expectation under the assumption of only a weak quadruple inequality instead of a strong one as in Theorem 2.

3.1. Bounded Spaces and Smooth Cost Function

Let $(\mathcal{Q}, d)$ be a metric space and use the notation $\overline{qp} = d(q,p)$. For obtaining convergence rates in probability for the Fréchet mean estimator, [PM19] use

$$ \overline{yq}^2 - \overline{yp}^2 = (\overline{yq} - \overline{yp})\,(\overline{yq} + \overline{yp}) \leq 2\overline{qp}\,\mathsf{diam}(\mathcal{Q}) $$

for all $q, p, y \in \mathcal{Q}$. In the proof of Theorem 1, we have replaced this bound by the weak quadruple inequality, i.e.,

$$ {}^{\mathfrak{c}}\overline{yq} - {}^{\mathfrak{c}}\overline{yp} - {}^{\mathfrak{c}}\overline{zq} + {}^{\mathfrak{c}}\overline{zp} \leq \mathfrak{a}(y,z)\mathfrak{b}(q,p)\,. $$

This generalizes the results by [PM19] as for bounded metric spaces $(\mathcal{Q}, d)$ and cost function $\mathfrak{c} = d^2$, the weak quadruple inequality holds with $\mathfrak{a}(y,z) = 4\,\mathsf{diam}(\mathcal{Q})$ and $\mathfrak{b} = d$:

$$ \overline{yq}^2 - \overline{yp}^2 - \overline{zq}^2 + \overline{zp}^2 \leq \left|\overline{yq}^2 - \overline{yp}^2\right| + \left|\overline{zq}^2 - \overline{zp}^2\right| \leq 4\overline{qp}\,\mathsf{diam}(\mathcal{Q})\,. $$

More generally, if we can show Lipschitz continuity in the second argument of the cost function, i.e., ${}^{\mathfrak{c}}\overline{yq} - {}^{\mathfrak{c}}\overline{yp} \leq \mathfrak{a}(y)\mathfrak{b}(q,p)$, then the quadruple inequality holds with data distance $\mathfrak{a}(y) + \mathfrak{a}(z)$ and descriptor metric $\mathfrak{b}$. But this might lead to an unnecessarily large bound. We will see in section 3.2.3 that at least for certain metric spaces, we can find a bound via the quadruple inequality that does not involve the diameter of the space and, thus, allows for meaningful results in unbounded spaces.

3.2. Relation to Inner Product and Cauchy–Schwartz Inequality

3.2.1. Inner Product Space

Let $(\mathcal{Q}, d)$ be a metric space such that d comes from an inner product $\langle \cdot\,,\,\cdot \rangle$ on $\mathcal{Q}$, i.e., $\mathcal{Q}$ is a subset of am inner product space and $d(y,q)^2 = \langle y-q\,,\,y-q\rangle$. Use $\mathcal{Y} = \mathcal{Q}$ and the squared metric as cost function, $\mathfrak{c} = d^2$. Then

$$\begin{aligned} {}^{\mathfrak{c}}\overline{yq} - {}^{\mathfrak{c}}\overline{zq} - {}^{\mathfrak{c}}\overline{yp} + {}^{\mathfrak{c}}\overline{zp} &= -2\,\langle y-z\,,\,q-p\rangle \\ &\leq 2\|q-p\|\|y-z\|\,. \end{aligned}$$

Here the Cauchy–Schwartz inequality gives rise to an instance of the weak quadruple inequality. The very general framework that we impose also allows for a more flexible bound: If $\mathcal{Q} \subseteq \mathbb{H}$ is the subset of an infinite dimensional, separable Hilbert space $\mathbb{H}$, we can use a weighted Cauchy–Schwartz inequality: Let $s = (s_k)_{k\in\mathbb{N}} \subseteq (0,\infty)$. Then

$${}^{\mathfrak{c}}\overline{yq} - {}^{\mathfrak{c}}\overline{zq} - {}^{\mathfrak{c}}\overline{yp} + {}^{\mathfrak{c}}\overline{zp} \leq 2\|y-z\|_{s^{-1}}\|q-p\|_s\,,$$

where $\|x\|_s^2 = \sum_{k=1}^{\infty} s_k^2 x_k^2$ with generalized Fourier coefficients $(x_k)_{k\in\mathbb{N}}$ with respect to a fixed orthonormal basis of $\mathbb{H}$.

For the strong quadruple inequality, we set $\xi = 1$, $\mathfrak{l}(q,p) = \|q-p\|$ and obtain

$$\begin{aligned} &\frac{{}^{\mathfrak{c}}\overline{yq} - {}^{\mathfrak{c}}\overline{ym} - {}^{\mathfrak{c}}\overline{zq} + {}^{\mathfrak{c}}\overline{zm}}{\mathfrak{l}(m,q)} - \frac{{}^{\mathfrak{c}}\overline{yp} - {}^{\mathfrak{c}}\overline{ym} - {}^{\mathfrak{c}}\overline{zp} + {}^{\mathfrak{c}}\overline{zm}}{\mathfrak{l}(m,p)} \\ &= -2\left\langle y-z, \frac{q-m}{\|q-m\|} - \frac{p-m}{\|p-m\|}\right\rangle \\ &\leq 2\|y-z\|\left\|\frac{q-m}{\|q-m\|} - \frac{p-m}{\|p-m\|}\right\|\,. \end{aligned}$$

Thus, the strong quadruple inequality hold with $\mathfrak{a}(y,z) = 2\|y-z\|$ and $\mathfrak{b}_m(q,p) = \left\|\frac{q-m}{\|q-m\|} - \frac{p-m}{\|p-m\|}\right\|$. The pseudo-metric $\mathfrak{b}_m$ first projects the points q and p onto the surface of unit ball around m and then measures their Euclidean distance.

The analogous result for the weighted Cauchy–Schwartz inequality is

$$\begin{aligned} &\frac{{}^{\mathfrak{c}}\overline{yq} - {}^{\mathfrak{c}}\overline{ym} - {}^{\mathfrak{c}}\overline{zq} + {}^{\mathfrak{c}}\overline{zm}}{\mathfrak{l}(m,q)} - \frac{{}^{\mathfrak{c}}\overline{yp} - {}^{\mathfrak{c}}\overline{ym} - {}^{\mathfrak{c}}\overline{zp} + {}^{\mathfrak{c}}\overline{zm}}{\mathfrak{l}(m,p)} \\ &\leq 2\|y-z\|_{s^{-1}}\left\|\frac{q-m}{\|q-m\|} - \frac{p-m}{\|p-m\|}\right\|_s\,. \end{aligned}$$

3.2.2. Bregman Divergence

Let $\mathcal{Q} \subseteq \mathbb{R}^r$ be a closed convex set. Let $\psi\colon \mathcal{Q} \to \mathbb{R}$ be a continuously differentiable and strictly convex function. The Bregman divergence $D_\psi\colon \mathcal{Q}\times\mathcal{Q} \to [0,\infty)$ associated with ψ for points $y,q \in \mathcal{Q}$ is defined as $D_\psi(y,q) = \psi(y) - \psi(q) - \langle \nabla\psi(q)\,,\,y-q\rangle$. It is the difference between the value of ψ at point y and the value of the first-order Taylor expansion of ψ around point q evaluated at point y. It is well-known, that the minimizer

m of $q \mapsto \mathbb{E}[D_\psi(Y,q)]$ for a random variable Y with $\mathbb{E}[D_\psi(Y,q)] < \infty$ for all $q \in \mathcal{Q}$ is the expectation $m = \mathbb{E}[Y]$, see [BGW05, Theorem 1]. The Bregman divergence $\mathfrak{c} = D_\psi$ fulfills the weak quadruple inequality:

$$\begin{aligned} D_\psi(y,q) - D_\psi(z,q) - D_\psi(y,p) + D_\psi(z,p) &= \langle \nabla\psi(q) - \nabla\psi(p)\,,\, y - z\rangle \\ &\leq \|\nabla\psi(q) - \nabla\psi(p)\|\|y-z\|\,. \end{aligned}$$

Similarly, we obtain a version of the strong quadruple inequality with $\xi = 1$, $\mathfrak{l}(q,p) = \|q-p\|$,

$$\begin{aligned} &\frac{\mathfrak{c}\overline{yq} - \mathfrak{c}\overline{ym} - \mathfrak{c}\overline{zq} + \mathfrak{c}\overline{zm}}{\mathfrak{l}(m,q)} - \frac{\mathfrak{c}\overline{yp} - \mathfrak{c}\overline{ym} - \mathfrak{c}\overline{zp} + \mathfrak{c}\overline{zm}}{\mathfrak{l}(m,p)} \\ &= \left\langle y - z, \frac{\nabla\psi(q) - \nabla\psi(m)}{\|q-m\|} - \frac{\nabla\psi(p) - \nabla\psi(m)}{\|p-m\|} \right\rangle \\ &\leq \|y-z\| \left\| \frac{\nabla\psi(q) - \nabla\psi(m)}{\|q-m\|} - \frac{\nabla\psi(p) - \nabla\psi(m)}{\|p-m\|} \right\| . \end{aligned}$$

3.2.3. Hadamard Spaces and Quasi-Inner Product

Let $(\mathcal{Q}, d)$ be a metric space. Use the notation $\overline{qp} := d(q,p)$. We use the squared metric as the cost function $\mathfrak{c}(y,q) = d(y,q)^2 = \overline{yq}^2$. One particularly nice version of the weak quadruple inequality with this cost function is

$$\overline{yq}^2 - \overline{yp}^2 - \overline{zq}^2 + \overline{zp}^2 \leq 2\,\overline{yz}\,\overline{qp}\,.$$

Let us call this inequality the *nice quadruple inequality.* As seen before, this holds for subsets of inner product spaces. It also plays an important role for geodesic metric spaces. In this section, we paraphrase some results of [BN08]. In particular, we state that the nice quadruple inequality characterizes CAT(0)-spaces.

Let $(\mathcal{Q}, d)$ be a metric space. A *curve* is a continuous mapping $\gamma\colon [a,b] \to \mathcal{Q}$, where $[a,b]$ is a closed interval. The *length* of the curve $\gamma\colon [a,b] \to \mathcal{Q}$ is

$$L(\gamma) := \sup\left\{ \sum_{i=1}^{I} d(\gamma(t_{i-1}), \gamma(t_i)) \,\middle|\, a = t_0 < t_1 < \cdots < t_I = b, I \in \mathbb{N} \right\}.$$

A curve $\gamma\colon [a,b] \to \mathcal{Q}$ is called a *geodesic* if $L(\gamma) = d(\gamma(a), \gamma(b))$. A metric space is called *geodesic*, if any two points $q, p \in \mathcal{Q}$ can be joined by a geodesic $\gamma\colon [a,b] \to \mathcal{Q}$ with $\gamma(a) = q, \gamma(b) = p$. A *midpoint* of two points $q, p \in \mathcal{Q}$ is a point $m \in \mathcal{Q}$ such that $\overline{qm} = \overline{pm} = \frac{1}{2}\overline{qp}$. A complete metric space is a geodesic space if and only if all pairs of points have a midpoint, see [Stu03, Proposition 1.2]. Now, let $(\mathcal{Q}, d)$ be a geodesic metric space. For any triple of points $a, b, c \in \mathcal{Q}$ one can construct a *comparison triangle* in the Euclidean plane with corners $a', b', c' \in \mathbb{R}^2$, such that $\overline{ab} = \|b' - a'\|$, $\overline{ac} = \|c' - a'\|$, and $\overline{bc} = \|c' - b'\|$. A geodesic metric space $(\mathcal{Q}, d)$ is called *CAT(0)* if and only if for every triple of points $a, b, c \in \mathcal{Q}$ with comparison triangle (a', b', c') following condition holds: For every point d on a geodesic connecting a and b, it holds $\overline{dc} \leq \|c' - d'\|$, where

$d' \in \mathbb{R}^2$ is the point on the edge of the comparison triangle between a' and b' such that $\|d' - a'\| = \overline{ad}$. A complete CAT(0)-space is called *Hadamard space* or *global NPC space* (**n**on**p**ositive **c**urvature).

A metric space $(\mathcal{Q}, d)$ is said to fulfill the *NPC-inequality* if and only if for all $y_1, y_2 \in \mathcal{Q}$ there exists a point $m \in \mathcal{Q}$ such that for all $q \in \mathcal{Q}$ it holds $\overline{mq}^2 \leq \frac{1}{2}\overline{y_1q}^2 + \frac{1}{2}\overline{y_2q}^2 - \frac{1}{4}\overline{y_1y_2}^2$. Then m is the midpoint of y_1 and y_2.

A characterization of CAT(0)-spaces can be found in [Stu03, Section 2]: A metric space is CAT(0) if and only if it fulfills the NPC-inequality.

Another characterization of CAT(0)-spaces by the nice quadruple inequality is given in [BN08, Corollary 3]: A geodesic space is CAT(0) if and only if it fulfills the nice quadruple inequality.

In [BN08], the authors define the *quadrilateral cosine* for $q, p, y, z \in \mathcal{Q}$ as

$$\operatorname{cosq}(\vec{yz}, \vec{qp}) := \frac{\overline{yq}^2 - \overline{yp}^2 - \overline{zq}^2 + \overline{zp}^2}{-2\,\overline{yz}\,\overline{qp}}\,.$$

Obviously, the statement $\operatorname{cosq}(\vec{yz}, \vec{qp}) \leq 1$ for all $q, p, y, z \in \mathcal{Q}$ is equivalent to the nice quadruple inequality. To further motivate this notation and compare it with inner product spaces, they introduce a *quasilinearization* of the metric space and a *quasi-inner product*: Define $\langle \vec{yz}\,,\, \vec{qp} \rangle_d = \operatorname{cosq}(\vec{yz}, \vec{qp})\, \|\vec{yz}\|_d \|\vec{qp}\|_d$, where $\|\vec{yz}\|_d := \overline{yz}$. Thus, the nice quadruple inequality can be viewed as the Cauchy–Schwartz inequality of the quasi-inner product.

3.3. Power Inequality

If the metric space $(\mathcal{Q}, d)$ fulfills the nice quadruple inequality, i.e, $\overline{yq}^2 - \overline{yp}^2 - \overline{zq}^2 + \overline{zp}^2 \leq 2\,\overline{yz}\,\overline{qp}$, where $\overline{yq} = d(y, q)$, then $(\mathcal{Q}, d^a)$, $a \in [\frac{1}{2}, 1]$, also fulfills a weak quadruple inequality with a suitably adapted bound. The implications of this result for the estimators of the corresponding Fréchet means are discussed in section 4.4.2.

According to [DD16], the metric d^a is called *power transform metric* or *snowflake transform metric*.

Theorem 3 (Power Inequality)**.** Let $(\mathcal{Q}, d)$ be a metric space. Use the short notation $\overline{qp} := d(q, p)$. Let $q, p, y, z \in \mathcal{Q}$, $a \in [\frac{1}{2}, 1]$. Assume

$$\overline{yq}^2 - \overline{yp}^2 - \overline{zq}^2 + \overline{zp}^2 \leq 2\,\overline{yz}\,\overline{qp}\,. \tag{4}$$

Then

$$\overline{yq}^{2a} - \overline{yp}^{2a} - \overline{zq}^{2a} + \overline{zp}^{2a} \leq 8a2^{-2a}\,\overline{yz}^{2a-1}\,\overline{qp}\,. \tag{5}$$

In particular, if the metric space $(\mathcal{Q}, d)$ fulfills the nice quadruple inequality and $a \in [\frac{1}{2}, 1]$, then the weak quadruple inequality for $\mathfrak{c} = d^{2a}$ is fulfilled with $\mathfrak{a} = 8a2^{-2a}d^{2a-1}$ and $\mathfrak{b} = d$.

Following the intermediate step Lemma 14 (appendix Appendix G) in the proof of Theorem 3, one can easily show a similar result if the constant on the right hand side

of equation (4) is larger than 2. Only the constant $8a2^{-2a}$ on the right hand side of equation (5) changes.

The theorem applies to subsets of Hadamard spaces. But note that it is not required that $\mathcal{Q}$ is geodesic, but can consist of only the points q, p, y, z. As a statement purely about metric spaces, it might be of interest outside the realm of statistics.

In Corollary 5 (section 3.2.3) it is used to show rates of convergence for the Fréchet mean estimator of the power transform metric d^a. There the asymmetry of the exponents of the factors on the right hand side of (5) is essential for proving the result under weak assumption.

Unfortunately, the only proof of this statement that the author was able to derive (see appendix G) is very long and does not give much insight into the problem as it mostly consists of distinguishing many cases and then using simple calculus. The author is convinced that a more appealing proof is possible.

The concave function $[\frac{1}{2}, 1] \to (0, \infty), a \mapsto 8a2^{-2a}$ is maximal at $a_0 = (2\ln(2))^{-1} \approx 0.721$ with $8a_0 2^{-2a_0} = \frac{4}{\mathrm{e}\ln(2)} \leq 2.123$. Thus, the constant factor in the bound is very close to 2, but 2 is not sufficient.

In appendix Appendix E, we show that $8a2^{-2a}$ is the optimal constant, and that we cannot extend Theorem 3 to $a > 1$ or $a < \frac{1}{2}$.

It is not known to the author whether the nice quadruple inequality in $(\mathcal{Q}, d)$ does or does not imply the nice quadruple inequality in $(\mathcal{Q}, d^a)$ for $a \in (\frac{1}{2}, 1)$, i.e.,

$$\overline{yq}^{2a} - \overline{yp}^{2a} - \overline{zq}^{2a} + \overline{zp}^{2a} \leq 2\,\overline{yz}^a\,\overline{qp}^a\,.$$

3.4. Weak Implies Strong

The weak quadruple inequality is well justified as a condition: Aside from allowing to establish rates in probability (Theorem 1), it can be interpreted as a form of Cauchy–Schwartz inequality (section 3.2.3), it is fulfilled in a large class of metric spaces (bounded metric spaces, Hadamard spaces, appendix Appendix B), and the power inequality (Theorem 3) implies even more applications with a nice interpretation in statistics (section 4.4.2).

The case for the strong quadruple inequality, which we use in Theorem 2 to establish rates in expectation, seems much weaker. Although it can be established in Hilbert spaces, see section 3.2.1, it is not directly clear whether we can have a suitable version for Hadamard spaces or a power inequality.

The next section examines the strong quadruple inequality in Hadamard spaces and concludes with a negative result. Thereafter, we discuss an approach to infer convergence rates in expectation when only assuming the weak quadruple inequality by showing that a weak quadruple inequality imply certain strong quadruple inequalities. This approach is executed to obtain Theorem 4 for convergence rates in expectation, where the result holds only asymptotically, in contrast to Theorem 2.

3.4.1. Projection Metric

In Euclidean spaces, we can take $\mathfrak{b}_m(q,p) = \left\| \frac{q-m}{\|q-m\|} - \frac{p-m}{\|p-m\|} \right\|$ as the strong quadruple metric. This pseudo-metric can be written down only depending on the metric (not the norm or vector space operations) as

$$d_m^{\mathsf{proj}}(q,p) := \sqrt{\frac{\overline{qp}^2 - (\overline{qm} - \overline{pm})^2}{\overline{qm}\,\overline{pm}}}\,, \quad d_m^{\mathsf{proj}}(q,p) = \mathfrak{b}_m(q,p)\,.$$

The metric $d_m^{\mathsf{proj}}(q,p)$ can be defined in any metric space. Unfortunately, it does not yield a strong quadruple inequality in non-Euclidean Hadamard spaces in the same way as in Euclidean spaces. See appendix Appendix D for details.

3.4.2. Power Metric

To establish rates of convergence in expectation for the $\mathfrak{c}$-Fréchet mean, given that a weak quadruple inequality holds, we first show that some version of the strong quadruple inequality is implied by the weak one, Lemma 1. Unfortunately, we obtain a strong quadruple distance $\mathfrak{b}_m$ such that the measure of entropy $\mathsf{entr}(\mathcal{Q}, \mathfrak{b}_m)$ might be infinite. To solve this problem, we define an increasing sequence of sets $\mathcal{Q}_n$ such that $\mathcal{Q}_n \subseteq \mathcal{Q}_{n+1}$ and $\bigcup_{n\in\mathbb{N}} \mathcal{Q}_n = \mathcal{Q}$ with distances $\mathfrak{b}_{m,n}$ such that the strong quadruple inequality is fulfilled on $\mathcal{Q}_n$ with strong quadruple distance $\mathfrak{b}_{m,n}$, and $\mathsf{entr}(\mathcal{Q}_n, \mathfrak{b}_{m,n})$ is finite and can be suitably controlled in n. This allows us to prove an asymptotic result for the rate of convergence in expectation, Theorem 4.

Lemma 1. Assume $(\mathcal{Q}, \mathcal{Y}, \mathfrak{a}, \mathfrak{b}, \mathfrak{c})$ fulfills the weak quadruple inequality. Let $\xi \in [0,1]$. Then

$$\frac{{}^{\mathfrak{c}}\overline{yq} - {}^{\mathfrak{c}}\overline{ym} - {}^{\mathfrak{c}}\overline{zq} + {}^{\mathfrak{c}}\overline{zm}}{\mathfrak{b}(q,m)^{\xi}} - \frac{{}^{\mathfrak{c}}\overline{yp} - {}^{\mathfrak{c}}\overline{ym} - {}^{\mathfrak{c}}\overline{zp} + {}^{\mathfrak{c}}\overline{zm}}{\mathfrak{b}(p,m)^{\xi}} \leq 2^{\xi}\, \mathfrak{a}(y,z)\, \mathfrak{b}(q,p)^{1-\xi} \tag{6}$$

for all $y, z, q, p, m \in \mathcal{Q}$ with $\mathfrak{b}(q,m), \mathfrak{b}(p,m) > 0$.

See appendix Appendix C for a proof. We would like to have ξ large, i.e., close to 1, to obtain the same rate of convergence in expectation as in probability. We achieve that by defining sequences $\xi_n \nearrow 1$ and $\mathcal{Q}_n \nearrow \mathcal{Q}$, and control the entropy of $\mathcal{Q}_n$ with respect to $\mathfrak{b}^{1-\xi_n}$.

To state the result, we have to modify the `Entropy` and the `Existence` condition. Recall the definition of the objective function $F(q) = \mathbb{E}[{}^{\mathfrak{c}}\overline{Yq}]$ and the empirical objective function $F_n(q) = \frac{1}{n}\sum_{i=1}^{n} {}^{\mathfrak{c}}\overline{Y_i q}$.

Assumptions.

`Existence'`:

It holds $\mathbb{E}[|\mathfrak{c}(Y,q)|] < \infty$ for all $q \in \mathcal{Q}$. Let $o \in \mathcal{Q}$. Define $R_n := n$ and $\mathcal{Q}_n := \mathrm{B}_{R_n}(o, \mathfrak{b})$. There are $m_n^{\mathcal{Q}_n} \in \arg\min_{q \in \mathcal{Q}_n} F_n(q)$ measurable and $m \in \arg\min_{q \in \mathcal{Q}} F(q)$.

`Small Entropy`:
There are $\beta, c_{\mathsf{e}} > 0$ such that for $\delta > 0$ large enough

$$\sqrt{\log N(\mathrm{B}_\delta(o, \mathfrak{b}), \mathfrak{b}, r)} \leq c_{\mathsf{e}} \log\left(\frac{\delta}{r}\right)^\beta$$

for all $r > 0$.

Note that the `Small Entropy` condition is much stronger than `Entropy`, which we assumed in Theorem 1. In Euclidean subspaces $\mathcal{Q} \subseteq \mathbb{R}^b$, it holds

$$N(r, \mathrm{B}_\delta(0, d), d) \leq \left(\frac{3\delta}{r}\right)^b$$

for all $R > r > 0$ [Pol90, section 4]. Thus, `Small Entropy` is fulfilled in Euclidean spaces.

Theorem 4 (Convergence rate in expectation)**.** In the *Abstract Setting* of section 2.1 with loss $\mathfrak{l} = \mathfrak{b}$, where $\mathfrak{b}$ is a pseudo-metric, and rate parameter $\xi = 1$, assume that following conditions hold: `Existence'`, `Growth` with $\gamma > 1$, `Weak Quadruple`, `Strong Moment` with $\kappa > \gamma - 1$, `Small Entropy`. Then

$$\mathbb{E}\Big[\mathfrak{b}(m, m_n^{\mathcal{Q}_n})^\kappa\Big] = \mathbf{O}\left(\left(n^{-\frac{1}{2}} \log(n)^\beta\right)^{\frac{\kappa}{\gamma-1}}\right).$$

See appendix Appendix A for the proof.

4. Application of the Abstract Results

We apply the abstract results of Theorems 1 to 4 in this section. We first consider two toy examples – Euclidean spaces, section 4.1 and infinite dimensional Hilbert spaces, section 4.2 – to better understand the result and compare them to optimal bounds. Then we discuss two more involved settings: The Fréchet mean for non-convex subsets of Euclidean spaces, section 4.3, and for Hadamard spaces, section 4.4.

4.1. Euclidean Spaces

Let $\mathcal{Q} \subseteq \mathbb{R}^b$ be convex with the Euclidean metric $d(p,q) = \|p - q\|$. Choose $\mathcal{Y} = \mathcal{Q}$, $\mathfrak{c} = d^2$, $\mathfrak{l} = d$, $\xi = 1$. Let Y be a $\mathcal{Q}$-valued random variable with $\mathbb{E}[\|Y\|^2] < \infty$. Then the Fréchet mean equals the expectation $m = \mathbb{E}[Y] \in \mathcal{Q}$. We can easily calculate

$$\mathbb{E}\Big[\|Y - q\|^2 - \|Y - m\|^2\Big] = \|q - m\|^2.$$

Thus, the `Growth` condition is fulfilled with $\gamma = 2$. The space has the strong quadruple inequality at every point with data distance $\mathfrak{a}(y,z) = 2\|y-z\|$ and strong quadruple distance $\mathfrak{b}_m(p,q) = \left\| \frac{q-m}{\|q-m\|} - \frac{p-m}{\|p-m\|} \right\|$, see section 3.2.1. Thus, Theorem 2 implies

$$\begin{aligned}
\mathbb{E}\left[\left\|\mathbb{E}[Y] - \frac{1}{n}\sum_{i=1}^{n} Y_i\right\|^2\right] &= \mathbb{E}\Big[\mathfrak{l}(m,m_n)^2\Big] \\
&\leq Cn^{-1}\,\mathsf{entr}_n(\mathcal{Q},\mathfrak{b}_m)^2\mathbb{E}\Big[\mathfrak{a}(Y',Y)^2\Big] \\
&\leq C'b\frac{1}{n}\mathbb{E}\Big[\|Y-\mathbb{E}[Y]\|^2\Big]\,,
\end{aligned}$$

where we used $N(r, \mathrm{B}_R(0,d), d) \leq \left(\frac{3R}{r}\right)^b$ for all $R > r > 0$ [Pol90, section 4] to fulfill `Strong Entropy`. The constants $C, C' > 0$ are universal. Compare this with the result that one obtains by direct calculations, i.e.,

$$\mathbb{E}\left[\left\|\mathbb{E}[Y] - \frac{1}{n}\sum_{i=1}^{n} Y_i\right\|^2\right] = \frac{1}{n}\mathbb{E}\Big[\|Y-\mathbb{E}[Y]\|^2\Big]\,.$$

We pay an extra dimension factor b when using the Fréchet mean approach instead of direct calculations. This comes from the use of the Cauchy–Schwartz inequality, which powers the strong quadruple inequality in Euclidean spaces.

4.2. Hilbert Spaces

Let $\mathbb{H}$ be an infinite dimensional Hilbert space and $\mathcal{Q} \subseteq \mathbb{H}$ convex. Let $d(p,q)^2 = \|p-q\|^2 = \langle p-q\,,\,p-q\rangle$. Choose $\mathcal{Y} = \mathcal{Q}$, $\mathfrak{c} = d^2$, $\mathfrak{l} = d$, $\xi = 1$. Let Y be a $\mathcal{Q}$-valued random variable with $\mathbb{E}[\|Y\|^2] < \infty$. As in the Euclidean case, the Fréchet mean m equals the expectation $\mathbb{E}[Y]$, the `Growth` condition holds with $\gamma = 2$, and the strong quadruple inequality is fulfilled with $\mathfrak{a}(y,z) = 2\|y-z\|$ and pseudometric $\mathfrak{b}_m(p,q) = \left\| \frac{q-m}{\|q-m\|} - \frac{p-m}{\|p-m\|} \right\|$.

Unfortunately, `Strong Entropy` is not fulfilled on $\mathbb{H}$ if $\dim(\mathbb{H}) = \infty$. By introducing a weight sequence, we can make $\mathfrak{b}_m$ smaller by making $\mathfrak{a}$ larger: Assume that the Hilbert space $\mathbb{H}$ is separable and thus admits a countable basis. Let $s = (s_k)_{k\in\mathbb{N}} \subseteq (0,\infty)$. In section 3.2.1, we derived that the strong quadruple condition holds with $\mathfrak{a}(y,z) = 2\|y-z\|_{s^{-1}}$ and $\mathfrak{b}_m^s(p,q) = \left\| \frac{q-m}{\|q-m\|} - \frac{p-m}{\|p-m\|} \right\|_s$. Then $\mathsf{entr}_n(\mathbb{H},\mathfrak{b}_m^s) \leq \gamma_2(\mathbb{H},\mathfrak{b}_m^s) \leq \gamma_2(\mathcal{E}_s, d)$, where

$$\mathcal{E}_s = \left\{ h \in \mathbb{H} \colon \sum_{k=1}^{\infty} \frac{h_k^2}{s_k^2} \leq 1 \right\}\,.$$

There is a universal constant $c > 0$ such that $\gamma_2(\mathcal{E}_s, d)^2 \leq c\sum_{k=1}^{\infty} s_k^2$, see [Tal14, Proposition 2.5.1]. As a condition on the variance term, we need

$$\mathbb{E}\Big[\|Y-\mathbb{E}[Y]\|_{s^{-1}}^2\Big] = \|\sigma\|_{s^{-1}}^2 = \sum_{k=1}^{\infty} \sigma_k^2 s_k^{-2} < \infty\,,$$

where $\sigma_k^2 := \mathbb{V}[Y_k]$ and $\sigma = (\sigma_k)_{k\in\mathbb{N}}$. Similar to the Euclidean case, Theorem 2 implies

$$\mathbb{E}\left[\left\|\mathbb{E}[Y] - \frac{1}{n}\sum_{i=1}^{n} Y_i\right\|^2\right] = \mathbb{E}\Big[\mathfrak{l}(m, m_n)^2\Big] \leq C\frac{1}{n}\|s\|_{\ell_2}^2\|\sigma\|_{s^{-1}}^2 ,$$

where $\|s\|_{\ell_2}^2 = \sum_{k=1}^{\infty} s_k^2$.

Direct calculations yield a better result:

$$\mathbb{E}\left[\left\|\mathbb{E}[Y] - \frac{1}{n}\sum_{i=1}^{n} Y_i\right\|^2\right] = \frac{1}{n}\|\sigma\|_{\ell_2}^2 .$$

As in the Euclidean case, we pay a factor related to the dimension for using the more generally applicable Fréchet mean approach instead of using the inner product for direct calculations.

4.3. Non-Convex Subsets

Assume we are in the setting of section 4.2 and the mentioned conditions for convergence are fulfilled. But now we want to take $\mathcal{Q} \subseteq \mathbb{H}$ not necessarily convex and $\mathcal{Y} = \mathbb{H}$. Assume that `Existence` of the Fréchet mean $m \in \mathcal{Q}$ is fulfilled. The expectation $\mu := \mathbb{E}[Y] \in \mathbb{H}$ might not be an element of Q. Then the Fréchet mean m is the closest projection of μ to $\mathcal{Q}$, in the sense that

$$\arg\min_{q\in\mathcal{Q}} \mathbb{E}[\|Y - q\|^2] = \arg\min_{q\in\mathcal{Q}} \|\mu - q\| .$$

To get the same rate as in section 4.2, we mainly need to be concerned with the `Growth` condition, as the quadruple condition holds in all subsets. For $q \in \mathbb{H}$, simple calculations show

$$\mathbb{E}\Big[\|Y - q\|^2 - \|Y - m\|^2\Big] = \|\mu - q\|^2 - \|\mu - m\|^2 .$$

We want to find a lower bound of this term in the form of $c_{\mathtt{g}}\|q - m\|^\gamma$ for constants $\gamma, c_{\mathtt{g}} > 0$. For $a > 1$, it holds

$$\begin{aligned} &a\,\|\mu - q\|^2 - a\,\|\mu - m\|^2 - \|q - m\|^2 \\ &= (a-1)\left\|q - \left(\mu + \frac{\mu - m}{a-1}\right)\right\|^2 - \frac{a^2}{a-1}\,\|m - \mu\|^2 . \end{aligned}$$

Thus, $\|\mu - q\|^2 - \|\mu - m\|^2 \geq \frac{1}{a}\,\|q - m\|^2$ if and only if

$$\left\|q - \left(\mu + \frac{\mu - m}{a-1}\right)\right\| \geq \frac{a}{a-1}\,\|\mu - m\| .$$

Equivalently, the `Growth` condition holds with $\gamma = 2$ and $c_{\mathtt{g}} \in (0, 1)$ if and only if

$$\left\|q - \left(\mu + \frac{c_{\mathtt{g}}}{1 - c_{\mathtt{g}}}\,(\mu - m)\right)\right\| \geq \frac{1}{1 - c_{\mathtt{g}}}\,\|\mu - m\|$$

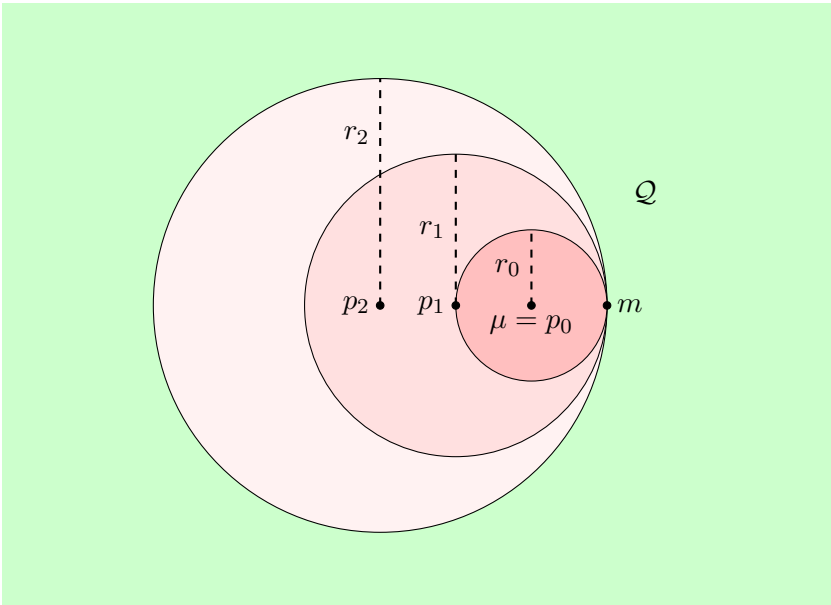

Figure 1: If $\mathcal{Q} \cap \mathrm{B}_{r_0}(p_0) \neq \emptyset$, the point m cannot be the Fréchet mean of a distribution with expectation μ. To fulfill the `Growth` condition, we need $\mathcal{Q} \cap \mathrm{B}_{r_1}(p_1) = \emptyset$ for a ball with larger radius $r_1 > r_0$ and adjusted center p_1. Increasing the radius further, $r_2 > r_1$, only improves the constant $c_{\mathtt{g}}$ of the `Growth` condition, but not the exponent γ.

for all $q \in \mathcal{Q}$, i.e., if and only if $\mathcal{Q} \cap \mathrm{B}_r(p) = \emptyset$, where $r = \frac{1}{1-c_{\mathrm{g}}} \|\mu - m\|$ and $p = \mu + \frac{1-c_{\mathrm{g}}}{c_{\mathrm{g}}} (\mu - m)$. Note that $\|p - m\| = r$. This is illustrated in Figure 1. We have answered the question of how $\mathcal{Q}$ may look like, given the location of μ and m. Possibly more interesting is the question of, given $\mathcal{Q}$, where may μ be located so that m can be estimated with the same rate as for convex sets. We will answer this question only informally via a description similar to a *medial axis transform* [CCM97]:

For simplicity assume $\mathcal{Q} = \mathbb{R}^2 \setminus A$, where A is a nonempty, open, and simply connected set with border ∂A that is parameterized by the continuous function $\gamma\colon [0,1] \to \partial A$. Roll a ball along the border on the inside of A. Make the ball as large as possible at any point so that it is fully contained in A and touches the border at point $\gamma(t)$. Denote the center of the ball as $c\colon [0,1] \to A$ and the radius as $r\colon [0,1] \to [0,\infty)$. Take $\epsilon \in (0,1)$ and trace the point $p_\epsilon\colon [0,1] \to A$ on the radius connecting the center of the ball $c(t)$ and the border $\gamma(t)$ such that it divides the radius into two pieces of length $\overline{p_\epsilon(t)c(t)} = \epsilon r(t)$ and $\overline{p_\epsilon(t)\gamma(t)} = (1-\epsilon)r(t)$. If μ lies on the outside of the set prescribed by $p\colon [0,1] \to A$, it can be estimated with the same rate as for convex sets. This is illustrated in Figure 2. The set of all centers $\mathcal{C} := \{c(t) \,|\, t \in [0,1]\}$, also called the *medial axis* ot *cut locus*, is critical: The closer μ is to $\mathcal{C}$, the larger the guaranteed error bound for the estimator. In particular, we cannot guarantee consistency of the estimator if $\mu \in \mathcal{C}$. A very similar phenomenon is described in [BP03, section 3]. The authors consider a Riemannian manifold $\mathcal{Q}$ that is embedded in an Euclidean space $\mathcal{Y}$. The *extrinsic mean* of a distribution on $\mathcal{Q}$ is the projection of the mean μ in $\mathcal{Y}$ to $\mathcal{Q}$. The points $\mathcal{C}$ are called *focal points*. It is shown [BP03, Theorem 3.3] that in many cases the *intrinsic mean*, i.e, the Fréchet mean in $\mathcal{Q}$ with respect to the Riemannian metric on $\mathcal{Q}$, is equal to the extrinsic mean, i.e, the Fréchet mean in $\mathcal{Q}$ with respect to the Euclidean metric on $\mathcal{Y}$.

The conditions described above are connected to the term reach of a set [Fed59]. The reach of $\mathcal{Q} \subseteq \mathbb{R}^b$ is the largest $\epsilon > 0$ (possibly ∞) such that $\inf_{q \in \mathcal{Q}} d(x,q) < \epsilon$ implies that $x \in \mathbb{R}^b$ has a unique projection to $\mathcal{Q}$, i.e., a unique point $x_{\mathcal{Q}}$ with $d(x, x_{\mathcal{Q}}) = \inf_{q \in \mathcal{Q}} d(x,q)$. If the distance of the mean μ to $\mathcal{Q}$ is less than the reach of $\mathcal{Q}$, then the `Growth` condition holds with $\gamma = 2$. Thus, the rate of convergence is upper bounded by $cn^{-\frac{1}{2}}$ for some $c > 0$. Note that convex sets have infinite reach and exhibit this upper bound for any distribution with finite second moment.

By considering the growth condition $\|\mu - q\|^2 - \|\mu - m\|^2 \geq c_{\mathrm{g}} \|q - m\|^\gamma$, one can also find examples of subspaces where the growth exponent for specific distributions is different from 2.

4.4. Hadamard Spaces

Let $(\mathcal{Q}, d)$ be a Hadamard space. A definition of Hadamard spaces is given in section 3.2.3. Use the notation $\overline{yq} = d(y,q)$. For our purposes the most notable property of Hadamard spaces is that they fulfill the nice quadruple property, i.e., $\overline{yq}^2 - \overline{yp}^2 - \overline{zq}^2 + \overline{zp}^2 \leq 2\,\overline{yz}\,\overline{qp}$. In the following subsections, we will see how this translates to convergence rates for the Fréchet mean estimator and use the power inequality to obtain results for a generalized Fréchet mean with cost function d^{2a} for $a \in [\frac{1}{2}, 1]$.

For an introduction to Hadamard spaces see [Bač14a]. A survey of recent developments

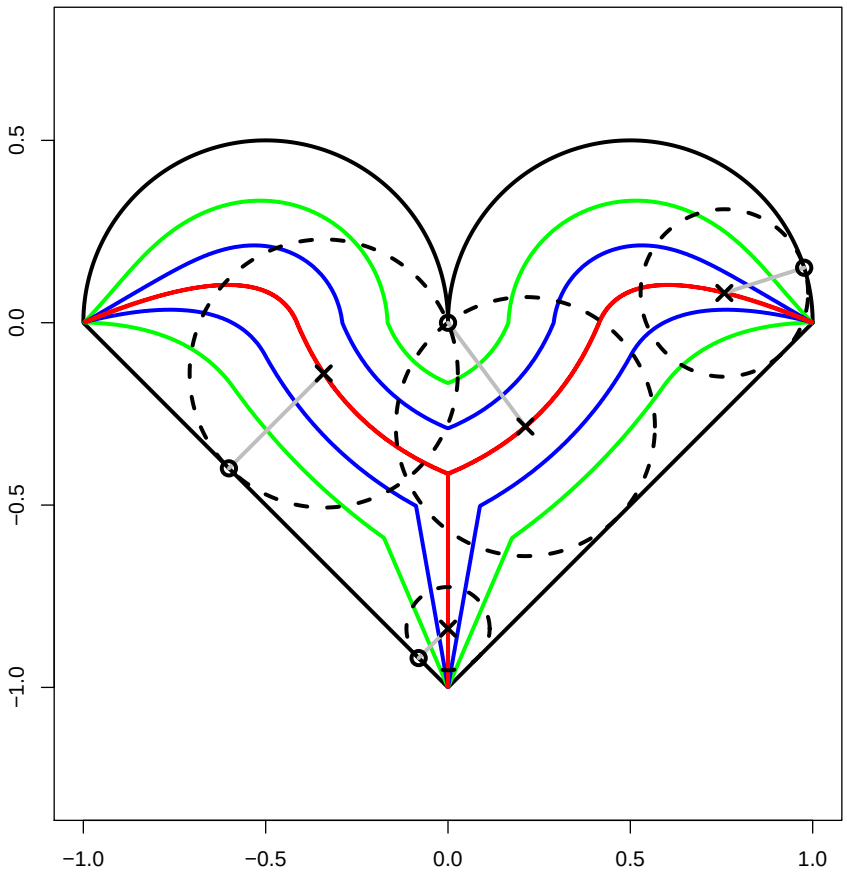

Figure 2: Let $A \subseteq \mathbb{R}^2$ be the set enclosed by the heart (solid black lines). Let $\mathcal{Y} = \mathbb{R}^2$ and $\mathcal{Q} = \mathbb{R}^2 \setminus A$. We consider a distribution on $\mathbb{R}^2$ with mean $\mu \in \mathcal{Y}$ and Fréchet mean $m \in \mathcal{Q}$ with respect to the Euclidean metric and the descriptor space $\mathcal{Q}$. The green, blue, and red lines show $p_\epsilon(t)$ for $\epsilon = 0.6, 0.3, 0$.

can be found in [Bac18]. In [BN08] the authors characterize Hadamard spaces by the nice quadruple inequality and discuss a quasilinearzation of these spaces by observing that the left hand side of the nice quadruple inequality behaves like an inner product to some extend. [Stu03] shows how some important theorems of probability theory in Euclidean spaces, like the law of large numbers and Jensen's inequality, translate to non-Euclidean Hadamard spaces. In [Stu02] martingale theory on Hadamard spaces is discussed.

Turning to more applied topics, [Bač14b] shows algorithms for calculating the Fréchet mean in Hadamard spaces with cost function d^{2a} for $a = \frac{1}{2}$ and $a = 1$. An important application of Hadamard spaces in Bioinformatics are phylogenetic trees [BHV01]. See also [Bac18, section 6.3] for a quick overview. Another application of Hadamard spaces is taking means in the manifold of positive definite matrices, e.g., in diffusion tensor imaging. But note that, as the underlying space is a differentiable manifold, one an use gradient-based approaches, see [PFA06].

Further examples of Hadamard spaces include Hilbert spaces, the Poincaré disc, complete metric trees, complete simply-connected Riemannian manifolds of nonpositive sectional curvature. See also [Stu03, section 3].

4.4.1. Fréchet Mean

Let $(\mathcal{Q}, d)$ be a Hadamard space. We use $\mathcal{Q}$ as data space as well as descriptor space, i.e., $\mathcal{Q} = \mathcal{Y}$. The cost function is $\mathfrak{c} = d^2$, the loss $\mathfrak{l} = d$. As described in section 3.2.3 the weak quadruple inequality holds with $\mathfrak{a} = 2d$ and $\mathfrak{b} = d$, i.e., $(\mathcal{Q}, d)$ fulfills the nice quadruple inequality. Let Y be a random variable with values in $\mathcal{Q}$. Let $Y_1, \dots, Y_n$ be iid copies of Y.

If $\mathbb{E}[d(Y, q)^2] < \infty$ for one $q \in \mathcal{Q}$, then it is also finite for every $q \in \mathcal{Q}$ and the Fréchet mean $m \in \arg\min_{q \in \mathcal{Q}} \mathbb{E}[d(Y, q)^2]$ exists and is unique, see [Stu03, Proposition 4.3]. The same holds true for the estimator $m_n \in \arg\min_{q \in \mathcal{Q}} \sum_{i=1}^n d(Y_i, q)^2$. Thus, `Existence` is fulfilled.

Here, we chose a second moment condition, because we will need it for estimation anyway. But note that choosing the cost function as $\mathfrak{c}(y, q) = d(y, q)^2 - d(y, o)^2$ for a fixed, arbitrary point $o \in \mathcal{Q}$ allows us to require only a finite first moment for `Existence` and the resulting Fréchet mean coincides with the d^2-Fréchet mean if the second moment is finite. This is described in more detail and utilized in [Stu03].

Furthermore, the `Growth`-condition holds in Hadamard spaces with $\gamma = 2$ and $c_{\mathtt{g}} = 1$, see [Stu03, Proposition 4.4]. Thus, we obtain following corollary of Theorem 1.

Corollary 3 (Convergence rate in probability)**.** Assume `Moment` with $\zeta = 2$ and

$\mathfrak{a} = 2d$ and `Entropy` with $\mathfrak{b} = d$ and $\alpha = \beta$. Define

$$\eta_{\beta,n} := \begin{cases} n^{-\frac{1}{2}} & \text{for } \beta < 1\,, \\ n^{-\frac{1}{2}} \log(n+1) & \text{for } \beta = 1\,, \\ n^{-\frac{1}{2\beta}} & \text{for } \beta > 1\,. \end{cases}$$

Then, for all $s > 0$, it holds

$$\mathbb{P}\Big(\eta_{\beta,n}^{-1} d(m, m_n) \geq s\Big) \leq c\,\mathbb{E}[d(Y, Y')^2]\, s^{-2}\,,$$

with a constant $c > 0$ depending only on β and c_{e}. In particular,

$$d(m, m_n) = \mathbf{O}_{\mathbb{P}}(\eta_{\beta,n})\ .$$

As described in section 3.2.3, it may be difficult to find a version of the strong quadruple inequality such that the same rate can be derived for convergence in expectation. Thus, instead of trying to apply Theorem 2, we utilize (i) Corollary 2 and (ii) Theorem 4, respectively.

Corollary 4.

(i) Let $\epsilon > 0$. Assume $\mathbb{E}\big[d(Y, Y')^{2+\epsilon}\big] < \infty$. Assume `Entropy` with $\mathfrak{b} = d$ and $\alpha = \beta < 1$. Then it holds

$$\mathbb{E}\Big[d(m, m_n)^2\Big] \leq c\,\mathbb{E}\Big[d(Y, Y')^{2+\epsilon}\Big]^{\frac{2}{2+\epsilon}} \frac{1}{\epsilon n}$$

for a constant $c > 0$ depending only on β.

(ii) Assume $\mathbb{E}\big[d(Y, Y')^2\big] < \infty$. Let $o \in \mathcal{Q}$. Assume `Small Entropy` with $\mathfrak{b} = d$. Let $\tilde{m}_n \in \arg\min_{q \in \mathrm{B}_n(o)} \sum_{i=1}^n d(Y_i, q)^2$. Then it holds

$$\mathbb{E}\Big[d(m, \tilde{m}_n)^2\Big] = \mathbf{O}\Big(\frac{1}{n} \log(n)^{2\beta}\Big)\ .$$

4.4.2. Power Fréchet Mean

We go beyond Hadamard spaces by utilizing the power inequality, Theorem 3. Let $(\mathcal{Q}, d)$ is a Hadamard space and $a \in [\frac{1}{2}, 1)$. Then $(\mathcal{Q}, d^a)$ is not Hadamard, but fulfills a weak quadruple inequality: Fix an arbitrary point $o \in \mathcal{Q}$. We use the cost function $\mathfrak{c}(y, q) = d^{2a}(y, q) - d^{2a}(y, o)$ and the loss $\mathfrak{l} = d$. Then the weak quadruple inequality holds with $\mathfrak{a}(y, z) = 8a2^{-2a} d(y, z)^{2a-1}$ and $\mathfrak{b} = d$.

We need to choose the cost function $d^{2a}(y, q) - d^{2a}(y, o)$ instead of $d^{2a}(y, q)$ to obtain a result with minimal moment requirement. To fulfill `Moment` we need $\mathbb{E}[d(Y, Y')^{2(2a-1)}] <$

∞ and for `Existence`, we need $\mathbb{E}[|\mathfrak{c}(Y,q)|] < \infty$. We fulfill both by assuming that $\mathbb{E}[d(Y,o)^{2(2a-1)}] < \infty$. Then the both conditions are satisfied: On one hand $\mathbb{E}[d(Y,Y')^{2(2a-1)}] \leq 2\mathbb{E}[d(Y,o)^{2(2a-1)}]$. On the other hand, using the tight power bound of Lemma 17 (appendix Appendix G),

$$\overline{yq}^{2a} - \overline{yo}^{2a} \leq 2a\,\overline{qo}\left(\frac{\overline{yq}+\overline{yo}}{2}\right)^{2a-1}$$

and thus

$$|\mathfrak{c}(Y,q)| \leq 2a\,\overline{qo}\left(\frac{\overline{qo}}{2}+\overline{Yo}\right)^{2a-1},$$

which implies $\mathbb{E}[|\mathfrak{c}(Y,q)|] < \infty$. But $\mathbb{E}[d(Y,q)^{2a}]$ might be infinite as $2a > 2(2a-1)$.

Theorem 1 with $\zeta = 2$ implies following corollary.

Corollary 5 (Rates in probability for power mean)**.** Assume:

(i) `Existence`: Let $a \in [\frac{1}{2}, 1]$. Let $o \in \mathcal{Q}$ be an arbitrary fixed point. Assume there are $m_n \in \arg\min_{q\in\mathcal{Q}} \frac{1}{n}\sum_{i=1}^n \left(\overline{Y_i q}^{2a} - \overline{Y_i o}^{2a}\right)$ measurable and $m \in \arg\min_{q\in\mathcal{Q}} \mathbb{E}[\overline{Yq}^{2a} - \overline{Yo}^{2a}]$.

(ii) `Growth`: There are constants $c_{\mathtt{g}} > 0, \gamma \in (1,\infty)$ such that $\mathbb{E}[\overline{Yq}^{2a}] - \mathbb{E}[\overline{Ym}^{2a}] \geq c_{\mathtt{g}} d(m,q)^{\gamma}$ for all $q \in \mathcal{Q}$.

(iii) `Moment`: $\mathbb{E}[\overline{Yq}^{2(2a-1)}] < \infty$ for one (and thus for all) $q \in \mathcal{Q}$.

(iv) `Entropy`: There is $\beta > 0$ such that

$$\sqrt{\log N(\mathrm{B}_\delta(m,d),d,r)} \leq c_{\mathtt{e}}\left(\frac{\delta}{r}\right)^{\beta}$$

for all $\delta, r > 0$.

Then, for all $s > 0$, it holds

$$\mathbb{P}\left(\eta_{\beta,n}^{-\frac{1}{\gamma-1}}\,\overline{mm_n} \geq s\right) \leq c\,\mathbb{E}[\overline{Yo}^{2(2a-1)}]\,s^{-2(\gamma-1)},$$

where $c > 0$ depends only on $\beta, \gamma, c_{\mathtt{e}}$. In particular,

$$d(m,m_n) = \mathbf{O}_{\mathbb{P}}\left(\eta_{\beta,n}^{-\frac{1}{\gamma-1}}\right).$$

Note that the moment condition becomes weaker as a gets smaller and vanishes for $a = \frac{1}{2}$, where, in the Euclidean case, the Fréchet mean is the median.

`Existence` of m_n and m is a purely technical condition, as one will usually only be able to minimize the objective functions up to an $\epsilon > 0$ and the set of ϵ-minimizers is always nonempty.

The `Growth` condition is more interesting. It seems possible to choose $\gamma = 2$ for all $a \in [\frac{1}{2}, 1]$ in many circumstances, at least under some conditions on the distribution of Y. But precises statements of this sort are unknown to the author. If γ really can be chosen independently of a, then the rate is the same for all $a \in [\frac{1}{2}, 1]$. In the Euclidean case, this is manifested in the fact that we can estimate median ($a = \frac{1}{2}$) and mean ($a = 1$) and all statistics "in between" ($a \in (\frac{1}{2}, 1)$) with the same rate (under some conditions), but with less restrictive moment assumptions for smaller powers a.

Similarly to the corollary above, we can apply Corollary 2 or Theorem 4 to obtain rates in expectation.

5. Further Research

The growth condition, especially for power Fréchet means, see section 4.4.2, needs to be studied further to get a better understanding of what properties a distribution must have, so that all power means can be estimated with the same rate.

In [Bač14a] the author describes algorithms for calculating means and medians in Hadamard spaces, i.e., power Fréchet means as in section 4.4.2 with $a \in \{\frac{1}{2}, 1\}$. As we have shown results also for $a \in (\frac{1}{2}, 1)$, it would be interesting to see, whether one can generalize the algorithms for $a = \frac{1}{2}$ and $a = 1$ to $a \in [\frac{1}{2}, 1]$.

We plan to use the results of this paper in a regression setting similar to [PM19]. We will show convergence rates for an orthogonal series-type regression estimator for the conditional Fréchet mean $m(x) \in \arg\min_{q \in \mathcal{Q}} \mathbb{E}[\mathfrak{c}(Y, q) \mid X = x]$.

Appendix A Proofs of Theorem 1, 2, and 4

A.1 Proof of Theorem 1 and Corollary 2

Define

$$\Delta_n(\delta) := \sup_{q \in \mathcal{Q}\colon \mathfrak{l}(m,q) \le \delta} F(q) - F(m) - F_n(q) + F_n(m)\,.$$

Results similar to following Lemma are well known in the M-estimation literature. The proof relies on the *peeling device*, see [Gee00].

Lemma 2 (Weak argmin transform)**.** Assume `Growth`. Let $\zeta \ge 1$. Assume that there are constants $\xi \in (0, \gamma)$, $h_n \ge 0$ such that $\mathbb{E}[\Delta_n(\delta)^\zeta] \le \left(h_n \delta^\xi\right)^\zeta$ for all $\delta > 0$. Then

$$\mathbb{P}(\mathfrak{l}(m, m_n) \ge s) \le c \left(h_n s^{-(\gamma - \xi)}\right)^\zeta\,,$$

where $c > 0$ depends only on $c_{\mathtt{g}}, \gamma, \xi, \zeta$.

Proof. Let $0 < a < b$. If $\mathfrak{l}(m, m_n) \in [a, b]$, we have

$$c_{\mathtt{g}} a^\gamma \le c_{\mathtt{g}} \mathfrak{l}(m, m_n)^\gamma \le F(m_n) - F(m) \le F(m_n) - F(m) - F_n(m_n) + F_n(m) \le \Delta_n(b)\,.$$

Let $s > 0$. For $k \in \mathbb{N}_0$, set $a_k := s2^k$. It holds

$$\begin{aligned}
&\mathbb{P}(\mathfrak{l}(m, m_n) \ge s) \\
&\le \sum_{k=0}^{\infty} \mathbb{P}(\mathfrak{l}(m, m_n) \in [a_k, a_{k+1}]) \\
&\le \sum_{k=0}^{\infty} \mathbb{P}\big(c_{\mathtt{g}} a_k^\gamma \le \Delta_n(a_{k+1})\big)\,.
\end{aligned}$$

We use Markov's inequality and the bound on $\mathbb{E}[\Delta_n(\delta)^\zeta]$ to obtain

$$\begin{aligned}
\mathbb{P}\big(c_{\mathtt{g}} a_k^\gamma \le \Delta_n(a_{k+1})\big) &\le \frac{\mathbb{E}\Big[\Delta_n(a_{k+1})^\zeta\Big]}{(c_{\mathtt{g}} a_k^\gamma)^\zeta} \\
&\le \left(\frac{h_n a_{k+1}^\xi}{c_{\mathtt{g}} a_k^\gamma}\right)^\zeta \\
&= \Big(2 c_{\mathtt{g}}^{-1} h_n s^{-(\gamma-\xi)} 2^{-k(\gamma-\xi)}\Big)^\zeta\,.
\end{aligned}$$

As $\gamma - \xi > 0$, we get

$$\begin{aligned}
\mathbb{P}(\mathfrak{l}(m, m_n) \ge s) &\le \Big(2 c_{\mathtt{g}}^{-1} h_n s^{-(\gamma-\xi)}\Big)^\zeta \sum_{k=0}^{\infty} 2^{-k\zeta(\gamma-\xi)} \\
&= \Big(2 c_{\mathtt{g}}^{-1} h_n s^{-(\gamma-\xi)}\Big)^\zeta \frac{1}{1 - 2^{-\zeta(\gamma-\xi)}}\,.
\end{aligned}$$

□

Lemma 3. Let $\zeta \ge 1$. Assume `Moment`, `Weak Quadruple`, and `Entropy`. Then

$$\mathbb{E}[\Delta_n(\delta)^\zeta] \le c\,\mathfrak{M}(\zeta)\,\Big(\delta^{\frac{\alpha}{\beta}}\,\eta_{\beta,n}\Big)^\zeta$$

where Y' is an independent copy of Y, $c > 0$ is a constant depending only on $\beta, c_{\mathtt{e}}, \zeta$, and

$$\eta_{\beta,n} := \begin{cases} n^{-\frac{1}{2}} & \text{for } \beta < 1\,, \\ n^{-\frac{1}{2}} \log(n+1) & \text{for } \beta = 1\,, \\ n^{-\frac{1}{2\beta}} & \text{for } \beta > 1\,. \end{cases}$$

Proof. Recall the notation ${}^{\mathfrak{c}}\overline{yq} := \mathfrak{c}(y,q)$, $F(q) = \mathbb{E}[{}^{\mathfrak{c}}\overline{Yq}]$, $F_n(q) = \frac{1}{n}\sum_{i=1}^n {}^{\mathfrak{c}}\overline{Y_i q}$. Define

$$Z_i(q) := \frac{1}{n}\left(\mathbb{E}\left[{}^{\mathfrak{c}}\overline{Yq} - {}^{\mathfrak{c}}\overline{Ym}\right] - {}^{\mathfrak{c}}\overline{Y_i q} + {}^{\mathfrak{c}}\overline{Y_i m}\right).$$

Thus, $\Delta_n(\delta) = \sup_{q\in \mathrm{B}_\delta(m,\mathfrak{l})}\sum_{i=1}^n Z_i(q)$. The `Moment` condition together with the `Weak Quadruple` condition imply that Z_i are integrable. Let $(Z_1', \dots, Z_n')$ be an independent copy of $(Z_1, \dots, Z_n)$, where $(Y_1', \dots, Y_n')$ is an independent copy of $(Y_1, \dots, Y_n)$. By `Weak Quadruple` it holds

$$\begin{aligned}
&n^2\left(Z_i(q) - Z_i(p) - Z_i'(q) + Z_i'(p)\right)^2\\
&= \left({}^{\mathfrak{c}}\overline{Y_i q} - {}^{\mathfrak{c}}\overline{Y_i p} - {}^{\mathfrak{c}}\overline{Y_i' q} + {}^{\mathfrak{c}}\overline{Y_i' p}\right)^2\\
&\le \mathfrak{b}(q,p)^2\mathfrak{a}(Y_i, Y_i')^2\,.
\end{aligned}$$

Furthermore,

$$\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^n \mathfrak{a}(Y_i, Y_i')^2\right)^{\frac{\zeta}{2}}\right] \le \mathfrak{M}(\zeta)\,.$$

Thus, Theorem 6 (appendix Appendix F) implies

$$\mathbb{E}[\Delta_n(\delta)^\zeta] \le c_1\mathfrak{M}(\zeta)\left(\frac{1}{\sqrt{n}}\,\mathsf{entr}_n(\mathrm{B}_\delta(m,\mathfrak{l}),\mathfrak{b})\right)^\zeta,$$

where entr_n is defined in Definition 1.

To bound $\mathsf{entr}_n(\mathrm{B}_\delta(m,\mathfrak{l}),\mathfrak{b})$ by applying Lemma 12 (appendix Appendix F), we need to find an upper bound on $\mathsf{diam}(\mathrm{B}_\delta(m,\mathfrak{l}),\mathfrak{b})$. Set $r_0 := (2c_{\mathsf{e}}\delta^\alpha)^{\frac{1}{\beta}}$. It fulfills $c_{\mathsf{e}}\frac{\delta^\alpha}{r_0^\beta} < \sqrt{\log(2)}$. Thus, `Entropy` implies $N(\mathrm{B}_\delta(m,\mathfrak{l}),\mathfrak{b},r_0) < 2$. As the covering number is an integer, $N(\mathrm{B}_\delta(m,\mathfrak{l}),\mathfrak{b},r_0) = 1$, which implies, $\mathsf{diam}(\mathrm{B}_\delta(m,\mathfrak{l}),\mathfrak{b}) \le 2r_0 =: D_\delta$. Rewriting the `Entropy`-condition in terms of D_δ yields

$$\sqrt{\log N(\mathrm{B}_\delta(m,\mathfrak{l}),\mathfrak{b},r)} \le c_\beta\left(\frac{D_\delta}{r}\right)^\beta$$

for a constant $c_\beta > 0$ depending only on β and c_{e}.

Together with Lemma 12 (appendix Appendix F) we get

$$\mathbb{E}[\Delta_n(\delta)^\zeta] \le c\mathfrak{M}(\zeta)\left(\delta^{\frac{\alpha}{\beta}}\eta_{\beta,n}\right)^\zeta$$

for a constant $c > 0$ depending only on $\beta, c_{\mathsf{e}}, \zeta$. □

Proof of Theorem 1. Combine Lemma 2 and Lemma 3. □

Proof of Corollary 2. Theorem 1 yields

$$\eta_{\beta,n}^{-\frac{\kappa}{\gamma-\frac{\alpha}{\beta}}}\,\mathbb{E}[\mathfrak{l}(m,m_n)^\kappa] = \int_0^\infty \mathbb{P}\left(\eta_{\beta,n}^{-\frac{1}{\gamma-\frac{\alpha}{\beta}}}\,\mathfrak{l}(m,m_n) \geq t^{\frac{1}{\kappa}}\right)\mathrm{d}t$$
$$\leq \int_0^\infty \max\Big(1, c\,\mathfrak{M}(\zeta)\,t^{-\xi}\Big)\,\mathrm{d}t\,.$$

In general for $a > 1$, $b > 0$, we have

$$\int \max(1, bt^{-a})\mathrm{d}t = \frac{a}{a-1}b^{\frac{1}{a}}\,.$$

The proof is concluded by applying this statement and noting that $\xi > 1$. □

A.2 Proof of Theorem 2

To state the next Lemma, which will be used to prove Theorem 2, we introduce an intermediate condition, which we call `Closeness`.

Assumptions.

`Closeness`:
There is $\xi \in (0,\gamma)$ and a random variable $H_n \geq 0$, such that

$$F(q) - F(m) - F_n(q) + F_n(m) \leq H_n\mathfrak{l}(m,q)^\xi \tag{7}$$

for all $q \in \mathcal{Q}$ almost surely.

Lemma 4. Assume `Closeness` and `Growth`, and let $\kappa > 0$. Then,

$$\mathbb{E}[\mathfrak{l}(m,m_n)^\kappa] \leq c\mathbb{E}\left[H_n^{\frac{\kappa}{\gamma-\xi}}\right],$$

where $c > 0$ depends only on $c_{\mathtt{g}}, \gamma, \xi, \kappa$.

Proof. We use `Growth` and the fact that m_n minimizes F_n to obtain

$$\begin{aligned} c_{\mathtt{g}}\mathfrak{l}(m,m_n)^\gamma &\leq F(m_n) - F(m) \\ &\leq F(m_n) - F(m) - F_n(m_n) + F_n(m) \\ &\leq H_n\mathfrak{l}(m,m_n)^\xi\,, \end{aligned}$$

where we applied the `Closeness` condition in the last step. Thus,

$$c_{\mathtt{g}}\mathfrak{l}(m,m_n)^{\gamma-\xi} \leq H_n\,,$$

which implies the claimed inequality. □

Define

$$X(q) := \frac{F_n(q) - F_n(m) - F(q) + F(m)}{\mathfrak{l}(m,q)^{\xi}} .$$

Lemma 5. Let $\zeta \geq 1$. Assume `Strong Moment` and `Strong Quadruple`. Then

$$\mathbb{E}\left[\sup_{q\in\mathcal{Q}} |X(q)|^{\zeta}\right] \leq c n^{-\frac{\zeta}{2}} \mathfrak{M}(\zeta) \min(\mathsf{entr}_n(\mathcal{Q}, \mathfrak{b}_m), \gamma_2(\mathcal{Q}, \mathfrak{b}_m))^{\zeta} ,$$

where $c > 0$ is a constant depending only on ζ. Additionally, assume `Strong Entropy`. Then

$$\mathbb{E}\left[\sup_{q\in\mathcal{Q}} |X(q)|^{\zeta}\right] \leq C\mathfrak{M}(\zeta) D^{\zeta} \eta_{n,\beta}^{\zeta} ,$$

where $C > 0$ is a constant depending only on $\zeta, \beta, c_{\mathsf{e}}$, and

$$\eta_{\beta,n} := \begin{cases} n^{-\frac{1}{2}} & \text{for } \beta < 1 , \\ n^{-\frac{1}{2}} \log(n+1) & \text{for } \beta = 1 , \\ n^{-\frac{1}{2\beta}} & \text{for } \beta > 1 . \end{cases}$$

Proof. Define

$$Z_i(q) := \frac{1}{n} \frac{{}^{\mathfrak{c}}\overline{Y_i q} - {}^{\mathfrak{c}}\overline{Y_i m} - \mathbb{E}\left[{}^{\mathfrak{c}}\overline{Y q} - {}^{\mathfrak{c}}\overline{Y m}\right]}{\mathfrak{l}(m,q)^{\xi}} .$$

Thus, $X(q) = \sum_{i=1}^n Z_i(q)$. The `Strong Moment` condition together with the `Strong Quadruple` condition imply that Z_i integrable. Let $(Z_1', \ldots, Z_n')$ be an independent copy of $(Z_1, \ldots, Z_n)$, where $(Y_1', \ldots, Y_n')$ is an independent copy of $(Y_1, \ldots, Y_n)$. By `Strong Quadruple` it holds

$$\begin{aligned} & n^2 \left(Z_i(q) - Z_i(p) - Z_i'(q) + Z_i'(p)\right)^2 \\ &= \left(\frac{{}^{\mathfrak{c}}\overline{Y_i q} - {}^{\mathfrak{c}}\overline{Y_i m} - {}^{\mathfrak{c}}\overline{Y_i' q} + {}^{\mathfrak{c}}\overline{Y_i' m}}{\mathfrak{l}(m,q)^{\xi}} - \frac{{}^{\mathfrak{c}}\overline{Y_i p} - {}^{\mathfrak{c}}\overline{Y_i m} - {}^{\mathfrak{c}}\overline{Y_i' p} + {}^{\mathfrak{c}}\overline{Y_i' m}}{\mathfrak{l}(m,p)^{\xi}} \right)^2 \\ &\leq \mathfrak{b}_m(q,p)^2 \mathfrak{a}(Y_i, Y_i')^2 . \end{aligned}$$

Furthermore,

$$\mathbb{E}\left[\left(\frac{1}{n}\sum_{i=1}^n \mathfrak{a}(Y_i, Y_i')^2\right)^{\frac{\zeta}{2}}\right] \leq \mathfrak{M}(\zeta)$$

with $\mathfrak{M}(\zeta) < \infty$ due to the assumption `Strong Moment`. Thus, Theorem 6 (appendix

Appendix F) implies

$$\mathbb{E}\left[\sup_{q\in\mathcal{Q}}|X(q)|^{\zeta}\right] \leq cn^{-\frac{\zeta}{2}}\mathfrak{M}(\zeta)\min(\mathsf{entr}_n(\mathcal{Q},\mathfrak{b}_m),\gamma_2(\mathcal{Q},\mathfrak{b}_m))^{\zeta}\ .$$

`Strong Entropy` together with Lemma 12 (appendix Appendix F) yield

$$\mathbb{E}\left[\sup_{q\in\mathcal{Q}}|X(q)|^{\zeta}\right] \leq C\mathfrak{M}(\zeta)\,(D\eta_{n,\beta})^{\zeta}$$

for a constant $C>0$ depending only on $\beta,\zeta,c_{\mathsf{e}}$. □

Proof of Theorem 2. Using $H_n := \sup_{q\in\mathcal{Q}}|X(q)|$ in Lemma 4 fulfills the `Closeness` condition by definition of X. Next, apply Lemma 5 with $\zeta := \frac{\kappa}{\gamma-\xi}$ to conclude the proof. □

A.3 Proof of Theorem 4

Lemma 6. The condition `Small Entropy` implies

$$\mathsf{entr}_n(\mathrm{B}_R(o,\mathfrak{b}),\mathfrak{b}^{1-\xi}) \leq cR^{1-\xi}(1-\xi)^{-\beta}$$

for $\xi\in(0,1)$, where $c>0$ depends only on β,c_{e}.

Proof. Obviously, it holds

$$\mathsf{entr}_n(Q,\mathfrak{b}^{1-\xi}) \leq \int_0^{\infty}\sqrt{\log N(Q,\mathfrak{b}^{1-\xi},r)}\mathrm{d}r$$

for any set $Q\subseteq\mathcal{Q}$. Furthermore,

$$N(Q,\mathfrak{b}^{1-\xi},r) = N(Q,\mathfrak{b},r^{\frac{1}{1-\xi}})\,,$$

which yields

$$\int_0^{\infty}\sqrt{\log N(Q,\mathfrak{b}^{1-\xi},r)}\mathrm{d}r = (1-\xi)\int_0^{\infty}s^{-\xi}\sqrt{\log N(Q,\mathfrak{b},s)}\mathrm{d}s$$

Thus, for $Q := \mathrm{B}_R(o,\mathfrak{b})$, we obtain, using the `Small Entropy` condition,

$$\mathsf{entr}_n(Q,\mathfrak{b}^{1-\xi}) \leq c_{\mathsf{e}}(1-\xi)\int_0^{R}r^{-\xi}\log\left(\frac{R}{r}\right)^{\beta}\mathrm{d}r\,.$$

To calculate the integral, we substitute $s := \frac{r}{R}$ and get

$$\int_0^R r^{-\xi} \log\left(\frac{R}{r}\right)^\beta \mathrm{d}r = R^{1-\xi} \int_0^1 s^{-\xi} \log\left(\frac{1}{s}\right)^\beta \mathrm{d}s\,.$$

For general $a \in (0,1), b > 0$ it holds

$$\int_0^1 x^{-a} \log\left(\frac{1}{x}\right)^b \mathrm{d}x = (1-a)^{-b-1}\Gamma(b+1)\,,$$

where $\Gamma(\cdot)$ is the Gamma function. Thus,

$$\int_0^1 s^{-\xi} \log\left(\frac{1}{s}\right)^\beta \mathrm{d}s \leq c_\beta (1-\xi)^{-\beta-1}$$

for a constant $c_\beta > 0$ depending only on β. Putting everything together, we obtain

$$\mathsf{entr}_n(Q, \mathfrak{b}^{1-\xi}) \leq cR^{1-\xi}(1-\xi)^{-\beta}\,.$$

□

Lemma 7. Set $\xi_n := 1 - \log(n)^{-1}$. Then

$$\left(n^{-\frac{1}{2}}\,(1-\xi_n)^{-\beta}\right)^{\frac{1}{\gamma-\xi_n}} \leq c_\gamma n^{-\frac{1}{2(\gamma-1)}} \log(n)^{\frac{\beta}{\gamma-1}}$$

where $c_\gamma > 0$ is a constant depending only on γ.

Proof. We have

$$\begin{aligned}\left(n^{-\frac{1}{2}}\,(1-\xi_n)^{-\beta}\right)^{\frac{1}{\gamma-\xi_n}} &= \left(n^{-\frac{1}{2(\gamma-\xi_n)}}\right) \log(n)^{\frac{\beta}{\gamma-\xi_n}} \\ &= \left(n^{-\frac{1}{2\left(z+\frac{1}{\log(n)}\right)}}\right) \log(n)^{\frac{\beta}{z+\frac{1}{\log(n)}}}\,,\end{aligned}$$

where $z = \gamma - 1$. We use

$$\log(n)^{\frac{\beta}{z+\frac{1}{\log(n)}}} \leq \log(n)^{\frac{\beta}{z}}\,,$$

$$n^{-\frac{1}{2(z+\frac{1}{\log(n)})}} = n^{-\frac{\log(n)}{2z\log(n)+2}} = \exp\left(-\frac{\log(n)^2}{2z\log(n)+2} + \frac{\log(n)}{2z}\right) n^{-\frac{1}{2z}}\,,$$

and

$$-\frac{\log(n)^2}{2z\log(n)+2}+\frac{\log(n)}{2z}=\frac{\log(n)}{2z(z\log(n)+1)}$$
$$\leq\frac{1}{2z^2}\,,$$

to obtain

$$\left(n^{-\frac{1}{2}}\,(1-\xi_n)^{-\beta}\right)^{\frac{1}{\gamma-\xi_n}}\leq\exp\left(\frac{1}{2z^2}\right)n^{-\frac{1}{2z}}\log(n)^{\frac{\beta}{z}}\,.$$

□

Proof of Theorem 4. For $n\in\mathbb{N}, n\geq 3$, set $\xi_n:=1-\log(n)^{-1}$, $Q_n:=\mathrm{B}_{R_n}(o,\mathfrak{b})$, and $R_n:=n$. For n large enough, the `Existence'` condition implies the existence of $m_n^{Q_n}\in\arg\min_{q\in Q_n}F_n(q)$ and $m^{Q_n}\in\arg\min_{q\in Q_n}F(q)$.

Theorem 2 implies

$$\mathbb{E}\Big[\mathfrak{b}(m^{Q_n},m_n^{Q_n})^\kappa\Big]\leq Cn^{-\frac{\kappa}{2(\gamma-\xi_n)}}\,\mathsf{entr}_n(Q_n,\mathfrak{b}^{1-\xi_n})^{\frac{\kappa}{\gamma-\xi_n}}\,\mathfrak{M}\left(\frac{\kappa}{\gamma-\xi_n}\right)\,,$$

for n large enough. Note, that $C>0$ can be chosen independently of n (even for ξ_n depending on n).

In `Strong Moment` we require $\kappa\geq\gamma-1$, because then $x\mapsto x^{\frac{\kappa}{\gamma-1}}$ is convex, which is needed for the symmetrization argument in the proof of Theorem 2. But, if $\kappa=\gamma-1$, then $\frac{\kappa}{\gamma-\xi_n}<1$, and Theorem 2 cannot be applied directly. For this technical reason, we assumed $\kappa>\gamma-1$, so that $\kappa\geq\gamma-\xi_n$ for n large enough.

By `Small Entropy` and Lemma 6 there is $c_\beta>0$ such that for $n\in\mathbb{N}$ large enough, it holds

$$\mathsf{entr}_n(\mathrm{B}_{R_n}(o,\mathfrak{b}),\mathfrak{b}^{1-\xi_n})\leq c_\beta R_n^{1-\xi_n}\,(1-\xi_n)^{-\beta}\,.$$

Using $R_n^{1-\xi_n}=n^{\frac{1}{\log(n)}}=\exp(1)$ together with Lemma 7, we obtain

$$\mathbb{E}\Big[\mathfrak{b}(m^{Q_n},m_n^{Q_n})^\kappa\Big]\leq C'\left(n^{-\frac{1}{2}}\log(n)^\beta\right)^{-\frac{\kappa}{\gamma-1}}\,\mathfrak{M}\left(\frac{\kappa}{\gamma-\xi_n}\right)\,.$$

As $\lim_{n\to\infty}\mathfrak{M}\left(\frac{\kappa}{\gamma-\xi_n}\right)=\mathfrak{M}\left(\frac{\kappa}{\gamma-1}\right)$, we have

$$\mathbb{E}\Big[\mathfrak{b}(m^{Q_n},m_n^{Q_n})^\kappa\Big]\leq C''n^{-\frac{\kappa}{2(\gamma-\xi_n)}}\left(R_n^{1-\xi_n}\,(1-\xi_n)^{-\beta}\right)^{\frac{\kappa}{\gamma-\xi_n}}\,\mathfrak{M}\left(\frac{\kappa}{\gamma-1}\right)\,.$$

Finally, there is a $n_0\in\mathbb{N}$ such that for all $n\geq n_0$ it holds $m\in Q_n$, which implies $m=m^{Q_n}$. Thus,

$$\mathbb{E}\Big[\mathfrak{b}(m,m_n^{Q_n})^\kappa\Big]=\mathbf{O}\left(\left(n^{-\frac{1}{2}}\log(n)^\beta\right)^{-\frac{\kappa}{\gamma-1}}\right)\,.$$

□

Appendix B Stability of Quadruple Inequalities

We present some trivial stability results for quadruple inequalities. The notation we use here is introduced in the beginning of section 3.

Subsets:
If $(\mathcal{Q}, \mathcal{Y}, \mathfrak{c}, \mathfrak{a}, \mathfrak{b})$ fulfills the weak quadruple inequality, then so does $(\mathcal{Q}', \mathcal{Y}', \mathfrak{c}, \mathfrak{a}, \mathfrak{b})$ with $\mathcal{Q}' \subseteq \mathcal{Q}$, $\mathcal{Y}' \subseteq \mathcal{Y}$.

Images:
Assume $(\mathcal{Q}, \mathcal{Y}, \mathfrak{c}, \mathfrak{a}, \mathfrak{b})$ fulfills the weak quadruple inequality and $f\colon \mathcal{Y}' \to \mathcal{Y}$, $g\colon \mathcal{Q}' \to \mathcal{Q}$. Then $(\mathcal{Q}', \mathcal{Y}', \mathfrak{c}', \mathfrak{a}', \mathfrak{b}')$ fulfills the weak quadruple inequality with $\mathfrak{c}'(y, q) = \mathfrak{c}(f(y), g(q))$, $\mathfrak{a}'(y, z) = \mathfrak{a}(f(y), f(z))$, $\mathfrak{b}'(q, p) = \mathfrak{b}(g(q), g(p))$.

Limits:
Let $(\mathcal{Q}, \mathcal{Y}, \mathfrak{c}_i, \mathfrak{a}_i, \mathfrak{b}_i)$ fulfill the weak quadruple inequality for $i \in \mathbb{N}$ and assume for all $q, p \in \mathcal{Q}$ and $y, z \in \mathcal{Y}$ the point-wise limits

$$\begin{aligned}
\mathfrak{a}(y, z) &:= \lim_{i\to\infty} \mathfrak{a}_i(y, z) \\
\mathfrak{b}(q, p) &:= \lim_{i\to\infty} \mathfrak{b}_i(q, p) \\
\mathfrak{c}(y, q) &:= \lim_{i\to\infty} \mathfrak{c}_i(y, q)
\end{aligned}$$

exist. Then $(\mathcal{Q}, \mathcal{Y}, \mathfrak{c}, \mathfrak{a}, \mathfrak{b})$ also fulfills the weak quadruple inequality.

Similar results hold for the strong quadruple inequality. For the following results it may not be so easy to obtain an analog for the strong quadruple inequality.

Product Spaces:
If $(\mathcal{Q}_i, \mathcal{Y}_i, \mathfrak{c}_i, \mathfrak{a}_i, \mathfrak{b}_i)$ fulfill the weak quadruple inequality for all $i \in \mathbb{N}$, then so does $(\mathcal{Q}, \mathcal{Y}, \mathfrak{c}, \mathfrak{a}, \mathfrak{b})$ where $\mathcal{Q} = \bigtimes_{i\in\mathbb{N}} \mathcal{Q}_i$, $\mathcal{Y} = \bigtimes_{i\in\mathbb{N}} \mathcal{Y}_i$, $\mathfrak{c} = \sum_{i=1}^{\infty} \mathfrak{c}_i$, $\mathfrak{a} = \|(\mathfrak{a}_i)_{i\in\mathbb{N}}\|_{\ell^2}$, $\mathfrak{b} = \|(\mathfrak{b}_i)_{i\in\mathbb{N}}\|_{\ell^2}$.

Proof. It holds

$$\begin{aligned}
{}^{\mathfrak{c}}\overline{yq} - {}^{\mathfrak{c}}\overline{zq} - {}^{\mathfrak{c}}\overline{yp} + {}^{\mathfrak{c}}\overline{zp} &= \sum_{i=1}^{\infty} \left({}^{\mathfrak{c}_i}\overline{y_i q_i} - {}^{\mathfrak{c}_i}\overline{z_i q_i} - {}^{\mathfrak{c}_i}\overline{y_i p_i} + {}^{\mathfrak{c}_i}\overline{z_i p_i}\right) \\
&\le \sum_{i=1}^{\infty} \mathfrak{a}_i(y_i, z_i)\mathfrak{b}_i(q_i, p_i) \\
&\le \mathfrak{a}(y, z)\mathfrak{b}(q, p)\,,
\end{aligned}$$

using the Cauchy–Schwartz inequality. □

Measure Spaces:
Let $(\Omega, \mathcal{A}, \mu)$ be a measure space. Assume $(\mathcal{Q}, \mathcal{Y}, \mathfrak{c}(\omega), \mathfrak{a}(\omega), \mathfrak{b}(\omega))$ fulfills the weak quadruple inequality for every $\omega \in \Omega$. Let $s, t > 0$ with $\frac{1}{s} + \frac{1}{t} = 1$. Let $L(\Omega, \mathcal{Q})$

be the set of measurable functions from Ω to $\mathcal{Q}$, define $L(\Omega, \mathcal{Y})$ analogously. For $q, p \in L(\Omega, \mathcal{Q})$, $y, z \in L(\Omega, \mathcal{Y})$, let

$$\begin{aligned}
\mathfrak{C}(y, q) &:= \int \mathfrak{c}(\omega, y(\omega), q(\omega))\, \mathrm{d}\mu(\omega)\,, \\
\mathfrak{A}(y, z) &:= \left(\int \mathfrak{a}(\omega; y(\omega), z(\omega))^t \, \mathrm{d}\mu(\omega) \right)^{\frac{1}{t}}\,, \\
\mathfrak{B}(q, p) &:= \left(\int \mathfrak{b}(\omega; q(\omega), p(\omega))^s \, \mathrm{d}\mu(\omega) \right)^{\frac{1}{s}}\,,
\end{aligned}$$

where we implicitly assume that the necessary measurablity and integrability conditions are fulfilled. Then

$$(L(\Omega, \mathcal{Q}), L(\Omega, \mathcal{Y}), \mathfrak{C}, \mathfrak{A}, \mathfrak{B})$$

also fulfills the quadruple inequality.

Proof. It holds

$$\begin{aligned}
&\mathfrak{C}(y, q) - \mathfrak{C}(z, q) - \mathfrak{C}(y, p) + \mathfrak{C}(z, p) \\
&= \int \mathfrak{c}(\omega; y(\omega), q(\omega)) - \mathfrak{c}(\omega; y(\omega), p(\omega)) \\
&\qquad - \mathfrak{c}(\omega; z(\omega), q(\omega)) + \mathfrak{c}(\omega; z(\omega), p(\omega)) \mathrm{d}\mu(\omega) \\
&\leq \int \mathfrak{a}(\omega; y(\omega), z(\omega)) \mathfrak{b}(\omega; q(\omega), p(\omega)) \mathrm{d}\mu(\omega) \\
&\leq \mathfrak{A}(y, z) \mathfrak{B}(q, p)\,,
\end{aligned}$$

by Hölder's inequality. □

Minima:

Let $(\mathcal{Q}, \mathcal{Y}, \mathfrak{c}, \mathfrak{a}, \mathfrak{b})$ fulfill the weak quadruple inequality. Let $\tilde{\mathcal{Y}} \subseteq 2^{\mathcal{Y}}$. Define the cost function $\mathfrak{C}\colon \tilde{\mathcal{Y}} \times \mathcal{Q} \to \mathbb{R}$ by $\mathfrak{C}(\mathbf{y}, q) = \inf_{y \in \mathbf{y}} \mathfrak{c}(y, q)$ and $\mathfrak{A}(\mathbf{y}, \mathbf{z}) = \sup_{y \in \mathbf{y}, z \in \mathbf{z}} \mathfrak{a}(y, z)$ assuming the infinma and suprema are finite. Then $(\mathcal{Q}, \tilde{\mathcal{Y}}, \mathfrak{C}, \mathfrak{A}, \mathfrak{b})$ fulfills the weak quadruple inequality.

Proof. Let $\mathbf{y}, \mathbf{z} \in \tilde{\mathcal{Y}}$ and $q, p \in \mathcal{Q}$. Assume there are $y_q, y_p \in \mathbf{y}$, $z_q, z_p \in \mathbf{z}$ such that $\mathfrak{C}(\mathbf{y}, q) = {}^{\mathfrak{c}}\overline{y_q q}$, $\mathfrak{C}(\mathbf{y}, p) = {}^{\mathfrak{c}}\overline{y_p p}$, $\mathfrak{C}(\mathbf{z}, q) = {}^{\mathfrak{c}}\overline{z_q q}$, and $\mathfrak{C}(\mathbf{z}, p) = {}^{\mathfrak{c}}\overline{z_p p}$. Then

$$\begin{aligned}
\mathfrak{C}(\mathbf{y}, q) - \mathfrak{C}(\mathbf{y}, p) - \mathfrak{C}(\mathbf{z}, q) + \mathfrak{C}(\mathbf{z}, p) &= {}^{\mathfrak{c}}\overline{y_q q} - {}^{\mathfrak{c}}\overline{y_p p} - {}^{\mathfrak{c}}\overline{z_q q} + {}^{\mathfrak{c}}\overline{z_p p} \\
&\leq {}^{\mathfrak{c}}\overline{y_p q} - {}^{\mathfrak{c}}\overline{y_p p} - {}^{\mathfrak{c}}\overline{z_q q} + {}^{\mathfrak{c}}\overline{z_q p} \\
&\leq \mathfrak{a}(y_p, z_q) \mathfrak{b}(q, p) \\
&\leq \mathfrak{A}(\mathbf{y}, \mathbf{z}) \mathfrak{b}(q, p)\,.
\end{aligned}$$

If the infima are not attained, one can follow the same proof with minimizing

sequences. □

In many interesting problems the setting is opposite to what was described before, i.e., $\mathfrak{C}\colon \mathcal{Y} \times \tilde{\mathcal{Q}} \to \mathbb{R}$, $(y, \mathbf{q}) \mapsto \inf_{q \in \mathbf{q}} \mathfrak{c}(y, q)$, where $\tilde{\mathcal{Q}} \subseteq 2^{\mathcal{Q}}$: the elements of the descriptor space are subsets and the elements of data space are points. Examples are k-means, where $\tilde{\mathcal{Q}}$ consists of k-tuples of points in $\mathcal{Q}$, or fitting hyperplanes. A quadruple inequality with $\sup_{q \in \mathbf{q}, p \in \mathbf{p}} \mathfrak{b}(q, p)$ as the descriptor distance can be established. Unfortunately, this is usually not useful, as the entropy condition cannot be fulfilled with distances of this type. The framework described in this article can still be applied using inequalities as for bounded spaces, see section 3.1. But we cannot directly use the advantage of quadruple inequalities over Lipschitz-continuity.

Appendix C Proof of Lemma 1

We first state and prove two simple lemmas for some simple arithmetic expressions and then use those for the proof of Lemma 1.

Lemma 8. Let $A, B \in \mathbb{R}$, $a, b, c, r \geq 0$, $s, t > 0$. Assume $t \geq s \;\Leftrightarrow\; b \geq a$. Assume $|A| \leq ra$, $|B| \leq rb$, $|A - B| \leq rc$. Then

$$\left| \frac{A}{s} - \frac{B}{t} \right| \leq r \frac{\min(s,t)c + |s-t| \min(a,b)}{st} .$$

Proof. For $t \geq s$, using the bound on A and on $A - B$ implies $\frac{A}{s} - \frac{B}{t} \leq r \frac{(t-s)a + sc}{st}$. Similarly, for $s \geq t$, we get $\frac{A}{s} - \frac{B}{t} \leq r \frac{tc + (s-t)b}{st}$ by using the bound on B and $A - B$. Together, we obtain

$$\frac{A}{s} - \frac{B}{t} \leq r \frac{\min(s,t)c + |s-t| \min(a,b)}{st} .$$

We finish the proof by pointing out the symmetry between (A, a, s) and (B, b, t). □

Lemma 9. Let $a, b, c > 0$, $\beta \in [0, 1]$. Assume $a \leq b$, $b \leq a + c$, $c \leq a + b$. Then

$$\frac{ca^\beta + \left(b^\beta - a^\beta\right) a}{a^\beta b^\beta} \leq 2^\beta c^{1-\beta} .$$

Proof. The statement is trivial for $\beta \in \{0, 1\}$. So let $\beta \in (0, 1)$.
Case I, $c \leq a$: Define $f(x) = 1 - x - (1 + x)^\beta (1 - x^{1-\beta})$. Then $f(0) = f(1) = 0$ and $f''(x) = -(1 - \beta)\beta x^{-\beta-1}(x + 1)^{\beta-2}(1 - x^{\beta+1}) \leq 0$ for $x \in (0, 1)$. Thus, $f(x) \geq 0$ for $x \in [0, 1]$. In particular $f(\frac{c}{a}) \geq 0$, which implies $a - c \geq (a + c)^\beta (a^{1-\beta} - c^{1-\beta}) \geq$

$b^\beta(a^{1-\beta} - c^{1-\beta})$. Thus,

$$\frac{ca^\beta + \left(b^\beta - a^\beta\right) a}{a^\beta b^\beta} \leq c^{1-\beta}\,.$$

<u>Case II, $c \geq a$:</u> As $1-\beta \leq 1$ and $c - a \geq 0$, we have $(c-a)^{1-\beta} + a^{1-\beta} \leq 2^\beta c^{1-\beta}$. Multiplying by $(c-a)^\beta$ and using $c - a \leq b$, we get $c - a \leq b^\beta \left(2^\beta c^{1-\beta} - a^{1-\beta}\right)$. Thus,

$$\frac{ca^\beta + \left(b^\beta - a^\beta\right) a}{a^\beta b^\beta} \leq 2^\beta c^{1-\beta}\,.$$

□

Proof of Lemma 1. Applying Lemma 8 to the left hand side of equation (6), yields

$$\frac{{}^{\mathfrak{c}}\overline{yq} - {}^{\mathfrak{c}}\overline{ym} - {}^{\mathfrak{c}}\overline{zq} + {}^{\mathfrak{c}}\overline{zm}}{\mathfrak{b}(q,m)^\xi} - \frac{{}^{\mathfrak{c}}\overline{yp} - {}^{\mathfrak{c}}\overline{ym} - {}^{\mathfrak{c}}\overline{zp} + {}^{\mathfrak{c}}\overline{zm}}{\mathfrak{b}(p,m)^\xi} \leq \mathfrak{a}(y,z)\,\tilde{\mathfrak{b}}_{m,\xi}(q,p)$$

where

$$\tilde{\mathfrak{b}}_{m,\xi}(q,p) = \frac{\min({}^{\mathfrak{b}}\overline{qm}, {}^{\mathfrak{b}}\overline{pm})^\xi\, {}^{\mathfrak{b}}\overline{qp} + \left|{}^{\mathfrak{b}}\overline{qm}^\xi - {}^{\mathfrak{b}}\overline{pm}^\xi\right| \min({}^{\mathfrak{b}}\overline{qm}, {}^{\mathfrak{b}}\overline{pm})}{{}^{\mathfrak{b}}\overline{qm}^\xi\, {}^{\mathfrak{b}}\overline{pm}^\xi}$$

with the short notation ${}^{\mathfrak{b}}\overline{qp} := \mathfrak{b}(q,p)$, for all $y, z, q, p, m \in \mathcal{Q}$. Applying Lemma 9 yields $\tilde{\mathfrak{b}}_{m,\xi}(q,p) \leq 2^\xi \mathfrak{b}(q,p)^{1-\xi}$. □

Appendix D Projection Metric Counter Example

We take a tripod $(\mathcal{Q}, d)$ as a simple example of a non-Euclidean Hadamard space, see [Stu03, Example 3.2], and show that it does not fulfill the strong quadruple inequality with the projection metric

$$d_m^{\mathsf{proj}}(q,p) := \sqrt{\frac{\overline{qp}^2 - (\overline{qm} - \overline{pm})^2}{\overline{qm}\,\overline{pm}}}\,.$$

Let $r > \varepsilon > 0$ and define y, z, q, p, o on a tripod as in Figure 3. We take $\mathfrak{c} = d^2$, $\xi = 1$, $\mathfrak{l} = d$, $\mathfrak{a} = Kd$, and $\mathfrak{b}_m = d_m^{\mathsf{proj}}$. Then

$$\frac{{}^{\mathfrak{c}}\overline{yq} - {}^{\mathfrak{c}}\overline{ym} - {}^{\mathfrak{c}}\overline{zq} + {}^{\mathfrak{c}}\overline{zm}}{\mathfrak{l}(q,m)} - \frac{{}^{\mathfrak{c}}\overline{yp} - {}^{\mathfrak{c}}\overline{ym} - {}^{\mathfrak{c}}\overline{zp} + {}^{\mathfrak{c}}\overline{zm}}{\mathfrak{l}(p,m)} = 2\varepsilon\,,$$

$$\mathfrak{a}(y,z)\mathfrak{b}_m(q,p) = K\varepsilon\sqrt{2\frac{\varepsilon}{r+\varepsilon}}\,.$$

If the strong quadruple inequality holds, then

$$K \geq \sqrt{2\frac{r+\varepsilon}{\varepsilon}} \xrightarrow{\varepsilon \searrow 0} \infty\,.$$

Thus, d_m^{proj} is not a suitable candidate for the strong quadruple distance in general Hadamard spaces.

	y	z	q	p	m
pod number	1	1	2	1	3
distance from center	ε	0	r	ε	r

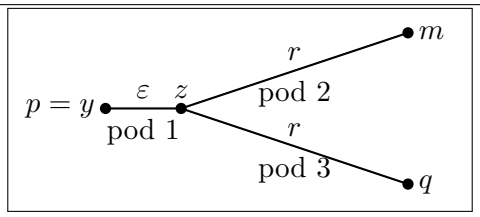

Figure 3: Tripod counter-example for the strong quadruple inequality

Case	$\overline{yq}$	$\overline{yp}$	$\overline{zq}$	$\overline{zp}$	$\overline{yz}$	$\overline{qp}$
(a)	$1-\epsilon$	$1-3\epsilon$	$1-2\epsilon$	1	$2-3\epsilon$	2ϵ
(b)	1	ϵ	$1-\epsilon$	2ϵ	2ϵ	1
(c)	2ϵ	ϵ	1	1	1	ϵ

Table 1: Distances of four points $y, z, q, p \in \mathcal{Q}$ for showing lower bounds of the constant in Theorem 3.

Appendix E Optimality of Power Inequality

We show that $8\alpha 2^{-2\alpha}$ is the optimal constant, and that we cannot extend Theorem 3 to $\alpha > 1$ or $\alpha < \frac{1}{2}$. Let $\epsilon \in (0,1)$ and $(\mathcal{Q}, d)$ be a metric space with $q, p, y, z \in \mathcal{Q}$ such that for each case below the distances have the values written down in Table 1. One can easily show that in all three cases the necessary triangle inequalities and the nice quadruple inequality hold.

(a) For $\alpha \in [\frac{1}{2}, 1]$, we have, using l'Hôpital's rule,

$$\begin{aligned}
&\lim_{\epsilon \searrow 0} \frac{\overline{yq}^{2\alpha} - \overline{yp}^{2\alpha} - \overline{zq}^{2\alpha} + \overline{zp}^{2\alpha}}{\overline{yz}^{2\alpha-1}\,\overline{qp}} \\
&= \lim_{\epsilon \searrow 0} \frac{(1-\epsilon)^{2\alpha} - (1-3\epsilon)^{2\alpha} - (1-2\epsilon)^{2\alpha} + 1}{(2\epsilon)(2-3\epsilon)^{2\alpha-1}} \\
&= \lim_{\epsilon \searrow 0} 2\alpha \frac{-(1-\epsilon)^{2\alpha-1} + 3(1-3\epsilon)^{2\alpha-1} + 2(1-2\epsilon)^{2\alpha-1}}{2(2-3\epsilon)^{2\alpha-1} - 6\epsilon(2\alpha-1)(2-3\epsilon)^{2\alpha-2}} \\
&= 8\alpha 2^{-2\alpha}\,.
\end{aligned}$$

Thus, the constant $8\alpha 2^{-2\alpha}$ in Theorem 3 is optimal.

(b) For $\alpha > 1$, we have, using l'Hôpital's rule,

$$\begin{aligned}
&\lim_{\epsilon \searrow 0} \frac{\overline{yq}^{2\alpha} - \overline{yp}^{2\alpha} - \overline{zq}^{2\alpha} + \overline{zp}^{2\alpha}}{\overline{yz}^{2\alpha-1}\,\overline{qp}} \\
&= \lim_{\epsilon \searrow 0} \frac{1 - \epsilon^{2\alpha} - (1-\epsilon)^{2\alpha} + (2\epsilon)^{2\alpha}}{(2\epsilon)^{2\alpha-1}} \\
&= \frac{2\alpha}{2\alpha - 1} \lim_{\epsilon \searrow 0} \frac{-\epsilon^{2\alpha-1} + (1-\epsilon)^{2\alpha-1} + 2(2\epsilon)^{2\alpha-1}}{2^{2\alpha-1}\epsilon^{2\alpha-2}} \\
&= \infty\,.
\end{aligned}$$

Thus, there is no power inequality in the form of Theorem 3 for $\alpha > 1$.

(c) For $\alpha \in (0, \frac{1}{2})$, we have

$$\begin{aligned}
&\lim_{\epsilon \searrow 0} \frac{\overline{yq}^{2\alpha} - \overline{yp}^{2\alpha} - \overline{zq}^{2\alpha} + \overline{zp}^{2\alpha}}{\overline{yz}^{2\alpha-1}\,\overline{qp}} \\
&= \lim_{\epsilon \searrow 0} \frac{(2\epsilon)^{2\alpha} - \epsilon^{2\alpha} - 1 + 1}{\epsilon} \\
&= \lim_{\epsilon \searrow 0} (2^{2\alpha} - 1)\epsilon^{2\alpha-1} \\
&= \infty\,.
\end{aligned}$$

Thus, there is no power inequality in the form of Theorem 3 for $\alpha < \frac{1}{2}$.

Appendix F Chaining

Recall the measures of entropy γ_2 and entr_n defined in Definition 1. We add another useful entry to this list.

Definition 2 (Bernoulli Bound)**.** For $T \subseteq \mathbb{R}^n$ define

$$b(T) := \inf\left\{ \sup_{t \in T_1} \|t\|_1 + \gamma_2(T_2) : \; T_1, T_2 \subseteq \mathbb{R}^n, T \subseteq T_1 + T_2 \right\},$$

where $\gamma_2(T_2) = \gamma_2(T_2, d_2)$ for the Euclidean metric d_2 on $\mathbb{R}^n$, $\|t\|_1 = \sum_{i=1}^n |t_i|$, and $T_1 + T_2 = \{t_1 + t_2 \colon t_1 \in T_1, t_2 \in T_2\}$.

We write down the Bernoulli bound for powers of the Bernoulli process. [BL14] show that the bound can be reversed (up to an universal constant). Thus, this step can be regarded as optimal.

Theorem 5 (Bernoulli bound)**.** Let $\sigma_1, \dots, \sigma_n$ be independent random signs, i.e., $\mathbb{P}(\sigma_i = \pm 1) = \frac{1}{2}$. For $t \in R^n$ set $\tilde{X}_t := \sum_{i=1}^n \sigma_i t_i$. Let $T \subseteq \mathbb{R}^n$. Let $\kappa \geq 1$. Then

$$\mathbb{E}\left[\sup_{t \in T} \left|\tilde{X}_t\right|^\kappa\right] \leq c_\kappa b(T)^\kappa \,,$$

where c_κ depends only on κ.

Proof. Let $T_1, T_2 \subseteq \mathbb{R}^n$ such that $T \subseteq T_1 + T_2$. As $(a+b)^\kappa \leq 2^{\kappa-1}(a^\kappa + b^\kappa)$ for all $a, b \geq 0$, we can split the supremum into two parts,

$$\mathbb{E}\left[\sup_{t \in T} \left|\tilde{X}_t\right|^\kappa\right] \leq 2^{\kappa-1}\left(\mathbb{E}\left[\sup_{t \in T_1} \left|\tilde{X}_t\right|^\kappa\right] + \mathbb{E}\left[\sup_{t \in T_2} \left|\tilde{X}_t\right|^\kappa\right]\right) .$$

The first term is bounded using the 1-norm, $\mathbb{E}\left[\sup_{t \in T_1} \left|\tilde{X}_t\right|^\kappa\right] \leq \sup_{t \in T_1} \|t\|_1^\kappa$. For the second we use Talagrand's generic chaining bound for the supremum of the subgaussian process $\mathbb{E}\left[\sup_{t \in T_2} \left|\tilde{X}_t\right|^\kappa\right] \leq c'_\kappa \gamma_2(T_2)^\kappa$, see [Tal14]. We obtain

$$\begin{aligned}\mathbb{E}\left[\sup_{t \in T} \left|\tilde{X}_t\right|^\kappa\right] &\leq c_\kappa \left(\sup_{t \in T_1} \|t\|_1^\kappa + \gamma_2(T_2)^\kappa\right) \\ &\leq c_\kappa \left(\sup_{t \in T_1} \|t\|_1 + \gamma_2(T_2)\right)^\kappa .\end{aligned}$$

□

Lemma 10 (Lipschitz connection)**.** Let $(\mathcal{Q}, b)$ be a pseudo-metric space. Assume there are function $f_i \colon \mathcal{Q} \to \mathbb{R}$ such that $|f_i(q) - f_i(p)| \leq a_i b(q,p)$. Let $T := \{(f_i(q))_{i=1,\dots,n} \colon q \in \mathcal{Q}\}$. Set $a = (a_1, \dots, a_n)$. Then

$$b(T) \leq C\|a\|_2 \min(\mathsf{entr}_n(\mathcal{Q}, b), \gamma_2(\mathcal{Q}, b)) \,,$$

where $C > 0$ is an universal constant.

Proof. For $\epsilon > 0$, choose $\mathcal{Q}_2$ to be an ϵ-covering of $\mathcal{Q}$ with respect to b, i.e., for all $q \in \mathcal{Q}$ there is a $p_q \in \mathcal{Q}_2$ such that $b(q, p_q) \leq \epsilon$. For $q \in \mathcal{Q}$ denote $t_q := (f_i(q))_{i=1,\dots,n} \in \mathbb{R}^n$. Define $T_2 := \{t_p \colon p \in \mathcal{Q}_2\}$ and $T_1 := \{t_q - t_{p_q} \colon q \in Q\}$. Then $T \subseteq T_1 + T_2$. The Lipschitz-condition implies $\|t_q - t_p\|_2 \leq \|a\|_2 b(q,p)$ for all $q, p \in \mathcal{Q}$. Thus,

$$\sup_{t \in T_1} \|t\|_1 \leq \sup_{q \in \mathcal{Q}} \sqrt{n}\|t_q - t_{p_q}\|_2 \leq \epsilon\sqrt{n}\|a\|_2 \,.$$

By the properties of γ_2, see [Tal14], we obtain

$$\gamma_2(T_2) \leq c\|a\|_2\gamma_2(\mathcal{Q}_2, b) \leq c'\|a\|_2 \int_\epsilon^\infty \sqrt{\log N(Q, b, r)}\mathrm{d}r$$

for universal constants $c, c' > 0$. Applying the two inequalities to the definition of $b(T)$ concludes the proof. □

Lemma 11 (Symmetrization)**.** Let $\mathcal{Q}$ be set. Let $Z_1, \ldots, Z_n$ be centered, independent, and integrable stochastic processes indexed by $\mathcal{Q}$. Let $\Phi\colon \mathbb{R} \to \mathbb{R}$ be a convex, non-decreasing function. Let $(Z_1', \ldots, Z_n')$ be an independent copy of $(Z_1, \ldots, Z_n)$. Let $\varepsilon_1, \ldots, \varepsilon_n$ be iid with $\mathbb{P}(\varepsilon_1 = \pm 1) = \frac{1}{2}$. Then

$$\mathbb{E}\left[\sup_{q\in\mathcal{Q}} \Phi\left(\sum_{i=1}^n Z_i(q)\right)\right] \leq \mathbb{E}\left[\sup_{q\in\mathcal{Q}} \Phi\left(\sum_{i=1}^n \varepsilon_i \left(Z_i(q) - Z_i'(q)\right)\right)\right],$$

assuming measurability of the involved terms.

The symmetrization lemma is well-known. The statement here is an intermediate step of from the proof of [VW96, 2.3.6 Lemma].

Theorem 6 (Empirical process bound)**.** Let $(\mathcal{Q}, b)$ be a separable pseudo-metric space. Let $Z_1, \ldots, Z_n$ be centered, independent, and integrable stochastic processes indexed by $\mathcal{Q}$ with a $q_0 \in \mathcal{Q}$ such that $Z_i(q_0) = 0$ for $i = 1, \ldots, n$. Let $(Z_1', \ldots, Z_n')$ be an independent copy of $(Z_1, \ldots, Z_n)$. Assume the following Lipschitz-property: There is a random vector A with values in $\mathbb{R}^n$ such that

$$\left|Z_i(q) - Z_i(p) - Z_i'(q) + Z_i'(p)\right| \leq A_i b(q, p)$$

for $i = 1, \ldots, n$ and all $q, p \in \mathcal{Q}$. Let $\kappa \geq 1$. Then

$$\mathbb{E}\left[\sup_{q\in\mathcal{Q}} \left|\sum_{i=1}^n Z_i(q)\right|^\kappa\right] \leq C\,\mathbb{E}[\|A\|_2^\kappa]\,\min(\mathsf{entr}_n(\mathcal{Q}, b), \gamma_2(\mathcal{Q}, b))^\kappa\,,$$

where $C > 0$ is an universal constant.

Proof. Use Lemma 11. Then apply Theorem 5 and Lemma 10 conditionally on $Z_1, \ldots, Z_n$. □

Lemma 12. Let $(\mathcal{Q}, b)$ be a pseudo-metric space. Let $D > 0$ such that $\mathsf{diam}(\mathcal{Q}, b) \leq$

$D < \infty$. Let $\beta > 0$. Assume

$$\sqrt{\log(N(\mathcal{Q}, b, r))} \leq c_{\mathsf{e}} \left(\frac{D}{r}\right)^{\beta}$$

for all $0 < r < D$.

(i) If $\beta < 1$ then $\mathsf{entr}_n(\mathcal{Q}, b) \leq \frac{c_{\mathsf{e}} D}{1-\beta}$.

(ii) If $\beta = 1$ then $\mathsf{entr}_n(\mathcal{Q}, b) \leq c'_{\mathsf{e}} D \log(n+1)$, where $c' > 0$ depends only on c_{e}.

(iii) If $\beta > 1$ then $\mathsf{entr}_n(\mathcal{Q}, b) \leq c_{\mathsf{e}}^{\frac{1}{\beta}} \frac{\beta}{\beta-1} n^{-\frac{1}{2\beta}+\frac{1}{2}}$.

In particular

$$n^{-\frac{1}{2}}\ \mathsf{entr}_n(\mathcal{Q}, b) \ \leq\ c\, D\, \eta_{\beta,n}\,,$$

where c depends only on c_{e} and β and

$$\eta_{\beta,n} := \begin{cases} n^{-\frac{1}{2}} & \text{for } \beta < 1\,, \\ n^{-\frac{1}{2}} \log(n+1) & \text{for } \beta = 1\,, \\ n^{-\frac{1}{2\beta}} & \text{for } \beta > 1\,. \end{cases}$$

The proof consists of calculating the entropy integral with the given bound on the covering numbers and, for $\beta \geq 1$, choosing the minimizing starting point of the integral $\epsilon > 0$.

Appendix G Proof of the Power Inequality, Theorem 3

Let $(\mathcal{Q}, d)$ be a metric space. Use the short notation $\overline{qp} := d(q, p)$. Let $q, p, y, z \in \mathcal{Q}$, $\alpha \in [\frac{1}{2}, 1]$. Assume

$$\overline{yq}^2 - \overline{yp}^2 - \overline{zq}^2 + \overline{zp}^2 \leq 2\, \overline{yz}\, \overline{qp}\,.$$

The goal of this section is to prove

$$\overline{yq}^{2\alpha} - \overline{yp}^{2\alpha} - \overline{zq}^{2\alpha} + \overline{zp}^{2\alpha} \leq 8\alpha 2^{-2\alpha}\, \overline{yz}^{2\alpha-1}\, \overline{qp}\,.$$

G.1 Arithmetic Form

Theorem 3 will be proven in the form of Lemma 13.

Lemma 13. Let $a, b, c \geq 0$, $r, s \in [-1, 1]$, and $\alpha \in [\frac{1}{2}, 1]$. Then

$$\begin{aligned} & a^{2\alpha} - c^{2\alpha} - \left(a^2 - 2rab + b^2\right)^{\alpha} + \left(c^2 - 2scb + b^2\right)^{\alpha} \\ & \leq 8\alpha 2^{-2\alpha} b \max(ra - sc, |a - c|)^{2\alpha-1}\,. \end{aligned}$$

The advantage of using Lemma 13 to prove Theorem 3 is, that we do not need to consider a system of additional conditions for describing that the real values in the inequality are distances, which have to fulfill the triangle inequality. The disadvantage is, that we loose the possibility for a geometric interpretation of the proof.

Lemma 14. Lemma 13 implies Theorem 3.

Proof. Three points from an arbitrary metric space can be embedded in the Euclidean plane so that the distances are preserved. Thus, the cosine formula of Euclidean geometry can be applied to the three points $y, p, q \in \mathcal{Q}$: It holds

$$\overline{yq}^2 = \overline{yp}^2 + \overline{qp}^2 - 2s\,\overline{yp}\,\overline{qp}\,,$$

where $s := \cos(\measuredangle ypq)$ with the angle $\measuredangle ypq$ in the Euclidean plane. Similarly

$$\overline{zq}^2 = \overline{zp}^2 + \overline{qp}^2 - 2r\,\overline{zp}\,\overline{qp}\,,$$

where $r := \cos(\measuredangle zpq)$. Thus,

$$\begin{aligned}
&\overline{yq}^{2\alpha} - \overline{yp}^{2\alpha} - \overline{zq}^{2\alpha} + \overline{zp}^{2\alpha} \\
&= \left(\overline{yp}^2 + \overline{qp}^2 - 2s\,\overline{yp}\,\overline{qp}\right)^{\alpha} - \left(\overline{zp}^2 + \overline{qp}^2 - 2r\,\overline{zp}\,\overline{qp}\right)^{\alpha} - \overline{yp}^{2\alpha} + \overline{zp}^{2\alpha} \\
&= \left(c^2 + b^2 - 2scb\right)^{\alpha} - \left(a^2 + b^2 - 2rab\right)^{\alpha} - c^{2\alpha} + a^{2\alpha}\,,
\end{aligned}$$

where $a := \overline{zp}$, $c := \overline{yp}$, $b := \overline{qp}$. Hence, Lemma 13 yields

$$\overline{yq}^{2\alpha} - \overline{yp}^{2\alpha} - \overline{zq}^{2\alpha} + \overline{zp}^{2\alpha} \leq 8\alpha 2^{-2\alpha} b \max(ra - sc, |a - c|)^{2\alpha - 1}\,. \tag{8}$$

The assumption of Theorem 3 states $\overline{yq}^2 - \overline{yp}^2 - \overline{zq}^2 + \overline{zp}^2 \leq 2\,\overline{yz}\,\overline{qp}$. This implies

$$2b\,(ra - sc) = \left(c^2 + b^2 - 2scb\right) - \left(a^2 + b^2 - 2rab\right) - c^2 + a^2 \leq 2b\,\overline{yz}\,.$$

Therefore, $ra - sc \leq \overline{yz}$ (or $b = 0$, but then $q = p$ and Theorem 3 becomes trivial). Furthermore, the triangle inequality implies $|a - c| = |\overline{zp} - \overline{yp}| \leq \overline{yz}$. Thus, we obtain

$$\max(ra - sc, |a - c|) \leq \overline{yz}\,. \tag{9}$$

Finally, (8) and (9) together yield

$$\overline{yq}^{2\alpha} - \overline{yp}^{2\alpha} - \overline{zq}^{2\alpha} + \overline{zp}^{2\alpha} \leq 8\alpha 2^{-2\alpha}\,\overline{qp}\,\overline{yz}^{2\alpha - 1}\,.$$

□

The remaining part of this section is dedicated to proving Lemma 13.

The proof of Lemma 13 can be described as *brute force*. We will distinguish many different cases, i.e., certain bounds on a, b, c, r, s, e.g., $a \leq c$ and $a > c$. In each case,

we try to simplify the inequality step by step until we can solve it easily. Mostly, the simplification consists of taking some derivative and showing that the derivative is always negative (or always positive). Then we only need to show the inequality at one extremal point. This process may have to be iterated. It is often not clear immediately which derivative to take in order to simplify the inequality. Even after finishing the proof there seems to be no deeper reason for distinguishing the cases that are considered. Thus, unfortunately, the proof does not create a deeper understanding of the result.

G.2 First Proof Steps and Outline of the Remaining Proof

We want to show Lemma 13 to prove Theorem 3. We refer to the left hand side of the inequality, $a^{2\alpha} - c^{2\alpha} - \left(a^2 - 2rab + b^2\right)^\alpha + \left(c^2 - 2scb + b^2\right)^\alpha$, as LHS. By RHS we, of course, mean the right hand side, $8\alpha 2^{-2\alpha} b \max(ra - sc, |a - c|)^{2\alpha - 1}$.

For $\max(ra - sc, |a - c|) = 0$ we have $a = c$ and $r \leq s$. Thus, LHS ≤ 0. If $\max(ra - sc, |a - c|) > 0$, LHS and RHS are continuous in all parameters. Thus, it is enough to show the inequality on a dense set. In particular, we can and will ignore certain special cases in the following which might introduce technical problems, e.g., "0^0".

We have to distinguish the cases $|a - c| = \max(ra - sc, |a - c|)$ and $ra - sc = \max(ra - sc, |a - c|)$. We further distinguish $a \geq c$ and $c \geq a$.

Lemma 15 ($ra - sc$ vs $|a - c|$)**.** Let $a, b, c \geq 0$, $r, s \in [-1, 1]$, and $\alpha \in [\frac{1}{2}, 1]$. Then

$$\begin{aligned} ra - sc \geq a - c \quad &\Leftrightarrow \quad s \leq (r-1)\frac{a}{c} + 1 \quad \Leftrightarrow \quad r \geq (s-1)\frac{c}{a} + 1\,, \\ ra - sc \geq c - a \quad &\Leftrightarrow \quad s \leq (r+1)\frac{a}{c} - 1 \quad \Leftrightarrow \quad r \geq (s+1)\frac{c}{a} - 1\,. \end{aligned}$$

Proof. It holds

$$\begin{aligned} ra - sc \geq a - c \quad &\Leftrightarrow \quad ra - a + c \geq sc \quad \Leftrightarrow \quad \frac{a(r-1) + c}{c} \geq s\,, \\ ra - sc \geq c - a \quad &\Leftrightarrow \quad ra - c + a \geq sc \quad \Leftrightarrow \quad \frac{a(r+1) - c}{c} \geq s\,, \\ ra - sc \geq a - c \quad &\Leftrightarrow \quad ra \geq a - c + sc \quad \Leftrightarrow \quad r \geq (s-1)\frac{c}{a} + 1\,, \\ ra - sc \geq c - a \quad &\Leftrightarrow \quad ra \geq c - a + sc \quad \Leftrightarrow \quad r \geq (s+1)\frac{c}{a} - 1\,. \end{aligned}$$

□

G.2.1 The Case $|a - c| \leq ra - sc$

Consider the case $ra - sc \geq |a - c|$. The next lemma shows convexity in r of the function "LHS minus RHS". This means, we only have to check values of r on the border of its domain.

Lemma 16 (Convexity in r)**.** Let $a, b, c \geq 0$, $s, r \in [-1, 1]$, $\alpha \in [\frac{1}{2}, 1]$. Assume $ra - sc \geq 0$. Define

$$F(r,s) := a^{2\alpha} - c^{2\alpha} - \left(a^2 - 2rab + b^2\right)^{\alpha} + \left(c^2 - 2scb + b^2\right)^{\alpha} - 8\alpha 2^{-2\alpha} b (ra - sc)^{2\alpha - 1} .$$

Then

$$\partial_r^2 F(r,s) \geq 0 .$$

Note, neither $\partial_s^2 F(r,s) \geq 0$ nor $\partial_s^2 F(r,s) \leq 0$ for all a, b, c, s, r.

Proof. Define

$$\ell(r,s) := a^{2\alpha} - c^{2\alpha} - \left(a^2 - 2rab + b^2\right)^{\alpha} + \left(c^2 - 2scb + b^2\right)^{\alpha} .$$

It holds

$$\partial_r \ell(r,s) = 2ab\alpha \left(a^2 - 2rab + b^2\right)^{\alpha - 1} .$$

Define $h(r,s) := 8\alpha 2^{-2\alpha} b \left(ra - sc\right)^{2\alpha - 1}$. It holds

$$\partial_r h(r,s) = 8\alpha(2\alpha - 1) 2^{-2\alpha} ba \left(ra - sc\right)^{2\alpha - 2} .$$

It holds $F(r,s) = \ell(r,s) - h(r,s)$. It holds

$$f(r) := \frac{\partial_r \ell(r,s) - \partial_r h(r,s)}{2ab\alpha} = \left(a^2 - 2rab + b^2\right)^{\alpha - 1} - (2\alpha - 1)\left(\frac{ra - sc}{2}\right)^{2\alpha - 2}$$

and

$$\partial_r f(r) = -2ab(\alpha - 1)\left(a^2 - 2rab + b^2\right)^{\alpha - 2} - \frac{1}{2} a(2\alpha - 1)(2\alpha - 2)\left(\frac{ra - sc}{2}\right)^{2\alpha - 3} .$$

It holds $2ab\left(a^2 - 2rab + b^2\right)^{\alpha - 2} \geq 0$ and $(\alpha - 1) \leq 0$. Thus,

$$-2ab(\alpha - 1)\left(a^2 - 2rab + b^2\right)^{\alpha - 2} \geq 0 .$$

It holds $\frac{1}{2} a(2\alpha - 1)\left(\frac{ra - sc}{2}\right)^{2\alpha - 3} \geq 0$ and $(2\alpha - 2) \leq 0$. Thus, $-\frac{1}{2} a(2\alpha - 1)(2\alpha - 2)\left(\frac{ra - sc}{2}\right)^{2\alpha - 3} \geq 0$. Hence, $\partial_r f(r) \geq 0$. Hence, $\partial_r^2 F(r,s) \geq 0$. □

G.2.2 The Case $|a - c| \geq ra - sc$

In the case $|a - c| \geq ra - sc$, the RHS does not depend on s or r. Thus, we maximize the LHS with respect to r and s and only need to show the inequality for this maximized term.

Define

$$\ell(r,s) := a^{2\alpha} - c^{2\alpha} - \left(a^2 - 2rab + b^2\right)^\alpha + \left(c^2 - 2scb + b^2\right)^\alpha .$$

It holds

$$\max_{s \geq s_0, r \leq r_0} \ell(r,s) = \ell(r_0, s_0) .$$

Distinguish the two cases $a \geq c$ and $a \leq c$.
<u>Case 1: $a \geq c$.</u> For fixed $r \in [-1,1]$, set $s = s_{\mathsf{min}}(r) = (r-1)\frac{a}{c} + 1$, cf Lemma 15. Define

$$\begin{aligned} f(r) &:= \ell(r, s_{\mathsf{min}}(r)) \\ &= a^{2\alpha} - c^{2\alpha} - \left(a^2 - 2rab + b^2\right)^\alpha + \left(c^2 - 2rab + 2ab - 2cb + b^2\right)^\alpha . \end{aligned}$$

Then

$$\frac{f'(r)}{2ab\alpha} = \left(a^2 - 2rab + b^2\right)^{\alpha-1} - \left(c^2 - 2rab + 2ab - 2cb + b^2\right)^{\alpha-1} .$$

<u>Case 1.1: $a^2 \leq c^2 + 2ab - 2cb$.</u> Then

$$\begin{aligned} a^2 - 2rab + b^2 &\leq c^2 - 2rab + 2ab - 2cb + b^2 , \\ \left(a^2 - 2rab + b^2\right)^{\alpha-1} &\geq \left(c^2 - 2rab + 2ab - 2cb + b^2\right)^{\alpha-1} . \end{aligned}$$

Thus, $f'(r) \geq 0$. In this case, we need to show

$$a^{2\alpha} - c^{2\alpha} - |a-b|^{2\alpha} + |c-b|^{2\alpha} = f(1) \leq 8\alpha 2^{-2\alpha} b(a-c)^{2\alpha-1} .$$

<u>Case 1.2: $a^2 \geq c^2 + 2ab - 2cb$.</u> Then

$$\begin{aligned} a^2 - 2rab + b^2 &\geq c^2 - 2rab + 2ab - 2cb + b^2 , \\ \left(a^2 - 2rab + b^2\right)^{\alpha-1} &\leq \left(c^2 - 2rab + 2ab - 2cb + b^2\right)^{\alpha-1} . \end{aligned}$$

Thus, $f'(r) \leq 0$. The relevant values are $r = r_{\mathsf{min}} = 1 - 2\frac{c}{a}$, with $s = s_{\mathsf{min}}(r) = -1$. In this case, we need to show

$$a^{2\alpha} - c^{2\alpha} - \left((a-b)^2 + 4cb\right)^\alpha + (c+b)^{2\alpha} = f(r_{\mathsf{min}}) \leq 8\alpha 2^{-2\alpha} b(a-c)^{2\alpha-1} .$$

<u>Case 2: $a \leq c$.</u> For fixed $r \in [-1,1]$, set $s = s_{\mathsf{min}}(r) = (r+1)\frac{a}{c} - 1$. Define

$$\begin{aligned} f(r) &:= \ell(r, s_{\mathsf{min}}(r)) \\ &= a^{2\alpha} - c^{2\alpha} - \left(a^2 - 2rab + b^2\right)^\alpha + \left(c^2 - 2rab - 2ab + 2cb + b^2\right)^\alpha . \end{aligned}$$

Then

$$\frac{f'(r)}{2ab\alpha} = \left(a^2 - 2rab + b^2\right)^{\alpha-1} - \left(c^2 - 2rab - 2ab + 2cb + b^2\right)^{\alpha-1} .$$

Case 2.1: $a^2 \leq c^2 - 2ab + 2cb$. Then

$$a^2 - 2rab + b^2 \leq c^2 - 2rab - 2ab + 2cb + b^2\,,$$
$$\left(a^2 - 2rab + b^2\right)^{\alpha-1} \geq \left(c^2 - 2rab - 2ab + 2cb + b^2\right)^{\alpha-1}\,.$$

Thus, $f'(r) \geq 0$. The critical value is $r = r_{\mathsf{max}} = 1$, with $s = s_{\mathsf{min}}(r) = 2\frac{a}{c} - 1$. In this case, we need to show

$$a^{2\alpha} - c^{2\alpha} - |a-b|^{2\alpha} + \left((c+b)^2 - 4ab\right)^{\alpha} = f(1) \leq 8\alpha 2^{-2\alpha} b(c-a)^{2\alpha-1}\,.$$

Case 2.2: $a^2 \geq c^2 - 2ab + 2cb$. This cannot happen for $a \leq c$.

Remark 1. Assume $a \geq c$. Then

$$a^2 \geq c^2 + 2ab - 2cb \quad \Leftrightarrow \quad a^2 - c^2 \geq 2b(a-c) \quad \Leftrightarrow \quad a + c \geq 2b\,.$$

G.2.3 Outline

Remark 2 (What we need to show)**.** Define

$$\begin{aligned} F(r,s) :=& a^{2\alpha} - c^{2\alpha} - \left(a^2 - 2rab + b^2\right)^{\alpha} + \left(c^2 - 2scb + b^2\right)^{\alpha} \\ &- 8\alpha 2^{-2\alpha} b(ra - sc)^{2\alpha-1}\,. \end{aligned}$$

(i) For the case $a \geq c$, we need $F(1,s) \leq 0$ for all $s \in [-1,1]$ (cf Lemma 16, includes case 1.1 of section G.2.2) and $F(1 - 2\frac{c}{a}, -1) \leq 0$ (case 1.2 of section G.2.2). Note $r = 1 - 2\frac{c}{a}, s = -1$ means $ra - sc = a - c$.

(ii) For the case $a \leq c$, we need $F(1,s) \leq 0$ for all $s \in [-1, 2\frac{a}{c} - 1]$ (cf Lemma 16, includes case 2.1 of section G.2.2). Note $r = 1, s = 2\frac{a}{c} - 1$ means $ra - sc = c - a$.

The part $F(1 - 2\frac{c}{a}, -1) \leq 0$ for $a \geq c$ is discussed in section G.7. The different case for $F(1,s) \leq 0$ ($a \leq c$ and $a \geq c$) are covered in the following way:

(a) $b \geq 2sc$: Lemma 20 (Merging Lemma) + Lemma 17 (Tight Power Bound)

(b) $b \leq 2sc$ and $sc \leq a - b$: Lemma 21 (Merging Lemma) + Lemma 17 (Tight Power Bound)

(c) $b \leq 2sc$ and $sc \geq a - b$ and $a \geq c$: Lemma 25

(d) $b \leq 2sc$ and $sc \geq a - b$ and $a \leq c$, $sc \leq 2a - c$ and $b \leq 2a - 2sc$: Lemma 26

(e) $b \leq 2sc$ and $sc \geq a - b$ and $a \leq c$, $sc \leq 2a - c$ and $b \geq 2a - 2sc$ and $a \leq b$: Lemma 29

(f) $b \leq 2sc$ and $sc \geq a - b$ and $a \leq c$, $sc \leq 2a - c$ and $b \geq 2a - 2sc$ and $a \geq b$: Lemma 31

The proofs consist of distinguishing many different cases and applying simple analysis methods in each case. Nonetheless, finding the poofs is often quite hard, as the inequalities are usually very tight and the right steps necessary for the proof are hard to guess.

As intermediate steps we can, in some cases, use two lemmas: the Tight Power Bound, see section G.3, and the Merging Lemma, see G.4. The remaining cases that cannot be solved via Tight Power Bound and Merging Lemma will be discussed in sections G.6 and G.7.

G.3 Tight Power Bound

Following lemma gives one very useful inequality in three different forms. It gives a hint to why the power $\ldots^{2\alpha-1}$ comes up in the RHS of Lemma 13.

Lemma 17 (Tight Power Bound)**.** Let $x, y \geq 0$.

(i) If $a \in [1, 2]$, $x \geq y$, then

$$2^a x^{a-1} y \quad \leq \quad (x+y)^a - (x-y)^a \quad \leq \quad 2a x^{a-1} y \,.$$

(ii) If $a \in [1, 2]$, then

$$(x+y)^a - |x-y|^a \quad \leq \quad 2a \min(xy^{a-1}, x^{a-1}y) \,.$$

(iii) If $a \in [1, 2]$, $x \geq y$, then

$$(x+y)^{a-1}(x-y) \quad \leq \quad x^a - y^a \quad \leq \quad a(x-y)\left(\frac{x+y}{2}\right)^{a-1} \,.$$

Note that this result is slightly stronger than the application of the mean value theorem to the function $x \mapsto x^a$, which yields $x^a - y^a \leq a(x-y)z^{a-1}$ for all $x \geq y \geq 0$ and $a > 0$, where $z \in [y, x]$.

Proof. Assume $x \geq y$. Set $z = \frac{y}{x} \in [0, 1]$. Define

$$f(z) = \frac{(1+z)^a - (1-z)^a}{z} \,.$$

If we can show $f(z) \leq 2a$, then

$$\begin{aligned} & (1+z)^a - (1-z)^a \leq 2az \\ \Rightarrow \quad & (x+zx)^a - (x-zx)^a \leq 2ax^a z \\ \Rightarrow \quad & (x+y)^a - (x-y)^a \leq 2ax^{a-1}y\,. \end{aligned}$$

It holds

$$f'(z) = \frac{g(z)}{z^2}\,,$$

where

$$g(z) = az\left((1+z)^{a-1} + (1-z)^{a-1}\right) - \left((1+z)^a - (1-z)^a\right)\,.$$

It holds

$$g'(z) = az(a-1)\left((1+z)^{a-2} - (1-z)^{a-2}\right) \leq 0\,.$$

Thus, $g(z) \leq g(0) = 0$. Thus, $f'(z) \leq 0$. Thus, for all $z_0 \in [0,1]$,

$$f(z_0) \leq \lim_{z \searrow 0} f(z) \overset{\mathsf{L'H}}{=} \lim_{z \searrow 0} \frac{a(1+z)^{a-1} + a(1-z)^{a-1}}{1} = 2a\,,$$

where $\mathsf{L'H}$ indicates the use of L'Hospital's rule. Furthermore, $f(z_0) \geq f(1) = 2^a$, which implies the lower bound. This finishes the proof for (i). The other parts follow immediately. □

G.4 Merging Lemma

In many cases (i.e., with additional assumption on a, b, c, r or s), we prove the inequality of Lemma 13 by applying first a merging lemma to the LHS to reduce the four summands to two summands of a specific form. Then we apply the Tight Power Bound. The Merging Lemma is discussed in this section.

G.4.1 Simple Merging Lemma

Lemma 18 (Simple Merging Lemma)**.** Let $\alpha \in [\frac{1}{2}, 1]$, $b \geq 0$, $a, c \in \mathbb{R}$. Then

$$|a|^{2\alpha} - |c|^{2\alpha} - |a-b|^{2\alpha} + |c-b|^{2\alpha} \leq 2^{1-2\alpha}\left((a-c+b)^{2\alpha} - |a-c-b|^{2\alpha}\right)\mathbb{1}_{a-c>0}\,.$$

Proof. For $\tilde{\alpha} \geq 1$, the function $\mathbb{R} \to \mathbb{R}$, $x \mapsto |x|^{\tilde{\alpha}} - |x-1|^{\tilde{\alpha}}$ is increasing. It holds $2\alpha \geq 1$. Thus, if $a \leq c$, then

$$|a|^{2\alpha} - |a-b|^{2\alpha} \leq |c|^{2\alpha} - |c-b|^{2\alpha}\,.$$

This shows the inequality for the case $a \leq c$.

Set $q := a - b$ and define

$$g(b) := |q+b|^{2\alpha} - |c|^{2\alpha} - |q|^{2\alpha} + |c-b|^{2\alpha} - 2\left(\left(\frac{q-c}{2}+b\right)^{2\alpha} - \left(\frac{q-c}{2}\right)^{2\alpha}\right).$$

It holds $g(0) = 0$ and

$$\frac{g'(b)}{2\alpha} = \mathsf{sgn}(q+b)|q+b|^{2\alpha-1} - \mathsf{sgn}(c-b)|c-b|^{2\alpha-1} - 2\left(\frac{q-c}{2}+b\right)^{2\alpha-1}.$$

<u>Case 1: $\mathsf{sgn}(q+b) = +1$, $\mathsf{sgn}(c-b) = +1$:</u>

$$\frac{g'(b)}{2\alpha} = (q+b)^{2\alpha-1} - (c-b)^{2\alpha-1} - 2\left(\frac{q-c}{2}+b\right)^{2\alpha-1},$$

$$(q+b)^{2\alpha-1} - (c-b)^{2\alpha-1} \leq (q+b-(c-b))^{2\alpha-1} \leq 2\left(\frac{q-c}{2}+b\right)^{2\alpha-1}.$$

<u>Case 2: $\mathsf{sgn}(q+b) = -1$, $\mathsf{sgn}(c-b) = -1$:</u>

$$\frac{g'(b)}{2\alpha} = (b-c)^{2\alpha-1} - (-q-b)^{2\alpha-1} - 2\left(\frac{q-c}{2}+b\right)^{2\alpha-1},$$

$$(b-c)^{2\alpha-1} - (-q-b)^{2\alpha-1} \leq (b-c-(-q-b))^{2\alpha-1} \leq 2\left(\frac{q-c}{2}+b\right)^{2\alpha-1}.$$

<u>Case 3: $\mathsf{sgn}(q+b) = +1$, $\mathsf{sgn}(c-b) = -1$:</u>

$$\frac{g'(b)}{2\alpha} = (q+b)^{2\alpha-1} + (b-c)^{2\alpha-1} - 2\left(\frac{q-c}{2}+b\right)^{2\alpha-1},$$

$$(q+b)^{2\alpha-1} + (b-c)^{2\alpha-1} \leq 2\left(\frac{q-c}{2}+b\right)^{2\alpha-1}.$$

<u>Case 4: $\mathsf{sgn}(q+b) = -1$, $\mathsf{sgn}(c-b) = +1$:</u>

$$\frac{g'(b)}{2\alpha} = -(-q-b)^{2\alpha-1} - (c-b)^{2\alpha-1} - 2\left(\frac{q-c}{2}+b\right)^{2\alpha-1},$$

$$-(-q-b)^{2\alpha-1} - (c-b)^{2\alpha-1} \leq 0.$$

<u>Together:</u> In every case, we have $g'(b) \leq 0$ and $g(0) = 0$. Thus, $g(b) \leq 0$. □

G.4.2 $ra - sc$–Merging Lemma

Lemma 19. Let $\alpha \in [0, 1]$.

(i) Let $b, c \geq 0$, $s \in [-1, 1]$. Assume $2sc \leq b$. Then

$$-c^{2\alpha} + (c^2 - 2scb + b^2)^\alpha \leq -|sc|^{2\alpha} + |sc - b|^{2\alpha}.$$

(ii) Let $a, b \geq 0$, $r \in [-1, 1]$. Assume $2ra \geq b$. Then

$$a^{2\alpha} - (a^2 - 2rab + b^2)^\alpha \leq |ra|^{2\alpha} - |ra - b|^{2\alpha}.$$

Proof. The function $t \mapsto (t+1)^\alpha - t^\alpha$, $t \geq 0$ is non-increasing for all $\alpha \in [0, 1]$. It holds $0 \leq s^2c^2 \leq c^2$ and $x := -2scb + b^2 \geq 0$. Thus,

$$\left(c^2 + x\right)^\alpha - \left(c^2\right)^\alpha \leq \left((sc)^2 + x\right)^\alpha - \left((sc)^2\right)^\alpha.$$

Thus,

$$(c^2 - 2scb + b^2)^\alpha - c^{2\alpha} \leq |-sc + b|^{2\alpha} - |sc|^{2\alpha}.$$

For the second part, set $x := 2rab - b^2$, $y := a^2 - x \geq 0$, $\tilde{y} := |ra|^2 - x \geq 0$. The condition $2ra \geq b$ implies $x \geq 0$. Thus, as before,

$$a^{2\alpha} - (a^2 - 2rab + b^2)^\alpha = (y + x)^\alpha - y^\alpha \leq (\tilde{y} + x)^\alpha - \tilde{y}^\alpha = |ra|^{2\alpha} - |ra - b|^{2\alpha}.$$

□

Lemma 20 ($ra - sc$–Merging Lemma)**.** Let $\alpha \in [\frac{1}{2}, 1]$. Let $a, b, c \geq 0$, $r, s \in [-1, 1]$.

(i) Assume $2ra \geq b$, $s \in \{-1, 1\}$. Then

$$\begin{aligned} &a^{2\alpha} - c^{2\alpha} - (a^2 - 2rab + b^2)^\alpha + (c^2 - 2scb + b^2)^\alpha \\ &\leq 2^{1-2\alpha}\Big((ra - sc + b)^{2\alpha} - |ra - sc - b|^{2\alpha}\Big)\mathbb{1}_{ra-sc>0}. \end{aligned}$$

(ii) Assume $b \geq 2sc$, $r \in \{-1, 1\}$. Then

$$\begin{aligned} &a^{2\alpha} - c^{2\alpha} - (a^2 - 2rab + b^2)^\alpha + (c^2 - 2scb + b^2)^\alpha \\ &\leq 2^{1-2\alpha}\Big((ra - sc + b)^{2\alpha} - |ra - sc - b|^{2\alpha}\Big)\mathbb{1}_{ra-sc>0}. \end{aligned}$$

(iii) Assume $2ra \geq b \geq 2sc$. Then

$$\begin{aligned} &a^{2\alpha} - c^{2\alpha} - (a^2 - 2rab + b^2)^\alpha + (c^2 - 2scb + b^2)^\alpha \\ &\leq 2^{1-2\alpha}\left((ra - sc + b)^{2\alpha} - |ra - sc - b|^{2\alpha}\right)\mathbb{1}_{ra-sc>0}. \end{aligned}$$

Proof. The lemma above and the simple merging lemma imply

$$\begin{aligned}
&a^{2\alpha} - c^{2\alpha} - (a^2 - 2rab + b^2)^\alpha + (c^2 - 2scb + b^2)^\alpha \\
&\leq (ra)^{2\alpha} - (sc)^{2\alpha} - (ra - b)^{2\alpha} + (sc - b)^{2\alpha} \\
&\leq 2^{1-2\alpha}\left((ra - sc + b)^{2\alpha} - |ra - sc - b|^{2\alpha}\right)\mathbb{1}_{ra-sc>0}\,.
\end{aligned}$$

□

G.4.3 $a - sc$–Merging Lemma

Lemma 20 covers the case $\frac{1}{2}b \geq sc$. The following lemma covers $\frac{1}{2}b \leq sc$ under the additional restriction $sc \leq a - b$.

Lemma 21 ($a - sc$–Merging Lemma)**.** Let $\alpha \in [\frac{1}{2}, 1]$. Let $a, b, c \geq 0$, $s \in [-1, 1]$. Assume $\frac{1}{2}b \leq sc \leq a - b$. Then

$$a^{2\alpha} - c^{2\alpha} - (a - b)^{2\alpha} + (c^2 - 2scb + b^2)^\alpha \leq 2^{1-2\alpha}\Big((a - sc + b)^{2\alpha} - (a - sc - b)^{2\alpha}\Big)\,.$$

Proof. Set $\delta := a - b$. Define

$$\begin{aligned}
f(\delta) = (\delta + b)^{2\alpha} - c^{2\alpha} - \delta^{2\alpha} + (c^2 - 2scb + b^2)^\alpha - \\
2\left(\left(\frac{\delta - sc}{2} + b\right)^{2\alpha} - \left(\frac{\delta - sc}{2}\right)^{2\alpha}\right)\,.
\end{aligned}$$

Then

$$\frac{f'(\delta)}{2\alpha} = (\delta + b)^{2\alpha-1} - \delta^{2\alpha-1} - \left(\frac{\delta - sc}{2} + b\right)^{2\alpha-1} + \left(\frac{\delta - sc}{2}\right)^{2\alpha-1}\,.$$

It holds

$$\begin{aligned}
\delta + b &\geq \delta\,, \\
\frac{\delta - sc}{2} &\leq \frac{\delta - sc}{2} + b\,, \\
(\delta + b) + \frac{\delta - sc}{2} &= \delta + \left(\frac{\delta - sc}{2} + b\right)\,.
\end{aligned}$$

Thus,

$$(\delta + b)^{2\alpha-1} + \left(\frac{\delta - sc}{2}\right)^{2\alpha-1} \leq \delta^{2\alpha-1} + \left(\frac{\delta - sc}{2} + b\right)^{2\alpha-1}\,.$$

Thus, $f'(\delta) \leq 0$.
The next lemma shows $f(sc) \leq 0$. Thus, $f(\delta) \leq 0$ for all $\delta \geq sc$. □

Lemma 22. Let $x, a, b, c \geq 0$. Assume $b \leq 2x$, $x + b \geq c$, $x \leq c$. Then

$$(x+b)^{2\alpha} + (c^2 - 2xb + b^2)^\alpha \leq c^{2\alpha} + x^{2\alpha} + 2b^{2\alpha}\,.$$

Proof. Define

$$\begin{aligned} g(x) &:= (x+b)^{2\alpha} + (c^2 - 2xb + b^2)^\alpha - c^{2\alpha} - x^{2\alpha} - 2b^{2\alpha}\,, \\ h(x) &:= \frac{g'(x)}{2\alpha} = (x+b)^{2\alpha-1} - x^{2\alpha-1} - b(c^2 - 2xb + b^2)^{\alpha-1}\,. \end{aligned}$$

It holds

$$h'(x) = (2\alpha-1)(x+b)^{2\alpha-2} - (2\alpha-1)x^{2\alpha-2} + 2(\alpha-1)b^2(c^2 - 2xb + b^2)^{\alpha-2}\,.$$

As $2\alpha - 2 \leq 0$ and $2\alpha - 1 \geq 0, (2\alpha-1)(x+b)^{2\alpha-2} - (2\alpha-1)x^{2\alpha-2} \leq 0$. As $\alpha - 1 \leq 0$, $2(\alpha-1)b^2(c^2 - 2xb + b^2)^{\alpha-2} \leq 0$. Thus, $h'(x) \leq 0$.
It holds $x \geq x_{\mathsf{min}} := \max(\frac{b}{2}, c - b)$. For checking $h(x_{\mathsf{min}}) \leq 0$ and $g(x_{\mathsf{min}}) \leq 0$, we distinguish $x_{\mathsf{min}} = c - b$ and $x_{\mathsf{min}} = \frac{b}{2}$.
<u>Case 1, $c - b \leq \frac{b}{2}$:</u>
If $c - b \leq \frac{b}{2}$, then $c \leq \frac{3}{2}b \leq (1+\sqrt{3})b$, $c^2 - 2cb - 2b^2 \leq 0$, and

$$\begin{aligned} h(c-b) &= c^{2\alpha-1} - (c-b)^{2\alpha-1} - b(c^2 - 2(c-b)b + b^2)^{\alpha-1} \\ &= c^{2\alpha-1} - (c-b)^{2\alpha-1} - b(c^2 - 2cb - b^2)^{\alpha-1} \\ &\leq b^{2\alpha-1} - b(c^2 - 2cb - b^2)^{\alpha-1} \\ &\leq b\left(\left(b^2\right)^{\alpha-1} - \left(c^2 - 2cb - b^2\right)^{\alpha-1}\right) \end{aligned}$$

And, thus, $h(c-b) \leq 0$ as $c^2 - 2cb - 2b^2 \leq 0$. Furthermore,

$$\begin{aligned} g(c-b) &= -(c-b)^{2\alpha} + (c^2 - 2cb - b^2)^\alpha - 2b^{2\alpha} \\ &= -(c-b)^{2\alpha} + (c^2 - 2cb - 2b^2 + b^2)^\alpha - 2b^{2\alpha} \\ &\leq -(c-b)^{2\alpha} + b^{2\alpha} - 2b^{2\alpha} \\ &= -(c-b)^{2\alpha} - b^{2\alpha} \\ &\leq 0\,. \end{aligned}$$

Thus, $g(x) \leq 0$ for all valid x.
<u>Case 2, $c - b \geq \frac{b}{2}$:</u>

If $c - b \geq \frac{b}{2}$, then $c \geq b$ and

$$\begin{aligned}
h\left(\frac{b}{2}\right) &= \left(\frac{3}{2}b\right)^{2\alpha-1} - \left(\frac{1}{2}b\right)^{2\alpha-1} - b(c^2)^{\alpha-1} \\
&\leq \left(\left(\frac{3}{2}\right)^{2\alpha-1} - \left(\frac{1}{2}\right)^{2\alpha-1}\right) b^{2\alpha-1} - b(c^2)^{\alpha-1} \\
&\leq b^{2\alpha-1} - b(c^2)^{\alpha-1} \\
&\leq b^{2\alpha-1} - b(b^2)^{\alpha-1} \\
&\leq 0\,,
\end{aligned}$$

$$\begin{aligned}
g\left(\frac{b}{2}\right) &= \left(\frac{3}{2}b\right)^{2\alpha} - c^{2\alpha} - \left(\frac{1}{2}b\right)^{2\alpha} + c^{2\alpha} - 2b^{2\alpha} \\
&\leq \left(\left(\frac{9}{4}\right)^{\alpha} - \left(\frac{1}{2}\right)^{\alpha} - 2\right) b^{2\alpha} \\
&\leq 0\,.
\end{aligned}$$

Thus, $g(x) \leq 0$ for all valid x. □

G.5 Application of Tight Power Bound and Merging Lemma

Whenever a Merging Lemma holds, we apply it as a first step and then use the Tight Power Bound, Lemma 17, to obtain

$$\begin{aligned}
&a^{2\alpha} - c^{2\alpha} - (a^2 - 2rab + b^2)^{\alpha} + (c^2 - 2scb + b^2)^{\alpha} \\
&\leq 2^{1-2\alpha}\left((ra - sc + b)^{2\alpha} - |ra - sc - b|^{2\alpha}\right) \\
&\leq 4\alpha 2^{1-2\alpha}(ra - sc)^{2\alpha-1} b\,.
\end{aligned}$$

In particular, we have finished the proof of Lemma 13 in following cases:

- $ra \geq sc$ and $s, r \in \{-1, 1\}$: Lemma 18,
- $2ra \geq b$ and $s \in \{-1, 1\}$; or $b \geq 2sc$ and $r \in \{-1, 1\}$; or $2ra \geq b \geq 2sc$: Lemma 20,
- $\frac{1}{2}b \leq sc \leq a - b$ and $r = 1$: Lemma 21.

G.6 The Case $ra - sc \geq |a - c|$

G.6.1 The Case $a \geq c$

First we prove to simple lemmas, before we solve this case.

Lemma 23. Let $a \geq b \geq 0$, $d \geq c \geq 0$, and $\alpha \in [0,1]$. Then

$$a^\alpha - b^\alpha - c^\alpha + d^\alpha \leq 2^{1-\alpha}(a-b-c+d)^\alpha .$$

Proof. As $a \geq b$, $d \geq c$, $\alpha \leq 1$,

$$a^\alpha - b^\alpha + d^\alpha - c^\alpha \leq (a-b)^\alpha + (d-c)^\alpha .$$

Furthermore, by concavity of $x \mapsto x^\alpha$,

$$(a-b)^\alpha + (d-c)^\alpha \leq 2^{1-\alpha}(a-b+d-c)^\alpha .$$

□

Lemma 24. Let $a \geq b \geq c \geq d \geq 0$, $a+d \geq b+c$, and $\alpha \in [0,1]$. Then

$$a^\alpha - b^\alpha - c^\alpha + d^\alpha \leq (a-b-c+d)^\alpha .$$

Proof. Define $f(x,y) = x^\alpha + y^\alpha - (x+y)^\alpha$ for $x, y \geq 0$. Then $\partial_x f(x,y) = \alpha(x^{\alpha-1} - (x+y)^{\alpha-1}) \geq 0$ and similarly $\partial_y f(x,y) \geq 0$. Set $\delta := a-b$ and $\epsilon := c-d$. The assumptions ensure $\delta \geq \epsilon \geq 0$. Then,

$$f(b,\delta) \geq f(b,\epsilon) \geq f(d,\epsilon) .$$

Thus,

$$\begin{aligned} 0 &\geq f(d,\epsilon) - f(b,\delta) \\ &= d^\alpha + \epsilon^\alpha - (d+\epsilon)^\alpha - b^\alpha - \delta^\alpha + (b+\delta)^\alpha \\ &= d^\alpha + \epsilon^\alpha - c^\alpha - b^\alpha - \delta^\alpha + a^\alpha . \end{aligned}$$

With this we get

$$\begin{aligned} d^\alpha - c^\alpha - b^\alpha + a^\alpha &\leq \delta^\alpha - \epsilon^\alpha \\ &\leq (\delta - \epsilon)^\alpha \\ &= (a-b-c+d)^\alpha . \end{aligned}$$

□

For $a \geq c$, the remaining case is solved by following lemma.

Lemma 25. Let $\alpha \in [0,1]$. Let $a, b, c \geq 0$, $s \in [-1,1]$. Assume $\frac{1}{2}b \leq sc$, $sc \geq a-b$,

and $a \geq c$. Then

$$a^{2\alpha} - c^{2\alpha} - (a-b)^{2\alpha} + (c^2 - 2scb + b^2)^\alpha \leq 2b^\alpha (a - sc)^\alpha \leq 2b(a - sc)^{2\alpha - 1}\,.$$

Proof. Because $a \geq c$ and $\frac{1}{2}b \leq sc$, we have $a - b \geq a - 2sc \geq a - 2c \geq -c$. Hence, $a^2 \geq \max(c^2, (a-b)^2)$. Thus, applying either Lemma 23 (if $c^2 - 2scb + b^2$ is larger then either c^2 or $(a-b)^2$) or Lemma 24 yields

$$\begin{aligned}
&a^{2\alpha} - c^{2\alpha} - (a-b)^{2\alpha} + (c^2 - 2scb + b^2)^\alpha \\
&\leq 2^{1-\alpha}\left(a^2 - c^2 - (a-b)^2 + (c^2 - 2scb + b^2)\right)^\alpha \\
&= 2b^\alpha (a - sc)^\alpha\,.
\end{aligned}$$

The condition $0 \leq a - sc \leq b$ implies

$$2b^\alpha (a - sc)^\alpha \leq 2b(a - sc)^{2\alpha - 1}\,.$$

□

G.6.2 The Case $a \leq c$

For the case $c \geq a$, we only need $ra - sc \geq c - a$ (for $r = 1$), i.e., $sc \leq 2a - c$. Assume $c \geq a$, $sc \geq a - b$, $\frac{1}{2}b \leq sc$, and $sc \leq 2a - c$. Then

$$\begin{aligned}
c^2 &\geq c^2 - 2scb + b^2 \geq (a-b)^2 \\
c^2 &\geq a^2 \geq (a-b)^2
\end{aligned}$$

We distinguish $\frac{1}{2}b \leq a - sc$ and $\frac{1}{2}b \geq a - sc$.

Lemma 26 ($\frac{1}{2}b \leq a - sc$)**.** Let $\alpha \in [0,1]$. Let $a, b, c \geq 0$, $s \in [-1, 1]$. Assume $\frac{1}{2}b \leq sc$, $sc \geq a - b$, $c \geq a$, $sc \leq 2a - c$, and $\frac{1}{2}b \leq a - sc$. Then

$$a^{2\alpha} - c^{2\alpha} - (a-b)^{2\alpha} + (c^2 - 2scb + b^2)^\alpha \leq 2b^\alpha (a - sc)^\alpha\,.$$

Proof. The conditions imply

$$\max\left(\frac{1}{2}b,\, a - b\right) \leq sc \leq \min\left(a - \frac{1}{2}b,\, 2a - c\right)\,.$$

In particular, $\frac{1}{2}b \leq a - \frac{1}{2}b$, and $a - b \leq 2a - c$. Thus, $a + b \geq c \geq a \geq b$.

Fix $a, b, c \geq 0$. Assume $c \geq a$. Let $x \in [0, a)$. Then $a - x > 0$ and $c^2 - 2xb + b^2 > 0$. Define

$$f(x) := a^{2\alpha} - c^{2\alpha} - (a-b)^{2\alpha} + (c^2 - 2xb + b^2)^\alpha - 2b^\alpha (a - x)^\alpha\,.$$

It holds

$$\frac{f'(x)}{2\alpha b} = -(c^2 - 2xb + b^2)^{\alpha-1} + b^{\alpha-1}(a-x)^{\alpha-1}\,.$$

Furthermore,

$$\begin{aligned}
& a - c \le 0 \le (c-b)^2 \\
\Rightarrow\quad & 2ab - 2cb \le 0 \le c^2 + b^2 - 2cb \\
\Rightarrow\quad & xb \le 2ab - cb \le c^2 + b^2 - cb \le c^2 + b^2 - ab \\
\Rightarrow\quad & ab - xb \le c^2 - 2xb + b^2 \\
\Rightarrow\quad & b^{\alpha-1}(a-x)^{\alpha-1} \ge (c^2 - 2xb + b^2)^{\alpha-1} \\
\Rightarrow\quad & f'(x) \ge 0\,.
\end{aligned}$$

We define

$$g(c) := f(a - \frac{1}{2}b) = a^{2\alpha} - c^{2\alpha} - (a-b)^{2\alpha} + (c^2 - 2ab + 2b^2)^{\alpha} - 2^{1-\alpha}b^{2\alpha}\,.$$

Thus,

$$\frac{g'(c)}{2\alpha c} = -c^{2\alpha-2} + \left(c^2 - 2ab + 2b^2\right)^{\alpha-1} \ge 0\,.$$

Define

$$\begin{aligned}
h(a,b) &:= g(a+b) \\
&= a^{2\alpha} - (a+b)^{2\alpha} - (a-b)^{2\alpha} + (a^2 + 3b^2)^{\alpha} - 2^{1-\alpha}b^{2\alpha} \\
&= a^{2\alpha} - (a+b)^{2\alpha} - (a-b)^{2\alpha} + (a^2 + 3b^2)^{\alpha} - 2\left(\frac{b^2}{2}\right)^{\alpha}\,.
\end{aligned}$$

The next lemma shows $h(a,b) \le 0$ for $a \ge b$. Thus, $g(c) \le 0$. Thus, $f(x) \le 0$. □

Lemma 27. Let $\alpha \in [0,1]$, $a, b \ge 0$. Assume $a \ge b$. Then

$$a^{2\alpha} + (a^2 + 3b^2)^{\alpha} \le (a+b)^{2\alpha} + (a-b)^{2\alpha} + 2\left(\frac{b^2}{2}\right)^{\alpha}\,.$$

Proof. Set $x = \frac{b}{a} \in [0,1]$. Define

$$f(\alpha, x) = 1 + \left(1 + 3x^2\right)^{\alpha} - (1+x)^{2\alpha} - (1-x)^{2\alpha} - 2\left(\frac{x^2}{2}\right)^{\alpha}\,.$$

It holds

$$\begin{aligned}\frac{\partial_\alpha f(\alpha,x)}{\alpha} &= \left(1+3x^2\right)^\alpha \log\left(1+3x^2\right) - (1+x)^{2\alpha}\log\left((1+x)^2\right)\\ &\quad - (1-x)^{2\alpha}\log\left((1-x)^2\right) - 2\left(\frac{x^2}{2}\right)^\alpha \log\left(\frac{x^2}{2}\right)\\ &=: g(x,\alpha)\,,\end{aligned}$$

$$\begin{aligned}\frac{\partial_\alpha g(\alpha,x)}{\alpha} &= \left(1+3x^2\right)^\alpha \log\left(1+3x^2\right)^2 - (1+x)^{2\alpha}\log\left((1+x)^2\right)^2\\ &\quad - (1-x)^{2\alpha}\log\left((1-x)^2\right)^2 - 2\left(\frac{x^2}{2}\right)^\alpha \log\left(\frac{x^2}{2}\right)^2\\ &\le \left(1+3x^2\right)^\alpha \log\left(1+3x^2\right)^2 - (1+x)^{2\alpha}\log\left((1+x)^2\right)^2\\ &=: h(x,\alpha)\,.\end{aligned}$$

For $x \in [0,1]$, it holds $1+3x^2 \le (1+x)^2$. Thus,

$$\left(\frac{1+3x^2}{(1+x)^2}\right)^\alpha \le 1 \le \left(\frac{\log((1+x)^2)}{\log(1+3x^2)}\right)^2 .$$

Thus,

$$\left(1+3x^2\right)^\alpha \log\left(1+3x^2\right)^2 \le (1+x)^{2\alpha}\log\left((1+x)^2\right)^2$$

Thus, $h(x,\alpha) \le 0$ and $\partial_\alpha g(\alpha,x) \le 0$. Thus, $g(\alpha,x) \ge g(1,x)$ and

$$\begin{aligned}g(1,x) &= \left(1+3x^2\right)\log\left(1+3x^2\right) - (1+x)^2\log\left((1+x)^2\right)\\ &\quad - (1-x)^2\log\left((1-x)^2\right) - 2\left(\frac{x^2}{2}\right)\log\left(\frac{x^2}{2}\right)\\ &=: \ell(x)\,.\end{aligned}$$

The next lemma shows $\ell(x) \ge 0$. Thus, $g(\alpha,x) \ge g(1,x) \ge 0$. Thus $\partial_\alpha f(\alpha,x) \ge 0$. Thus, $f(\alpha,x) \le f(1,x)$ and

$$f(1,x) = 1 + \left(1+3x^2\right) - (1+x)^2 - (1-x)^2 - 2\left(\frac{x^2}{2}\right) = 0\,.$$

Thus, $f(\alpha,x) \le 0$. □

Lemma 28. Let $x \in \mathbb{R}$. Define

$$f(x) := \left(1+3x^2\right)\log\left(1+3x^2\right) - (1+x)^2\log\left((1+x)^2\right) - (1-x)^2\log\left((1-x)^2\right) - x^2\log\left(\frac{x^2}{2}\right) .$$

Then

$$f(x) \geq 0 .$$

Proof. Let us first calculate some derivatives:

$$\begin{aligned}
f(x) &= x^2\log\left(\frac{2\left(1+3x^2\right)^3}{x^2\left(1-x^2\right)^2}\right) - 4x\log\left(\frac{1+x}{1-x}\right) + \log\left(\frac{1+3x^2}{\left(1-x^2\right)^2}\right) , \\
f'(x) &= 2x\log\left(\frac{2\left(1+3x^2\right)^3}{x^2\left(1-x^2\right)^2}\right) - 4\log\left(\frac{1+x}{1-x}\right) , \\
f''(x) &= 2\log\left(\frac{2\left(1+3x^2\right)^3}{x^2\left(1-x^2\right)^2}\right) - \frac{12}{1+3x^2} , \\
f'''(x) &= \frac{4\left(9x^4+24x^2-1\right)}{x(1-x)(1+x)\left(3x^2+1\right)^2} , \\
f^{(4)}(x) &= \frac{4\left(81x^8+324x^6-186x^4+36x^2+1\right)}{x^2\left(1-x^2\right)^2\left(3x^2+1\right)^3} .
\end{aligned}$$

We consider the cases $x \in [0, \frac{1}{10}]$ and $x \in [\frac{1}{10}, 1]$ separately, and start with the latter. For $x_0 \in (0,1)$, define

$$g_{x_0}(x) := f(x_0) + f'(x_0)\left(x-x_0\right) + \frac{1}{2}f''(x_0)\left(x-x_0\right)^2 + \frac{1}{6}f'''(x_0)\left(x-x_0\right)^3 .$$

Then the Taylor-Expansion for $x \in (0,1)$ is

$$f(x) = g_{x_0}(x) + \frac{1}{24}f^{(4)}(\xi(x,x_0))\left(x-x_0\right)^4 ,$$

with suitable $\xi(x,x_0)$. One can show that $81x^4+324x^3-186x^2+36x^1+1 > 0$ for all $x \geq 0$. In particular, $f^{(4)}(x) \geq 0$ for $x \in [0,1]$. Thus,

$$f(x) \geq g_{x_0}(x) .$$

We use $x_0 = \frac{1}{3}$:

$$\begin{aligned} f\left(\frac{1}{3}\right) &= \frac{10}{3}\log(3) - \frac{47}{9}\log(2)\,, \\ f'\left(\frac{1}{3}\right) &= 2\log(3) - \frac{10}{3}\log(2)\,, \\ f''\left(\frac{1}{3}\right) &= -9 + 2\log(2) + 6\log(3)\,, \\ f'''\left(\frac{1}{3}\right) &= \frac{27}{2}\,. \end{aligned}$$

One can show that $g_{\frac{1}{3}}(x) \geq 0$ for $x \geq \frac{1}{10}$. Thus, $f(x) \geq 0$ for $x \in [\frac{1}{10}, 1]$.

The case $x \in [0, \frac{1}{10}]$ is left. One can show

$$4\log\left(\frac{1+x}{1-x}\right) \leq 10x \leq 2x\log\left(\frac{2\left(1+3x^2\right)^3}{x^2\left(1-x^2\right)^2}\right)$$

for $x \in [0, \frac{1}{10}]$. This implies $f'(x) \geq 0$. Together with $f(0) = 0$, this yields $f(x) \geq 0$ for $x \in [0, \frac{1}{10}]$. □

Lemma 29 ($\frac{1}{2}b \geq a - sc$ and $a \leq b$)**.** Let $\alpha \in [\frac{1}{2}, 1]$. Let $a, b, c \geq 0$, $s \in [-1, 1]$. Assume

$$0 \leq c - a \leq a - sc \leq \frac{b}{2} \leq sc \leq 2a - c \leq a \leq c$$

and $a \leq b$. Then

$$a^{2\alpha} - c^{2\alpha} - |a-b|^{2\alpha} + (c^2 - 2scb + b^2)^\alpha \leq 2b(a - sc)^{2\alpha - 1}\,.$$

Proof. Define

$$f(a,b,y,w) := a^{2\alpha} - (y+a)^{2\alpha} - (b-a)^{2\alpha} + ((y+a)^2 - 2wb + b^2)^\alpha - 2b(a-w)^{2\alpha-1}\,.$$

It holds

$$\partial_y f(a,b,y,w) = -2\alpha(y+a)^{2\alpha-1} + \alpha\left(2(y+a)\right)\left((y+a)^2 - 2wb + b^2\right)^{\alpha-1}\,.$$

Because of $\frac{b}{2} \leq w$, we have

$$(y+a)^2 \geq (y+a)^2 - 2wb + b^2\,.$$

Thus $\partial_y f(a,b,y,w) \geq 0$. Thus, for $y \in [0, a-w]$, it holds $f(a,b,y,w) \leq f(a,b,a-w,w)$.

It holds

$$
\begin{aligned}
&f(a,b,a-w,w)\\
&= a^{2\alpha} - (2a-w)^{2\alpha} - (b-a)^{2\alpha} + ((2a-w)^2 - 2wb + b^2)^{\alpha} - 2b(a-w)^{2\alpha-1}\\
&=: g(a,b,w)\,,
\end{aligned}
$$

$$
\begin{aligned}
&\partial_b g(a,b,w)\\
&= -2\alpha(b-a)^{2\alpha-1} + 2\alpha(b-w)((2a-w)^2 - 2wb + b^2)^{\alpha-1} - 2(a-w)^{2\alpha-1}\\
&\le 2\alpha h(a,b,w)\,,
\end{aligned}
$$

$$
h(a,b,w) = -(b-a)^{2\alpha-1} + (b-w)((2a-w)^2 - 2wb + b^2)^{\alpha-1} - (a-w)^{2\alpha-1}\,.
$$

The conditions $0 \le a-w \le \frac{b}{2} \le w$ and $a \le b$ imply $w \le a \le b$. It holds,

$$
\begin{aligned}
&\partial_a^2 h(a,b,w)\\
=&-(2\alpha-1)(2\alpha-2)(b-a)^{2\alpha-3}\\
&+4(\alpha-1)(\alpha-2)(2a-w)^2(b-w)((2a-w)^2 - 2wb + b^2)^{\alpha-3}\\
&-(2\alpha-1)(2\alpha-2)(a-w)^{2\alpha-3}\\
\ge&0\,,
\end{aligned}
$$

$$
\begin{aligned}
h(w,b,w) &= -(b-w)^{2\alpha-1} + (b-w)((2w-w)^2 - 2wb + b^2)^{\alpha-1} - (w-w)^{2\alpha-1}\\
&= -(b-w)^{2\alpha-1} + (b-w)^{2\alpha-1} = 0\,,
\end{aligned}
$$

$$
h(b,b,w) = -(b-b)^{2\alpha-1} + (b-b)((2b-w)^2 - 2wb + b^2)^{\alpha-1} - (b-b)^{2\alpha-1} \le 0\,.
$$

Thus, $h(a,b,w) \le 0$ for all $a \in [w,b]$. Thus, $\partial_b g(a,b,w) \le 0$. The conditions for g are $0 \le a-w \le \frac{b}{2} \le w \le a \le b$. As $a \le b$ and $\frac{b}{2} \le w$, we have $a \le 2w$ and thus $a \ge 2a-2w$.

$$
\begin{aligned}
g(a,a,w) &= a^{2\alpha} - (2a-w)^{2\alpha} - (a-a)^{2\alpha} + ((2a-w)^2 - 2wa + a^2)^{\alpha} -\\
&\quad 2a(a-w)^{2\alpha-1}\\
&= a^{2\alpha} - (2a-w)^{2\alpha} + (5a^2 - 6aw + w^2)^{\alpha} - 2a(a-w)^{2\alpha-1}\,.
\end{aligned}
$$

Set $w = a-y$, $y \in [0,a]$. It holds

$$
\begin{aligned}
\ell(a,y) &:= a^{2\alpha} + (5a^2 - 6a(a-y) + (a-y)^2)^{\alpha} - (a+y)^{2\alpha} - 2ay^{2\alpha-1}\\
&= a^{2\alpha} + (4ay + y^2)^{\alpha} - (a+y)^{2\alpha} - 2ay^{2\alpha-1}\,.
\end{aligned}
$$

It holds $g(a,a,w) = \ell(a,y)$. Under the condition $0 \le y \le a$, it holds$\ell(a,y) \le 0$, cf next lemma. □

Lemma 30. Let $\alpha \in [\frac{1}{2}, 1]$, $a, y \geq 0$. Assume $y \leq a$. Then

$$a^{2\alpha} + (4ay + y^2)^\alpha \leq (a+y)^{2\alpha} + 2ay^{2\alpha-1}\,.$$

Proof. For $y \leq a$ define

$$g(a,y) := 2a^{2\alpha-1} + 4^\alpha y^\alpha a^{\alpha-1} - 2(a+y)^{2\alpha-1} - 2y^{2\alpha-1}\,.$$

It holds

$$\partial_a g(a,y) = 2(2\alpha-1)a^{2\alpha-2} + 4^\alpha(\alpha-1)y^\alpha a^{\alpha-2} - 2(2\alpha-1)(a+y)^{2\alpha-2}\,,$$

$$\partial_y\partial_a g(a,y) = 4^\alpha(\alpha-1)\alpha y^{\alpha-1}a^{\alpha-2} - 2(2\alpha-1)(2\alpha-2)(a+y)^{2\alpha-3}\,.$$

Set $z := \frac{y}{a} \in [0,1]$. Then

$$\frac{\partial_y\partial_a g(a,y)}{a^{2\alpha-3}} = (\alpha-1)\left(4^\alpha\alpha z^{\alpha-1} - 4(2\alpha-1)(1+z)^{2\alpha-3}\right)\,.$$

One can show

$$\frac{4^\alpha\alpha}{4(2\alpha-1)} \geq 1 \geq \frac{(1+z)^{2\alpha-3}}{z^{\alpha-1}}\,,$$

$$4^\alpha\alpha z^{\alpha-1} \geq 4(2\alpha-1)\,(1+z)^{2\alpha-3}\,.$$

Thus, $\partial_y\partial_a g(a,y) \leq 0$. Thus, $\partial_a g(a,y) \leq \partial_a g(a,0)$

$$\partial_a g(a,0) = 2(2\alpha-1)a^{2\alpha-2} - 2(2\alpha-1)(a)^{2\alpha-2} = 0$$

Thus, $\partial_a g(a,y) \leq 0$. Thus, $g(a,y) \leq g(y,y)$, and

$$g(y,y) = \left(2 + 4^\alpha - 2^{2\alpha} - 2\right)y^{2\alpha-1} = 0\,.$$

Thus, $g(a,y) \leq 0$. It holds

$$\begin{aligned}
&a^{2\alpha} + (4ay+y^2)^\alpha - (a+y)^{2\alpha} - 2ay^{2\alpha-1}\\
&\leq a^{2\alpha} + (4a)^\alpha y^\alpha + y^{2\alpha} - (a+y)^{2\alpha} - 2ay^{2\alpha-1}\\
&=: f(a,y)\,,
\end{aligned}$$

$$\partial_a f(a,y) = 2\alpha a^{2\alpha-1} + 4^\alpha\alpha y^\alpha a^{\alpha-1} - 2\alpha(a+y)^{2\alpha-1} - 2y^{2\alpha-1} \leq \alpha g(a,y) \leq 0\,.$$

Thus,

$$f(a,y) \leq f(y,y) = (1 + 5^\alpha - 4^\alpha - 2)\,y^{2\alpha} \leq 0\,.$$

□

Lemma 31 ($\frac{1}{2}b \geq a - sc$ and $a \geq b$)**.** Let $\alpha \in [\frac{1}{2}, 1]$. Let $a, b, c \geq 0$, $s \in [-1, 1]$. Assume

$$0 \leq c - a \leq a - sc \leq \frac{b}{2} \leq sc \leq 2a - c \leq a \leq c$$

and $a \geq b$. Then

$$a^{2\alpha} - c^{2\alpha} - (a-b)^{2\alpha} + (c^2 - 2scb + b^2)^\alpha \leq 2b(a - sc)^{2\alpha - 1}\,.$$

Proof. Define

$$f(a, b, c, w) := a^{2\alpha} - c^{2\alpha} - (a-b)^{2\alpha} + (c^2 - 2wb + b^2)^\alpha - 2b(a-w)^{2\alpha-1}\,.$$

It holds

$$\partial_c f(a, b, c, w) = 2\alpha \left(-c^{2\alpha-1} + c(c^2 - 2wb + b^2)^{\alpha-1}\right)\,.$$

Because of $2w \geq b$, it holds $\partial_c f(a, b, c, w) \leq 0$.

$$f(a, b, a, w) = -(a-b)^{2\alpha} + (a^2 - 2wb + b^2)^\alpha - 2b(a-w)^{2\alpha-1}\,.$$

Set $x := a - w$. The conditions for x are $0 \leq x \leq \frac{b}{2} \leq w$ and $b \leq x + w$. Define

$$g(x, b, w) := -(x + w - b)^{2\alpha} + ((x+w)^2 - 2wb + b^2)^\alpha - 2bx^{2\alpha-1}\,.$$

It holds

$$\partial_w g(x, b, w) = -2\alpha(x + w - b)^{2\alpha-1} + 2\alpha(x + w - b)((x+w)^2 - 2wb + b^2)^{\alpha-1}\,.$$

It holds

$$(x + w - b)^2 - \left((x+w)^2 - 2wb + b^2\right) = -2bx \leq 0\,.$$

Thus,

$$-2\alpha(x + w - b)^{2\alpha-2} + 2\alpha\left((x+w)^2 - 2wb + b^2\right)^{\alpha-1} \leq 0\,.$$

As $a \geq b$ and thus $x + w - b \geq 0$, $\partial_w g(x, b, w) \leq 0$. It holds $w \geq b - x \geq \frac{b}{2}$ and

$$g(x, b, b - x) = (2xb)^\alpha - 2bx^{2\alpha-1} \leq 0$$

because $x \leq b$. □

G.7 The Case $|a - c| \geq ra - sc$

Lemma 32 (Case 1.2)**.** Let $\alpha \in [\frac{1}{2}, 1]$. Let $a, b, c \geq 0$. Assume $a \geq c$ and $a + c \geq 2b$.

Then

$$a^{2\alpha} - c^{2\alpha} - \left((a-b)^2 + 4cb\right)^{\alpha} + (c+b)^{2\alpha} \leq 8\alpha 2^{-2\alpha} b(a-c)^{2\alpha-1}\,.$$

This lemma follows from the next two lemmas, which split the proof of this inequality into two cases.

Lemma 33 (Case 1.2, Merging)**.** Let $\alpha \in [\frac{1}{2}, 1]$. Let $a, b, c \geq 0$. Assume $a \geq b + c$. Then

$$a^{2\alpha} - c^{2\alpha} - \left((a-b)^2 + 4cb\right)^{\alpha} + (c+b)^{2\alpha} \leq 2^{1-2\alpha}\left((a-c+b)^{2\alpha} - (a-c-b)^{2\alpha}\right)$$

Proof. Set $\delta := a - b \geq c \geq 0$ and define

$$f(b) := (\delta+b)^{2\alpha} - c^{2\alpha} - \left(\delta^2 + 4cb\right)^{\alpha} + (c+b)^{2\alpha} - 2^{1-2\alpha}\left((\delta-c+2b)^{2\alpha} - (\delta-c)^{2\alpha}\right)\,.$$

Then

$$\frac{f'(b)}{2\alpha} = (\delta+b)^{2\alpha-1} - 2c\left(\delta^2+4cb\right)^{\alpha-1} + (c+b)^{2\alpha-1} - 2^{2-2\alpha}(\delta-c+2b)^{2\alpha-1}\,,$$

$$\frac{f'(b)}{2\alpha} \leq (\delta+2b+c)^{2\alpha-1} - (\delta+2b-c)^{2\alpha-1} - 2c\left(\delta^2+4cb\right)^{\alpha-1} =: g(c)\,,$$

$$\begin{aligned} g'(c) = (2\alpha-1)(\delta+2b+c)^{2\alpha-2} + (2\alpha-1)(\delta+2b-c)^{2\alpha-2} - \\ 2\left(\delta^2+4cb\right)^{\alpha-1} - 8cb(\alpha-1)\left(\delta^2+4cb\right)^{\alpha-2}\,, \end{aligned}$$

$$\begin{aligned} &g''(c) \\ &= (2\alpha-1)(2\alpha-2)(\delta+2b+c)^{2\alpha-3} - (2\alpha-1)(2\alpha-2)(\delta+2b-c)^{2\alpha-3} - A\,, \end{aligned}$$

where

$$\begin{aligned} A :=& 2(4b)(\alpha-1)\left(\delta^2+4cb\right)^{\alpha-2} 8b(\alpha-1)\left(\delta^2+4cb\right)^{\alpha-2} + \\ & 8cb(4b)(\alpha-1)(\alpha-2)\left(\delta^2+4cb\right)^{\alpha-3} \\ =& 16b(\alpha-1)\left(\delta^2+4cb\right)^{\alpha-3}\left(\delta^2+4cb+2cb(\alpha-2)\right)\,. \end{aligned}$$

Thus, $g''(c) \geq 0$. It holds $0 \leq c \leq \delta$ and

$$g(0) = (\delta+2b)^{2\alpha-1} - (\delta+2b)^{2\alpha-1} - 0 = 0\,,$$

$$g(\delta) = (2\delta + 2b)^{2\alpha-1} - (2b)^{2\alpha-1} - 2\delta\left(\delta^2 + 4\delta b\right)^{\alpha-1} =: h(\delta)\,,$$

$$h'(\delta) = 2(2\alpha-1)(2\delta+2b)^{2\alpha-2} - 2\left(\delta^2+4\delta b\right)^{\alpha-1} - 2\delta(\alpha-1)(2\delta+4b)\left(\delta^2+4\delta b\right)^{\alpha-2}\,.$$

For $\alpha \in [\frac{1}{2}, 1]$, we have

$$\begin{aligned}
&\left(\delta^2+4\delta b\right)^{\alpha-1} + \delta(\alpha-1)(2\delta+4b)\left(\delta^2+4\delta b\right)^{\alpha-2} \\
&= \left(\delta^2+4\delta b\right)^{\alpha-2}\left(\delta^2+4\delta b+\delta(\alpha-1)(2\delta+4b)\right) \\
&\geq \left(\delta^2+4\delta b\right)^{\alpha-2}\left(\delta^2+4\delta b+-\delta(1\delta+2b)\right) \geq 0\,.
\end{aligned}$$

Thus,

$$h'(\delta) \leq 0\,.$$

It holds

$$h(0) = (2b)^{2\alpha-1} - (2b)^{2\alpha-1} - 0 = 0\,.$$

Thus, $h(\delta) \leq 0$, thus, $g(b=\delta) \leq 0$, thus $g(b) \leq 0$, thus $f'(b) \leq 0$

$$f(0) = (\delta)^{2\alpha} - c^{2\alpha} - (\delta)^{2\alpha} + (c)^{2\alpha} - 2^{1-2\alpha}\left((\delta-c)^{2\alpha} - (\delta-c)^{2\alpha}\right) = 0\,.$$

Thus, $f(b) \leq 0$. □

Lemma 34 (Case 1.2, $b \geq a-c$)**.** Let $\alpha \in [\frac{1}{2}, 1]$. Let $a, b, c \geq 0$. Assume $b \geq a-c \geq 2\max(0, b-c)$. Then

$$a^{2\alpha} - c^{2\alpha} - \left((a-b)^2+4cb\right)^{\alpha} + (c+b)^{2\alpha} \leq 2b^{2\alpha-1}(a-c) \leq 2b(a-c)^{2\alpha-1}\,.$$

Proof. As $b \geq a-c$ and $2\alpha-1 \in [0,1]$, it holds $b^{2\alpha-1}(a-c) \leq b(a-c)^{2\alpha-1}$. Define

$$f(a) := a^{2\alpha} - c^{2\alpha} - \left((a-b)^2+4cb\right)^{\alpha} + (c+b)^{2\alpha} - 2b^{2\alpha-1}(a-c)\,.$$

It holds

$$\begin{aligned}
f'(a) &= 2\alpha a^{2\alpha-1} - 2\alpha(a-b)\left((a-b)^2+4cb\right)^{\alpha-1} - 2b^{2\alpha-1} \\
\text{$a\geq b,c$ and $2\alpha-2\leq 0$} \qquad &\leq 2\alpha a^{2\alpha-1} - 2\alpha(a-b)\,(a+b)^{2\alpha-2} - 2b^{2\alpha-1} \\
&= 2\left(\alpha a^{2\alpha-1} - \alpha\frac{a-b}{a+b}\,(a+b)^{2\alpha-1} - b^{2\alpha-1}\right)\,.
\end{aligned}$$

Set $x = \left(\frac{a-b}{a+b}\right)^{\frac{1}{2\alpha-1}} \leq 1$, $y = \alpha^{\frac{1}{2\alpha-1}} \leq 1$. Then

$$\begin{aligned}
\alpha a^{2\alpha-1} - \alpha \frac{a-b}{a+b}(a+b)^{2\alpha-1} - b^{2\alpha-1} &\leq (ya)^{2\alpha-1} - (xya + xyb)^{2\alpha-1} - b^{2\alpha-1} \\
&\leq (ya - xya - xyb)^{2\alpha-1} - b^{2\alpha-1} \\
&\leq (ya - xya - xyb - b)^{2\alpha-1} \\
&= ((y - xy)a - (xy+1)b)^{2\alpha-1} \\
&\leq (a-b)^{2\alpha-1} \\
&\leq 0\,.
\end{aligned}$$

Thus, $f'(a) \leq 0$. Thus, only need to show $f(b) \leq 0$. Assume $b \geq c$. Then

$$\begin{aligned}
f(b) &= b^{2\alpha} - c^{2\alpha} - (4cb)^{\alpha} + (c+b)^{2\alpha} - 2b^{2\alpha-1}(b-c) \\
&= -c^{2\alpha} - (4cb)^{\alpha} + (c+b)^{2\alpha} - b^{2\alpha} + 2b^{2\alpha-1}c \\
&\leq (c+b)^{2\alpha} - b^{2\alpha} - c^{2\alpha} - (4^{\alpha} - 2)\,(cb)^{\alpha} \\
&= (c+b)^{2\alpha} - (b^{\alpha} - c^{\alpha})^2 - 4^{\alpha}\,(cb)^{\alpha}\,.
\end{aligned}$$

Thus, the next lemma implies $f(b) \leq 0$. □

Lemma 35. Let $\alpha \in [\frac{1}{2}, 1]$, $x, y \geq 0$. Then

$$(x+y)^{2\alpha} - (x^{\alpha} - y^{\alpha})^2 \leq (4xy)^{\alpha}\,.$$

We need two further lemmas before we prove this inequality.

Lemma 36. For $s \in [0, \frac{1}{2}]$ it holds

$$\frac{1-s}{s} \leq \frac{\log(s)}{\log(1-s)}\,.$$

Proof. Define

$$f(s) := s\log(s) - (1-s)\log(1-s)\,.$$

It hold

$$f''(s) = \frac{1}{s} - \frac{1}{1-s}\,.$$

Thus, $f''(s) \geq 0$ for $s \leq \frac{1}{2}$. It holds $f(0) = f(\frac{1}{2}) = 0$. Thus, $f(s) \leq 0$. Thus,

$$s\log(s) \leq (1-s)\log(1-s)\,.$$

Because of $\log(1-s) \leq 0$, thus implies

$$\frac{1-s}{s} \leq \frac{\log(s)}{\log(1-s)}\,.$$

□

Lemma 37. Let $a, b \geq 0$, $x \in [1,2]$. Define

$$f(x) := \frac{a^x - b^x}{(a+b)^x}\,.$$

Assume $a \geq b$. Then $f''(x) \leq 0$. In particular,

$$\inf_{x\in[1,2]} f(x) = f(1) = f(2) = \frac{a^2-b^2}{(a+b)^2} = \frac{a-b}{a+b}\,.$$

Proof. It holds

$$f''(x) = (a+b)^{-x}\left(a^x \log\left(\frac{a}{a+b}\right)^2 - b^x \log\left(\frac{b}{a+b}\right)^2\right)\,.$$

Set $s = \frac{b}{a+b}$. Then $1-s = \frac{a}{a+b}$. Then Lemma 36 implies

$$\frac{a}{b} \leq \frac{\log\left(\frac{b}{a+b}\right)}{\log\left(\frac{a}{a+b}\right)}\,.$$

Thus,

$$\left(\frac{a}{b}\right)^x \leq \left(\frac{a}{b}\right)^2 \leq \left(\frac{\log\left(\frac{b}{a+b}\right)}{\log\left(\frac{a}{a+b}\right)}\right)^2\,.$$

Thus,

$$a^x \log\left(\frac{a}{a+b}\right)^2 \leq b^x \log\left(\frac{b}{a+b}\right)^2\,.$$

Thus, $f''(x) \leq 0$. □

Proof of Lemma 35. For $z \geq 1$ define

$$f(z) := \left(z + 2 + z^{-1}\right)^\alpha - z^\alpha - z^{-\alpha}\,.$$

We will show that $f(z) \leq 4^\alpha - 2$. This implies

$$\frac{(z+1)^{2\alpha} - z^{2\alpha} - 1}{z^\alpha} \leq 4^\alpha - 2\,.$$

Thus,

$$(z+1)^{2\alpha} \leq (4^\alpha - 2)\, z^\alpha + z^{2\alpha} + 1\,.$$

By setting $z = \frac{x}{y}$ for $x \geq y$, we obtain

$$(x+y)^{2\alpha} - (x^\alpha - y^\alpha)^2 \leq (4xy)^\alpha\,.$$

The condition $x \geq y$ can be dropped because of symmetry. It remains to show that $f(z) \leq 4^\alpha - 2$ is indeed true. It holds $f(1) = 4^\alpha - 2$. To finish the proof, we will show $f'(z) \leq 0$. Define

$$g(z) := (z^2 - 1)(z+2)^{2\alpha} - (z+1)^2(z^{2\alpha} - 1)\,.$$

Then

$$f'(z)\frac{z^{\alpha+2}\left(z + 2 + z^{-1}\right)}{\alpha} = g(z)\,.$$

We show $g(z) \leq 0$, and therefore $f'(z) \leq 0$, by applying Lemma 37 with $a = z$, $b = 1$, and $x = 2\alpha$:

$$\frac{z^x - 1^x}{(z+1)^x} \geq \frac{z^2 - 1^2}{(z+1)^2}\,,$$

which implies

$$\left(z^{2\alpha} - 1\right)(z+1)^2 \geq \left(z^2 - 1\right)(z+1)^{2\alpha}\,.$$

□

According to Remark 2, we have now finally finished to proof of Lemma 13 and therefore of Theorem 3.

Index of Notation

D, 11
F, 7
F_n, 7
$N(Q, d, r)$, 8
R_n, 19
Y', 8
Y_i, 7
Δ_n, 28
$\Sigma\mathcal{Y}$, 7
α, 8
$\mathrm{B}_0(m, \mathfrak{l})$, 10
$\mathrm{B}_r(o, d)$, 7
β, 8, 11, 19
$\mathsf{diam}(A, \mathfrak{b})$, 11
$\dot{\mathcal{Q}}$, 10
$\mathsf{entr}_n(Q, \mathfrak{b})$, 11
$\eta_{\beta,n}$, 9
γ, 7
$\gamma_2(\mathcal{Q}, \mathfrak{b})$, 11
κ, 10
$\mathcal{Q}$, 7
$\mathcal{Q}_n$, 19

$\mathcal{Y}$, 7
$\mathfrak{M}(\zeta)$, 8, 10
$\mathfrak{a}$, 7, 10
$\mathfrak{b}$, 7
$\mathfrak{b}_m$, 10
$\mathfrak{c}$, 7
$\mathfrak{l}$, 7
$\overline{qp}$, 8
${}^{\mathfrak{c}}\overline{yq}$, 7
ξ, 10
ζ, 8
$c_{\mathtt{e}}$, 8
$c_{\mathtt{g}}$, 7
m, 7
$m_n^{Q_n}$, 19
m_n, 7
o, 19

Index of Assumptions

`Closeness`, 31
`Entropy`, 8
`Existence'`, 18
`Existence`, 7
`Growth`, 7
`Moment`, 8
`Small Entropy`, 19
`Strong Entropy`, 10
`Strong Moment`, 10
`Strong Quadruple`, 10
`Weak Quadruple`, 7

References

[AGP18] Adil Ahidar-Coutrix, Thibaut Le Gouic, and Quentin Paris. "Convergence rates for empirical barycenters in metric spaces: curvature, convexity and extendible geodesics". In: (2018). arXiv: `1806.02740 [math.ST]`.

[Bač14a] Miroslav Bačák. "Computing medians and means in Hadamard spaces". In: *SIAM J. Optim.* 24.3 (2014), pp. 1542–1566. ISSN: 1052-6234. DOI: `10.1137/140953393`. URL: `https://doi.org/10.1137/140953393`.

[Bač14b] Miroslav Bačák. *Convex analysis and optimization in Hadamard spaces.* Vol. 22. De Gruyter Series in Nonlinear Analysis and Applications. De Gruyter, Berlin, 2014, pp. viii+185. ISBN: 978-3-11-036103-2; 978-3-11-036162-9. DOI: `10.1515/9783110361629`. URL: `https://doi.org/10.1515/9783110361629`.

[Bac18] Miroslav Bacak. *Old and new challenges in Hadamard spaces.* 2018. arXiv: `1807.01355 [math.FA]`.

[BDG07] Eustasio del Barrio, Paul Deheuvels, and Sara van de Geer. *Lectures on empirical processes.* EMS Series of Lectures in Mathematics. Theory and statistical applications, With a preface by Juan A. Cuesta Albertos and Carlos Matrán. European Mathematical Society (EMS), Zürich, 2007, pp. x+254. ISBN: 978-3-03719-027-2. DOI: `10.4171/027`. URL: `https://doi.org/10.4171/027`.

[BFW17] Dirk Banholzer, Jörg Fliege, and Ralf Werner. "On almost sure rates of convergence for sample average approximations". In: (2017). http://www.optimization-online.org/DB_HTML/2017/01/5834.html.

[BGW05] Arindam Banerjee, Xin Guo, and Hui Wang. "On the optimality of conditional expectation as a Bregman predictor". In: *IEEE Trans. Inform. Theory* 51.7 (2005), pp. 2664–2669. ISSN: 0018-9448. DOI: 10.1109/TIT.2005.850145. URL: https://doi.org/10.1109/TIT.2005.850145.

[BHV01] Louis J. Billera, Susan P. Holmes, and Karen Vogtmann. "Geometry of the space of phylogenetic trees". In: *Adv. in Appl. Math.* 27.4 (2001), pp. 733–767. ISSN: 0196-8858. DOI: 10.1006/aama.2001.0759. URL: https://doi.org/10.1006/aama.2001.0759.

[BL14] Witold Bednorz and Rafał Latała. "On the boundedness of Bernoulli processes". In: *Ann. of Math. (2)* 180.3 (2014), pp. 1167–1203. ISSN: 0003-486X. DOI: 10.4007/annals.2014.180.3.8. URL: https://doi.org/10.4007/annals.2014.180.3.8.

[BLO18] D. Barden, H. Le, and M. Owen. "Limiting behaviour of Fréchet means in the space of phylogenetic trees". In: *Ann. Inst. Statist. Math.* 70.1 (2018), pp. 99–129. ISSN: 0020-3157. DOI: 10.1007/s10463-016-0582-9. URL: https://doi.org/10.1007/s10463-016-0582-9.

[BN08] I. D. Berg and I. G. Nikolaev. "Quasilinearization and curvature of Aleksandrov spaces". In: *Geom. Dedicata* 133 (2008), pp. 195–218. ISSN: 0046-5755. DOI: 10.1007/s10711-008-9243-3. URL: https://doi.org/10.1007/s10711-008-9243-3.

[Bou02] Olivier Bousquet. "A Bennett concentration inequality and its application to suprema of empirical processes". In: *C. R. Math. Acad. Sci. Paris* 334.6 (2002), pp. 495–500. ISSN: 1631-073X. DOI: 10.1016/S1631-073X(02)02292-6. URL: https://doi.org/10.1016/S1631-073X(02)02292-6.

[BP03] Rabi Bhattacharya and Vic Patrangenaru. "Large sample theory of intrinsic and extrinsic sample means on manifolds. I". In: *Ann. Statist.* 31.1 (2003), pp. 1–29. ISSN: 0090-5364. DOI: 10.1214/aos/1046294456. URL: https://doi.org/10.1214/aos/1046294456.

[CCM97] Hyeong In Choi, Sung Woo Choi, and Hwan Pyo Moon. "Mathematical theory of medial axis transform". In: *Pacific J. Math.* 181.1 (1997), pp. 57–88. ISSN: 0030-8730. DOI: 10.2140/pjm.1997.181.57. URL: https://doi.org/10.2140/pjm.1997.181.57.

[DD16] Michel Marie Deza and Elena Deza. *Encyclopedia of distances*. Fourth. Springer, Berlin, 2016, pp. xxii+756. ISBN: 978-3-662-52843-3; 978-3-662-52844-0. DOI: 10.1007/978-3-662-52844-0. URL: https://doi.org/10.1007/978-3-662-52844-0.

[DM17] Paromita Dubey and Hans-Georg Müller. "Fréchet Analysis of Variance for Random Objects". In: (2017). arXiv: 1710.02761 [math.ST].

[Fed59] Herbert Federer. "Curvature measures". In: *Trans. Amer. Math. Soc.* 93 (1959), pp. 418–491. ISSN: 0002-9947. DOI: 10.2307/1993504. URL: https://doi.org/10.2307/1993504.

[Fré48] Maurice Fréchet. "Les éléments aléatoires de nature quelconque dans un espace distancié". In: *Ann. Inst. H. Poincaré* 10 (1948), pp. 215–310. URL: http://www.numdam.org/item?id=AIHP_1948__10_4_215_0.

[Gee00] Sara A. van de Geer. *Applications of empirical process theory.* Vol. 6. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2000, pp. xii+286. ISBN: 0-521-65002-X.

[Huc11] Stephan F. Huckemann. "Intrinsic inference on the mean geodesic of planar shapes and tree discrimination by leaf growth". In: *Ann. Statist.* 39.2 (2011), pp. 1098–1124. ISSN: 0090-5364. DOI: 10.1214/10-AOS862. URL: https://doi.org/10.1214/10-AOS862.

[Nye11] Tom M. W. Nye. "Principal components analysis in the space of phylogenetic trees". In: *Ann. Statist.* 39.5 (2011), pp. 2716–2739. ISSN: 0090-5364. DOI: 10.1214/11-AOS915. URL: https://doi.org/10.1214/11-AOS915.

[PFA06] Xavier Pennec, Pierre Fillard, and Nicholas Ayache. "A Riemannian Framework for Tensor Computing". In: *International Journal of Computer Vision* 66.1 (Jan. 2006). doi: 10.1007/s11263-005-3222-z, pp. 41–66. ISSN: 1573-1405. DOI: 10.1007/s11263-005-3222-z.

[PM19] Alexander Petersen and Hans-Georg Müller. "Fréchet regression for random objects with Euclidean predictors". In: *Ann. Statist.* 47.2 (2019), pp. 691–719. ISSN: 0090-5364. DOI: 10.1214/17-AOS1624. URL: https://doi.org/10.1214/17-AOS1624.

[Pol90] David Pollard. *Empirical processes: theory and applications.* Vol. 2. NSF-CBMS Regional Conference Series in Probability and Statistics. Institute of Mathematical Statistics, Hayward, CA; American Statistical Association, Alexandria, VA, 1990, pp. viii+86. ISBN: 0-940600-16-1.

[Stu02] Karl-Theodor Sturm. "Nonlinear martingale theory for processes with values in metric spaces of nonpositive curvature". In: *Ann. Probab.* 30.3 (2002), pp. 1195–1222. ISSN: 0091-1798. DOI: 10.1214/aop/1029867125. URL: https://doi.org/10.1214/aop/1029867125.

[Stu03] Karl-Theodor Sturm. "Probability measures on metric spaces of nonpositive curvature". In: *Heat kernels and analysis on manifolds, graphs, and metric spaces (Paris, 2002).* Vol. 338. Contemp. Math. Amer. Math. Soc., Providence, RI, 2003, pp. 357–390. DOI: 10.1090/conm/338/06080. URL: https://doi.org/10.1090/conm/338/06080.

[Tal14] Michel Talagrand. *Upper and lower bounds for stochastic processes*. Vol. 60. Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge. A Series of Modern Surveys in Mathematics [Results in Mathematics and Related Areas. 3rd Series. A Series of Modern Surveys in Mathematics]. Modern methods and classical problems. Springer, Heidelberg, 2014, pp. xvi+626. ISBN: 978-3-642-54074-5; 978-3-642-54075-2. DOI: 10.1007/978-3-642-54075-2. URL: https://doi.org/10.1007/978-3-642-54075-2.

[VW96] Aad W. van der Vaart and Jon A. Wellner. *Weak convergence and empirical processes*. Springer Series in Statistics. With applications to statistics. Springer-Verlag, New York, 1996, pp. xvi+508. ISBN: 0-387-94640-3. DOI: 10.1007/978-1-4757-2545-2. URL: https://doi.org/10.1007/978-1-4757-2545-2.

[Zie77] Herbert Ziezold. "On expected figures and a strong law of large numbers for random elements in quasi-metric spaces". In: *Transactions of the Seventh Prague Conference on Information Theory, Statistical Decision Functions, Random Processes and of the Eighth European Meeting of Statisticians (Tech. Univ. Prague, Prague, 1974), Vol. A*. 1977, pp. 591–602.