Austrian Journal of Statistics
MMMMMM YYYY, Volume VV, Issue II, 1–xx.
http://www.ajs.or.at/

Extended Order-p Means and Modes of Continuous Densities

Martin Welk
UMIT TIROL – Private University of Health Sciences and Health Technology

Abstract

The order-p mean of a finite multiset of real numbers is the minimiser of the sum of p-th powers of the distances to all numbers in the multiset. From Gauß' and Legendre's work on least squares estimation and the arithmetic mean ($p = 2$), this concept has developed in progressing generality. First, it was established for positive integer p, with $p = 1$ yielding the median as proven by Glaisher and Fechner. The further generalisation to arbitrary positive real p was considered by Jackson, Barral Souto and Fréchet who also established that the limits $p \to 0$ and $p \to \infty$ yield the mode and mid-range value of the data, respectively. Replacing sums by integrals, order-p means can be defined for continuous densities of real numbers; Fréchet contributed substantially to the study of this concept. Whereas the inclusion of the arithmetic mean and median in the scale of order-p means as well as the mid-range limit transfer straightforward to the continuous case, this is not the case for the mode limit, as was pointed out by Fréchet. Inspired by applications in signal and image processing, we present in this work an extension of order-p means of continuous densities to negative exponents between -1 and 0. We show that for densities with a unique global maximum (mode) the order-p means approach the mode for $p \to -1$. We demonstrate the relation between the extended order-p means and modes by numerical examples.

Keywords: order-p means, mode.

1. Introduction

Given a positive real parameter p, the order-p mean of a finite multiset $\mathcal{X}$ of real numbers is defined as the minimiser of the sum of p-th powers of distances to all numbers in the multiset,

$$\mu_p(\mathcal{X}) = \operatorname*{argmin}_{\mu\in\mathbb{R}} \sum_{x\in\mathcal{X}} |\mu - x|^p \ . \tag{1}$$

The concept can easily be generalised to weighted means and to continuous distributions described by densities. Its history goes back to works by Laplace (1774) (with $p = 1$), Legendre (1805) and Gauss (1809) (with least-squares minimisation, $p = 2$). Due to the important contributions of Fréchet (1946, 1948a,b,c) to their theory, order-p means are sometimes also termed Fréchet means. However, as this term is sometimes used in a wider or narrower sense, we stick with the name order-p means. We will give a few more historic facts in Section 2.3. For a significantly more detailed account we recommend a paper by Armatte (2006) which was also of great help to the author of the present work in tracking down historical facts and locating references.

As noted already by Legendre and Gauß, the case $p = 2$ yields the arithmetic mean of $\mathcal{X}$; Glaisher (1872) and Fechner (1878) showed that for $p = 1$ the median is obtained. In a report by Barral Souto (1938) it was proven that for any given $\mathcal{X}$, the limit of order-p means for $p \to 0$ is the mode of $\mathcal{X}$, i.e. the most frequent value (provided this is unique). Jackson (1921) found that the limit for $p \to \infty$ yields the mid-range, i.e. the arithmetic mean of the smallest and largest number in $\mathcal{X}$. By definition, order-p means are shift-equivariant and scale-equivariant, i.e. a shift or scaling applied to the input data before taking the mean is equivalent to applying the shift or scale after taking the mean.

As to nomenclature, order-p means are not to be confused with the other famous parametrised family of means, viz. the power means of order $p \in \mathbb{R}$ which are defined only for positive (or if $p > 0$, nonnegative) numbers, and which are only scale-equivariant but not shift-invariant. However, both families of means stand in an obvious relation as the order-p mean can also be stated, equivalent to (1), as minimiser of the power-p mean of distances from μ to the given numbers.

For continuous distributions, the special cases $p = 1$ and $p = 2$ carry over in an obvious way, as does the limit for $p \to \infty$ (with an "essential" mid-range value obtained from the supremum and infimum of the support of the density). More difficult is the case $p \to 0$: A short remark in Jordan (1927) without proof suggests that this again leads to the mode of a distribution. This intuition was proven wrong by Fréchet (1948b,c). Fréchet's analysis brought out, instead, that order-p means of a continuous distribution tend for $p \to 0$ to a limit μ_0 which minimises the geometric mean of distances from the given values. This is in perfect agreement with the well-known facts about power means where the geometric mean fits in in place of the otherwise undefined case $p = 0$.

Applications in data analysis. Order-p means of univariate data continue to be used in data analysis, see e.g. Livadiotis (2017) who is concerned with the case $p \geq 1$ and derives approximation laws.

Many authors consider the generalisation to data on Riemannian manifolds, see e.g. Arnaudon and Miclo (2014); Afsari (2011); Cazals, Delmas, and O'Donnell (2020), with applications in clustering, see Cabanes, Barbaresco, Arnaudon, and Bigot (2019); Cazals, Delmas, and O'Donnell (2021), or geodesic principal component analysis, see Pennec (2020). Whereas often the case $p \geq 1$ receives the main interest, also $p \in (0, 1)$ is explicitly included in the investigations e.g. of Cabanes *et al.* (2019). An adaptation of the Weiszfeld algorithm for multivariate L^1 medians Weiszfeld (1937) to Riemannian order-p means with $1 \leq p < 2$ is considered in Aftab, Hartley, and Trumpf (2015).

Application in signal and image processing. Starting with the introduction of the median filter in signal processing by Tukey (1971), estimators like order-p means have also been used to construct a class of local filters for signals and images. Based on the nomenclature from robust statistics, see Huber (1981), where order-p means belong to the class of M-estimators, such filters are often called M-smoothers. M-smoothers based on order-p means were introduced by Torroba, Cap, Rabal, and Furlan (1994). In this context, the relation between continuous distributions (assumed as underlying measured signals and images) and discrete data (which are actually processed) again becomes relevant.

In image processing, it is interesting to relate local filters based on discrete procedures to such that are derived from partial differential equations (PDEs). For example, the median filter asymptotically approximates, via a reformulation of the median filtering procedure for images over continuous domains, the curvature motion PDE, as shown by Guichard and Morel (1997). Extensions of this result to M-smoothers based on other order-p means were investigated first in Griffin (2000); Griffin and Lillholm (2003). Taking up this line of research recently, we could establish PDE limits for all $p > 0$ and the mode filter of images modelled by functions of sufficient regularity over the continuous domain (Welk and Weickert 2019, 2021). In agreement with Fréchet's findings, the limit $p \to 0$ does not yield the PDE of the mode filter.

Extended class of order-p means. Surprisingly, setting formally $p = -1$ in the M-smoother PDE, however, yields the same PDE as the mode filter. This triggered us to consider whether order-p means for $p < 0$ could make sense, such as to allow a limit transition $p \to -1$. Indeed, in the continuous case (but not for discrete data) such an extension is possible for $-1 < p < 0$. Due to the decreasing (instead of increasing) behaviour of the penalisers $|\mu - x|^p$ for negative p, the objective function needs to be modified by a factor $\operatorname{sgn}(p)$ to allow minimisation. We could prove (Welk and Weickert 2019) that the so defined M-smoothers asymptotically approximate PDEs perfectly matching the PDE family for positive p, and indeed yield for $p \to -1$ the PDE approximated by mode filtering as limit.

It appears intriguing to try to generalise order-p means even further to exponents $p \leq -1$. Indeed, the image filtering PDE from (Welk and Weickert 2019) could formally be considered for such p; note that already Gabor (1965) proposes an image filtering equation that corresponds to a time step of that PDE with $p = -2$, see also Lindenbaum, Fischer, and Bruckstein (1994). However, following the lines of existing definitions, we see no meaningful way to such a generalisation of the order-p means.

Goals of this work. The evidence accumulated so far can be seen as a hint that the mode could fit into an extended family of order-p means of continuous distributions at the parameter value $p = -1$. The PDE limit results from Welk and Weickert (2019, 2021), however, are not sufficient to prove this since the same PDE can be approximated by multiple local filters.

In this work, we undertake to close this gap. We will prove that the extended family of order-p means indeed converges to the mode filter for $p \to -1$, thereby complementing the picture drafted by earlier work on order-p means up to Fréchet.

We point out that the focus of this paper is theoretical. The design of possible data analysis applications based on the extended class of order-p means should be discussed by experts in suitable application fields.

Structure of the paper. In Section 2 we recall the existing results about order-p means and add some historic context. Section 3 we discuss the extension of order-p means to negative parameters $p > -1$ and prove our main result. Section 4 is devoted to demonstrations of the result by a combination of analytical and numerical reasoning for sufficiently accessible example functions, with part of the derivations placed in Appendix A. A summary is presented in Section 5.

2. Concept of order-p means

In this section we present essential facts about order-p means for $p > 0$, thereby filling in somewhat more detail into what has been outlined in the introduction. Throughout this paper, we denote by $\mathbb{R}^+$ and $\mathbb{R}_0^+$ the sets of positive and nonnegative real numbers, respectively.

2.1. The discrete case

The basic definition of order-p means has been given above (1).

Relation to power means. Given a real number $p \neq 0$, one can define the p-th power mean M_p of a finite multiset $\mathcal{X}$ of positive (or, for $p > 0$, even nonnegative) real numbers as

$$M_p(\mathcal{X}) = \left(\frac{1}{\#\mathcal{X}} \sum_{x \in \mathcal{X}} x^p \right)^{1/p} \tag{2}$$

where $\#\mathcal{X}$ is the cardinality of $\mathcal{X}$. This family, too, contains the arithmetic mean as a special case but for $p = 1$. Further important special cases include $p = 2$, the quadratic mean, and $p = -1$, the harmonic mean. For any given $\mathcal{X}$, power means $M_p(\mathcal{X})$ increase monotonically with p, strictly monotonically if $\mathcal{X}$ contains at least two distinct numbers. It is also well-known that for $p \to 0$, $M_p(\mathcal{X})$ approaches the geometric mean, which therefore can be denoted as $M_0(\mathcal{X})$ to fill in the gap in the parameter range, see e.g. (Bullen 2003, p. 175), (Gustin 1950),

$$M_0(\mathcal{X}) = \left(\prod_{x\in\mathcal{X}} x \right)^{1/\#\mathcal{X}} . \tag{3}$$

The two parametrised families of means are related since by an obvious reformulation the order-p mean can be stated as the minimiser of the p-th power mean of distances from μ to the given data,

$$\mu_p(\mathcal{X}) = \operatorname*{argmin}_{\mu\in\mathbb{R}} M_p(|\mu - \mathcal{X}|) , \tag{4}$$

where $|\mu - \mathcal{X}|$ denotes the multiset obtained from $\mathcal{X}$ by replacing each element x with $|\mu - x|$.

Despite superficial similarities, the two families of means address significantly different situations: Power means (with the exception of the arithmetic mean, $p = 1$, and those of odd integer p) are suitable only for positive numbers and are equivariant to scaling (1-homogeneous), i.e. for any $\alpha > 0$ one has

$$M_p(\alpha\mathcal{X}) = \alpha\, M_p(\mathcal{X}) \tag{5}$$

where $\alpha\mathcal{X} := \{\alpha x : x \in \mathcal{X}\}$ denotes the multiset in which each element is scaled with α. Order-p means, in contrast, are defined for data from the entire real line, and are equivariant not only to scaling but also to translations (thus, affine equivariant).

Weighted means. It is straightforward to generalise (1), (2) and (3) into weighted means. We state this exemplarily for the order-p means. Let $\mathcal{X}$ be a set of real numbers equipped with a weight function $w : \mathcal{X} \to \mathbb{R}_0^+$ with $\sum\limits_{x\in\mathcal{X}} w(x) = 1$. We can lift here the restriction that $\mathcal{X}$ needs to be finite. The condition that w sums up to unit total weight entails that the support – the set of values $x \in \mathcal{X}$ with $w(x) > 0$ – is countable.

The weighted order-p mean is then given by

$$\mu_p(\mathcal{X}, w) = \operatorname*{argmin}_{\mu\in\mathbb{R}} \sum_{x\in\mathcal{X}} |\mu - x|^p w(x) . \tag{6}$$

All statements above about special cases, limits, and the relation between order-p means and power means remain valid for weighted means.

2.2. Means in the continuous domain

We turn now to replace discrete (multi-) sets $\mathcal{X}$ of values by continuous distributions on the real axis. More precisely, we consider piecewise continuous densities $\gamma : \mathbb{R} \to \mathbb{R}_0^+$ with $\int\limits_{\mathbb{R}} \gamma(x)\,\mathrm{d}x = 1$. For such a density γ we can easily state an analogon of (1):

$$\mu_p(\gamma) = \operatorname*{argmin}_{\mu\in\mathbb{R}} \int\limits_{\mathbb{R}} |\mu - x|^p \gamma(x)\,\mathrm{d}x . \tag{7}$$

Similarly, for piecewise continuous densities $\gamma : \mathbb{R}^+ \to \mathbb{R}_0^+$ on the positive real axis with $\int\limits_{\mathbb{R}^+} \gamma(x)\,\mathrm{d}x = 1$, the p-th power mean and the geometric mean can be stated as

$$M_p(\gamma) = \left(\int\limits_{\mathbb{R}^+} x^p \gamma(x)\,\mathrm{d}x \right)^{1/p} , \qquad M_0(\gamma) = \exp\left(\int\limits_{\mathbb{R}^+} \ln x \cdot \gamma(x)\,\mathrm{d}x \right) . \tag{8}$$

Again, the order-p mean can be reformulated as minimiser of the p-th power mean of distances: If $\gamma_{\mu,+}$ denotes the density obtained from γ through shifting it by μ and folding the negative and positive parts of the real axis,

$$\gamma_{\mu,+}(x) := \gamma(\mu + x) + \gamma(\mu - x) , \qquad x \geq 0 , \tag{9}$$

we have for $p > 0$ that

$$\mu_p(\gamma) = \operatorname*{argmin}_{\mu \in \mathbb{R}} M_p(\gamma_{\mu,+}) . \tag{10}$$

2.3. A brief historical excursus

The concept of order-p means has developed since around 1800 with gradually increasing generality. With the works of Legendre (1805) and Gauss (1809), least-squares approximation was established early in the 19th century as a method for estimating quantities from noisy data. Both authors noticed that the minimisation of the sum of squared distances to given numbers, or equivalently, minimisation of the quadratic mean of deviations, yields the arithmetic mean. In Legendre's work, the reasoning already points to the concept of the continuous generalisation. In turn, Gauß already mentions the possibility to use other (namely, even positive integer) exponents p instead of 2, justifying the choice $p = 2$ pragmatically (Gauss 1809, p. 221), and also mentions the mid-range limit for $p \to \infty$. However, with the further theory developing around Gauß' normal distribution that leads to least-squares approximation as maximum likelihood estimator, and Moivre-Laplace's theorem that supports the assumption of normal distribution for a range of contexts, discussion focused on least squares and arithmetic mean for the next decades.

Decades before Legendre's and Gauß' work, Boscovich and Maire (1770) had already brought up the minimisation of the sum of absolute deviations, as in order-1 means, for the estimation of linear function coefficients from noisy measurements. Laplace (1774) had linked it to what is now known as the Laplace distribution, and also identified the order-1 mean of three numbers as the middle of these, i.e. their median, but apparently stopped short of generalising this to larger data sets. After the least-squares method was established, Laplace, too, turned to advocating $p = 2$ (Laplace 1820, p. 318pp.).

It was Ellis (1844) who brought back to attention that squares of deviations are not a priori the only choice, and that it needs justification to prefer these over other increasing functions of absolute deviations. Glaisher (1872) then considered the minimisation of the sum of absolute deviations, thus, the order-p mean with $p = 1$, and was able to prove that it coincides with the median of the values for any finite multiset of numbers. A few years later, Fechner (1878) studied the concept of order-p means for integer $p > 0$ and proved again that $p = 1$ yields the median.

Discrete order-p means with non-integer p seem to make their first appearance with Jackson (1921) who was mostly interested in the limit $p \to 1$ as a way of approximating the median. In their own right, order-p means with non-integer p were studied by Barral Souto (1938) who also proved that discrete order-p means for $p \to 0$ approach the mode of a discrete real multiset. Without a proof, this connection was also mentioned before in a short remark in (Jordan 1927, p. 172).

Fréchet studied order-p means and related means in several papers (Fréchet 1946, 1948a,b,c). In (Fréchet 1948b,c), he considers the limit $p \to 0$ and carries out a detailed analysis involving different kinds of distributions of input data as well as different concepts of the limit involved. Besides, he also extends in (Fréchet 1948c) on $p \to \infty$ which results in the mid-range value of the given data. For $p \to 0$, Fréchet distinguishes between (finite) discrete distributions and general distributions which comprise such with continuous densities as considered in our present work as well as mixed distributions in which a continuous density is combined with an enumerable set of values with positive probability. In the discrete case he considers three possible interpretations of the limit $p \to 0$, two of which lead to reasonable limits, and agree in reproducing the mode of the discrete distribution as the relevant limit. He finds that for purely continuous distributions the order-p means for $p \to 0$ approach a value for which the geometric mean of absolute distances to the input data is minimised. In contrast, for mixed distributions, the discrete component dominates for $p \to 0$, effectively making the mode of the discrete component the limit value.

3. Extended family of order-p means

In this section we focus on the extension of continuous order-p means of a continuous distribution to the extended parameter range $p > -1$, $p \neq 0$, and prove our approximation result for the mode value. We emphasise that this generalisation is not applicable to the case of discrete data (finite multisets) or mixed discrete–continuous distributions as the objective functions that are minimised by order-p means with negative p will have poles (diverging to $-\infty$) at any values with positive probability.

3.1. Definition of extended order-p means

The following definition follows Welk and Weickert (2019, 2021).

Definition 1 (Continuous order-p mean). Let $\gamma : \mathbb{R} \to \mathbb{R}_0^+$ be a continuous density, i.e. a piecewise continuous function with $\int_{\mathbb{R}} \gamma(x)\,\mathrm{d}x = 1$.

The order-p mean $\mu_p(\gamma)$ for $p > -1$, $p \neq 0$ is given by

$$\mu_p(\gamma) := \operatorname*{argmin}_{\mu \in \mathbb{R}} E_{p,\gamma}(\mu) \tag{11}$$

where

$$E_{p,\gamma}(\mu) := \operatorname{sgn}(p) \cdot \int_{\mathbb{R}} |\mu - x|^p \gamma(x)\,\mathrm{d}x\ . \tag{12}$$

Since for $p < 1$ the objective function (12) can be non-convex, it is possible that the global minimum is attained at multiple locations. Densities for which this happens are, however, non-generic, i.e. turn by small perturbations into such with unique global minimisers. We do therefore not consider these bifurcation cases in detail.

Remark 1. The reformulation (10) of order-p means via power means transfers verbatim to $-1 < p < 0$. Note that the negative exponent $1/p$ in the definition of M_p has the same effect of inverting the monotonicity of the penalisers $|\mu - x|^p$ as the sign factor in (12).

Moreover, by the inclusion of the geometric mean for $p = 0$ in (8), the definition of order-p means via (10) can also be used to close the gap in Def. 1, thus leading to a concept of order-p means for all real $p > -1$. Note that the insertion of this case is in accordance with Fréchet's limit result (Fréchet 1948b,c).

3.2. Mode limit

Lemma 1. *Let γ be as in Definition 1. Let $\delta, \varepsilon > 0$ be given. Let $m, \mu \in \mathbb{R}$ be given such that for all $z, \zeta \in \mathbb{R}$ with $|z - m| \le \varepsilon$, $|\zeta - \mu| \le \varepsilon$ one has*

$$\gamma(z) > \gamma(\zeta) + \delta\ . \tag{13}$$

Then there exists some $\eta > 0$ such that for all p, $-1 < p < -1 + \eta$ one has $E_{p,\gamma}(m) < E_{p,\gamma}(\mu)$.

Proof. Consider p with $-1 < p < 0$. By definition

$$E_{p,\gamma}(\mu) - E_{p,\gamma}(m) = -\int_{-\infty}^{\infty} |\mu - z|^p \gamma(z)\,\mathrm{d}z + \int_{-\infty}^{\infty} |m - z|^p \gamma(z)\,\mathrm{d}z$$
$$= I_1 + I_2 + I_3 + I_4 + I_5 + I_6 + I_7 + I_8 \;, \tag{14}$$

$$I_1 = -\int_{-\infty}^{\mu-\varepsilon} (\mu - z)^p \gamma(z)\,\mathrm{d}z \;, \qquad I_2 = -\int_{\mu-\varepsilon}^{\mu} (\mu - z)^p \gamma(z)\,\mathrm{d}z \;, \tag{15}$$

$$I_3 = -\int_{\mu}^{\mu+\varepsilon} (z - \mu)^p \gamma(z)\,\mathrm{d}z \;, \qquad I_4 = -\int_{\mu+\varepsilon}^{\infty} (z - \mu)^p \gamma(z)\,\mathrm{d}z \;, \tag{16}$$

$$I_5 = \int_{-\infty}^{m-\varepsilon} (m - z)^p \gamma(z)\,\mathrm{d}z \;, \qquad I_6 = \int_{m-\varepsilon}^{m} (m - z)^p \gamma(z)\,\mathrm{d}z \;, \tag{17}$$

$$I_7 = \int_{m}^{m+\varepsilon} (z - m)^p \gamma(z)\,\mathrm{d}z \;, \qquad I_8 = \int_{m+\varepsilon}^{\infty} (z - m)^p \gamma(z)\,\mathrm{d}z \;. \tag{18}$$

By the decreasing monotonicity of $|z|^p$, and non-negativity of γ we have

$$I_1 + I_4 \ge -\varepsilon^p \left(\int_{-\infty}^{\mu-\varepsilon} + \int_{\mu+\varepsilon}^{\infty} \right) \gamma(z)\,\mathrm{d}z \ge -\varepsilon^p \int_{-\infty}^{\infty} \gamma(z)\,\mathrm{d}z = -\varepsilon^p \;. \tag{19}$$

Nonnegativity of the integrands in I_5, I_8 yields

$$I_5 + I_8 \ge 0 \;. \tag{20}$$

By the hypothesis (13) we have

$$I_2 + I_6 = \int_{-\varepsilon}^{0} (-z)^p \big(\gamma(m + z) - \gamma(\mu + z)\big)\,\mathrm{d}z \ge \delta \int_{-\varepsilon}^{0} (-z)^p\,\mathrm{d}z = \frac{\delta}{p+1}\varepsilon^{p+1} \;, \tag{21}$$

$$I_3 + I_7 = \int_{0}^{\varepsilon} z^p \big(\gamma(m + z) - \gamma(\mu + z)\big)\,\mathrm{d}z \ge \delta \int_{0}^{\varepsilon} z^p\,\mathrm{d}z = \frac{\delta}{p+1}\varepsilon^{p+1} \;. \tag{22}$$

Combining all estimates yields

$$E_{p,\gamma}(\mu) - E_{p,\gamma}(m) \ge \varepsilon^p \left(\frac{2\varepsilon\delta}{p+1} - 1 \right) \;. \tag{23}$$

Clearly, for any given $\delta, \varepsilon > 0$ the expression on the right-hand side of the last inequality becomes positive for p sufficiently close to -1, which concludes the proof of the lemma. □

Proposition 2. *Let γ be as in Definition 1, and let γ possess a unique mode, i.e. a local maximum $m \in \mathbb{R}$ such that for any other local maximum m' of γ one has $\gamma(m') < \gamma(m) - \tau$ with some global constant τ.*

Then $\mu_p(\gamma)$ approaches m for $p \to -1$ in the following sense: For each $\varepsilon > 0$ there exists some $\eta > 0$ such that for all p, $-1 < p < -1 + \eta$ one has $|\mu_p(\gamma) - m| \le \varepsilon$.

Proof. Given ε as assumed in the proposition, let

$$\tilde{\delta} := \max_{z \in [m-\varepsilon, \varepsilon]} \big(\gamma(m) - \gamma(z)\big) \;. \tag{24}$$

If $\tilde{\delta} > \delta$, reduce ε until $\tilde{\delta} \le \delta$. Let further

$$\delta^* := \min_{z\in\mathbb{R}\setminus[m-\varepsilon,m+\varepsilon]} \big(\gamma(m) - \gamma(z)\big)\ . \tag{25}$$

Certainly one has $0 < \delta^* \le \tilde{\delta} \le \delta$. Since there is no local maximum of γ outside m with a value closer than δ to $\gamma(m)$, one can find a positive $\varepsilon^* < \varepsilon/2$ such that

$$\max_{z\in[m-\varepsilon^*,m+\varepsilon^*]} \big(\gamma(m) - \gamma(z)\big) < \tfrac{1}{3}\delta^* \quad \text{and} \tag{26}$$

$$\min_{z\in\mathbb{R}\setminus[m-\varepsilon+\varepsilon^*,m+\varepsilon-\varepsilon^*]} \big(\gamma(m) - \gamma(z)\big) > \tfrac{2}{3}\delta^*\ . \tag{27}$$

Using the values ε^* and $\frac{1}{3}\delta^*$ for ε and δ in its hypothesis, Lemma 1 yields the desired result. □

3.3. Complementary remarks

Heuristically, the mode limit of order-p means for $p \to -1$ can be understood also on the level of penalisers. If one rescales the penaliser from (12) as $-\frac{1}{p+1}|\mu - x|^p$, one can establish that it weakly converges for $p \to -1$ to the negative Dirac impulse $-\delta(\mu - x)$. Thus, the minimisation of its integral analogous to (12) becomes the minimisation of $-\gamma(\mu)$, leading by definition to the mode of γ.

This is similar to the situation in the discrete setting where the penalisers $|\mu - x|^p$ for $p \to +0$ converge to a function $s(\mu - x)$ with $s(0) = 0$ and $s(z) = 1$ for $z \neq 0$, see Barral Souto (1938), turning the objective function into a jump function that counts for each μ the number of data values from the finite multiset that are different from μ, and is again minimised by the mode.

In the context of signal and image processing where the discrete data being processed are in fact samples from continuous-scale distributions, it is important to understand why the simple transfer from the discrete to the continuous setting that works well for $p \ge 1$ fails for the mode limit. For this aspect we refer to (Welk and Weickert 2021, Sect. 5) where it has been discussed that a proper model of the sampling of signal and image data indeed links the discrete limit $p \to +0$ to a finite-volume discretisation of the continuous-scale limit $p \to -1$.

Multivariate data. The concept of order-p means can in principle be adapted to data from the Euclidean space $\mathbb{R}^d$ with $d > 1$ by using the penaliser $|\boldsymbol{\mu} - \boldsymbol{x}|^p$ where $|\boldsymbol{\mu} - \boldsymbol{x}|$ denotes the Euclidean distance of $\boldsymbol{\mu}, \boldsymbol{x} \in \mathbb{R}^d$.

It is straightforward that $p = 2$ again leads to the arithmetic mean or centre of gravity of the input data whereas $p = 1$ corresponds to the so-called L^1 median (Gini and Galvani 1929; Weber 1909), both in the discrete and continuous case. Also the discrete mode approximation result for $p \to +0$ transfers in a straightforward way. For continuous densities over $\mathbb{R}^d$, the order-p mean can be defined for $p > -d$, and the penalisers weakly converge for $p \to -d$ to a negative Dirac impulse. It can therefore be expected that d-dimensional order-p means approximate the mode of a continuous density in this case.

It should, however, be noted that for data from $\mathbb{R}^d$ already the L^1 median is only one of several multidimensional generalisations of the median, see Small (1990); Welk (2019) and further references therein. As pointed out in these sources, it is controversial whether, and for which applications, the L^1 median is actually a suitable choice, mainly because of its reliance on a Euclidean structure on $\mathbb{R}^d$ which is not always justified in applications. This also puts into question how meaningful the straightforward generalisation of order-p means is for multivariate data from $\mathbb{R}^d$. An important drawback of the L^1 median is its lack of equivariance with regard to affine transformations of the data space. Regarding order-p means, it is easy to check that whereas in the one-dimensional case they are equivariant under affine transformations for any p, this is true in $d > 1$ dimensions only for $p = 2$ and the presumable mode limit case $p \to -d$.

Widening the view further to d-dimensional differentiable manifolds, the generalisation of order-p means as minimisers of a sum of p-th powers of metric distances clearly makes sense for Riemannian manifolds because of their locally Euclidean structure, which has indeed been exploited in data analysis literature, see Cabanes *et al.* (2019) and the other references mentioned already in the introduction. In this setting, also the extension to $-d < p < 0$ with a mode-type limit for $p \to -d$ could be worth considering.

As a thorough discussion of these questions goes beyond the scope of the present work, we keep the focus here on the one-dimensional case and defer further investigation of multivariate order-p means to future work.

4. Demonstrations of the theoretical results

This section aims at illustrating the theoretical result obtained in the last section on simple examples by a combination of analytical and numerical means. First we will consider a particularly simple example for which the limit property can be demomstrated by analytical means. After this, we will consider a slightly richer set of examples based on B-spline basis functions for which the objective functions can be derived in a purely analytical manner, and evaluated by numerical calculations with high precision.

4.1. Analysis of a toy example

Modified from (Welk and Weickert 2021) we consider the following (non-normalised) density function:

$$\gamma(x) = \begin{cases} 1 - \lambda x^2 \;, & x \in [-1+r, 1+r] \;, \\ 0 & \text{else} \end{cases} \tag{28}$$

where λ and r are parameters with $0 < r \ll 1$, $0 < \lambda < (1+r)^{-1}$. The mode of this density is obviously 0.

The order-2 mean of γ is the arithmetic mean of x weighted with γ, i.e.

$$\mu_2(\gamma) = \frac{\int\limits_{-1+r}^{1+r} (1-\lambda x^2) x \,\mathrm{d}x}{\int\limits_{-1+r}^{1+r} (1-\lambda x^2) \,\mathrm{d}x} = \frac{\left[\frac{x^2}{2} - \lambda \frac{x^4}{4}\right]_{-1+r}^{1+r}}{\left[x - \lambda \frac{x^3}{3}\right]_{-1+r}^{1+r}} = \frac{(1-\lambda)r - \lambda r^3}{1 - \lambda/3 - \lambda r^2} \;. \tag{29}$$

Since $r \ll 1$, we have by neglecting $\mathrm{O}(r^3)$ contributions

$$\mu_2(\gamma) \approx \frac{1-\lambda}{1-\lambda/3} r \;. \tag{30}$$

The order-1 mean of γ is the median of x weighted with γ, which can be described as the bisector of the integral over γ. We have therefore $\mu_1(\gamma) = \mu$ where

$$\int\limits_{-1+r}^{\mu} 1 - \lambda x^2 \,\mathrm{d}x = \int\limits_{\mu}^{1+r} 1 - \lambda x^2 \,\mathrm{d}x \tag{31}$$

which after solving the integrals and sorting terms becomes

$$\mu - \frac{\lambda}{3}\mu^3 = (1-\lambda) r - \frac{\lambda}{3} r^3 \;. \tag{32}$$

Due to $r \ll 1$, we have by neglecting the $\mathrm{O}(r^3)$ term

$$\mu_1(\gamma) \approx (1-\lambda) r \;. \tag{33}$$

For general $p > -1$, $p \neq 0$, the order-p mean of γ can be approximately obtained by adapting the derivation from (Welk and Weickert 2021) as

$$\mu_p(\gamma) = \frac{1-\lambda}{1-\frac{p-1}{p+1}\lambda} r + \mathrm{O}(r^3) \tag{34}$$

which obviously matches the previous findings for μ_2 and μ_1.

For $p \to -1$, one sees that $\mu_p(\gamma)$ approaches 0 (up to $\mathrm{O}(r^3)$) in agreement with the expectation that this limit should yield the mode of γ.

For $p \to 0$, one has (for any fixed λ and r)

$$\lim_{p\to 0} \mu_p(\gamma) = \frac{1-\lambda}{1+\lambda} r + \mathrm{O}(r^3)\ . \tag{35}$$

According to the finding from Fréchet (1948c) the order-0 mean of γ is the minimiser μ of the geometric mean of $|\mu - x|$ weighted with γ, i.e.

$$\mu_0(\gamma) = \operatorname*{argmin}_{\mu} \int_{\mathbb{R}} \gamma(x) \ln|\mu - x| \, \mathrm{d}x\ . \tag{36}$$

Due to differentiability of the involved functions we can determine the minimiser μ by

$$\begin{aligned} 0 &= \frac{\mathrm{d}}{\mathrm{d}\mu} \left(\int_{-1+r}^{\mu} (1-\lambda x^2) \ln(\mu - x) \, \mathrm{d}x + \int_{\mu}^{1+r} (1-\lambda x^2) \ln(x-\mu) \, \mathrm{d}x \right) \\ &= 2\lambda(r+\mu) + (1-\lambda\mu^2)\big(\ln(1-r+\mu) - \ln(1+r-\mu)\big)\ . \end{aligned} \tag{37}$$

For details of the derivation of this condition see Appendix A.1.

For $r \ll 1$, one has $\mu \sim r$ which allows to use the Taylor expansion of ln at 1 to derive

$$\mu_0(\gamma) \approx \frac{1-\lambda}{1+\lambda} r \tag{38}$$

in agreement with the expression obtained above for $\lim_{p\to 0} \mu_p(\gamma)$.

4.2. Numerical test of the mode limit

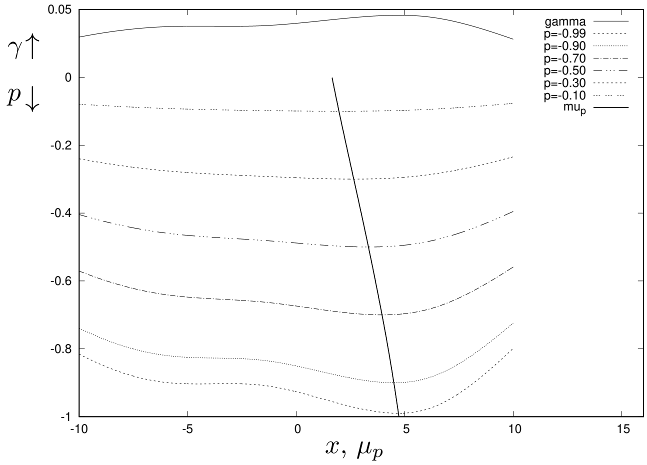
To construct numerical examples we use the cubic B-spline basis function

$$\begin{aligned} B_3(x) = {} & \frac{1}{6}(2+x)^3 \chi_{[-2,-1)}(x) + \left(\frac{2}{3} - x^2 - \frac{1}{2}x^3\right) \chi_{[-1,0)}(x) \\ & + \left(\frac{2}{3} - x^2 + \frac{1}{2}x^3\right) \chi_{[0,1)}(x) + \frac{1}{6}(2-x)^3 \chi_{[1,2]}(x) \end{aligned} \tag{39}$$

where χ_S denotes the characteristic function of a set $S \subset \mathbb{R}$. Note that B_3 is twice continously differentiable, has total weight 1, is supported on $[-2, 2]$, and piecewise cubic.

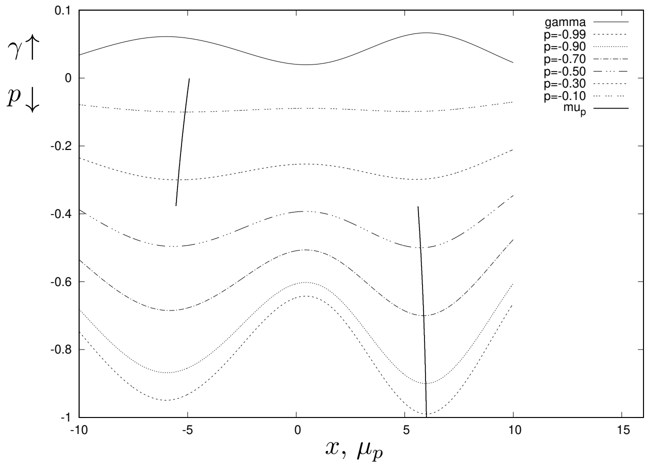
For our numerical experiments, we will use convex combinations of scaled and shifted cubic B-spline basis functions as densities. A density of this class is characterised by a positive integer n and three finite sequences $(\alpha_i)_{i=1,\dots,n}$ (weights), $(\xi_i)_{i=1,\dots,n}$ (centres), $(\sigma_i)_{i=1,\dots,n}$ (width) of real numbers with $\alpha_i > 0$ and $\sigma_i > 0$ for $i = 1, \dots, n$ and $\sum_{i=1}^{n} \alpha_i = 1$. The density then reads as

$$\gamma(x) = \sum_{i=1}^{n} \frac{\alpha_i}{\sigma_i} B_3\left(\frac{x - \xi_i}{\sigma_i}\right)\ . \tag{40}$$



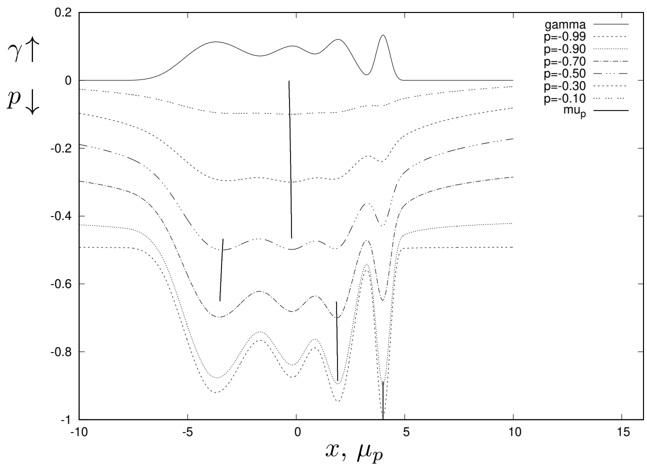
(a), $n = 2$

i	α_i	ξ_i	σ_i
1	0.55	-6	10.0
2	0.45	6	7.5



(b), $n = 2$

i	α_i	ξ_i	σ_i
1	0.55	-6	6.0
2	0.45	6	4.5



(c), $n = 5$

i	α_i	ξ_i	σ_i
1	0.30	-4	2.0
2	0.13	-2	2.0
3	0.20	0	1.5
4	0.17	2	1.0
5	0.10	4	0.5

Figure 1: Plots of example densities γ composed of B-spline basis functions according to (40) with selected objective functions $E_p(x)$ and order-p means μ_p for parameters $p \in (-1, 0)$. On top of each diagram the density γ is shown (solid line); objective functions $E_p(x)$ (dotted and dash-dotted lines) are rescaled and shifted such that the global minimum of E_p is placed at height p where also the curve of μ_p (thick solid line) passes through.

The objective function (12) from the definition of the extended order-p means, $-1 < p < 0$, for this density is given by

$$E_{p,\gamma}(\mu) = \sum_{i=1}^{n} \alpha_i \sigma_i^p \, E_{p,B_3}\left(\frac{\mu - \xi_i}{\sigma_i}\right) . \tag{41}$$

Herein, E_{p,B_3} is the objective function for the B-spline basis function (39). For its numerical evaluation we use the analytical expression (49) stated in Appendix A.2.

Fig. 1 shows numerical results computed with three different densities of this type. Each subfigure shows a density γ and objective functions $E_p(x)$ for selected values of $p \in (0,1)$ as computed by evaluation of the analytic expressions (40) and (41), respectively, for parameters shown in the figure. Minimisers μ_p were computed by grid search with resolution 0.001 on the x and p axes, and are shown in the same frame. In Fig. 1(a) it can be seen that μ_p slowly varies from μ_0 to μ_{-1} (the mode) as p decreases from 0 to -1. For the density in Fig. 1(b) the continuous evolution of μ_p is interrupted by a jump from one minimum of E_p to another at $p \approx -0.378$. The third example in Fig. 1(c) demonstrates that for sufficiently complex densities, several jumps of this kind can occur (here, at $p \approx -0.467$, $p \approx -0.652$ and $p \approx -0.887$).

5. Summary

Motivated by applications in signal and image processing, we have proposed an extension of order-p means for continuous densities to exponents p between 0 and -1. This generalisation, which has no counterpart in the case of discrete and mixed distributions, takes the family of order-p means to a presumptive endpoint since it appears to be the most general formulation of order-p means possible in the case of one-dimensional continuous densities.

We have shown that order-p means in this generalised sense approach the mode of a continuous density for $p \to -1$. Our result complements the picture drawn by earlier works by Jackson, Barral Souto and Fréchet who had shown that order-p means approach the mode for $p \to 0$ for discrete and mixed distributions but not for continuous distributions. Analytical and numerical examples were presented to demonstrate the limit result.

A generalisation to multivariate continuous densities is possible but entangled with further questions about the appropriate multivariate median concept, and equivariance properties. Therefore this direction of research has been left for future work.

We assume that the extended family of order-p means can be made useful in the one or other way in data analysis applications. However, we do not attempt at designing a particular application of this kind here because we believe that this is a task more competently done by dedicated data analysis experts.

A. Detailed calculations

In this Appendix, we provide detailed calculations for the examples from Section 4.

A.1. Derivation of the minimiser in the toy example

We derive Equation (37) of Section 4.1.

By an obvious rewrite of (36), we have

$$\mu_0(\gamma) = \operatorname*{argmin}_{\mu} \int_{-1+r}^{\mu} (1 - \lambda x^2) \ln(\mu - x) \, \mathrm{d}x + \int_{\mu}^{1+r} (1 - \lambda x^2) \ln(x - \mu) \, \mathrm{d}x \, . \tag{42}$$

Taking derivatives, we have

$$
\begin{aligned}
0 &= \frac{\mathrm{d}}{\mathrm{d}\mu}\left(\int\limits_{-1+r}^{\mu}(1-\lambda x^2)\ln(\mu-x)\,\mathrm{d}x + \int\limits_{\mu}^{1+r}(1-\lambda x^2)\ln(x-\mu)\,\mathrm{d}x\right)\\
&= \frac{\mathrm{d}}{\mathrm{d}\mu}\left(\int\limits_{0}^{1-r+\mu}(1-\lambda(\mu-x)^2)\ln x\,\mathrm{d}x + \int\limits_{0}^{1+r-\mu}(1-\lambda(\mu+x)^2)\ln x\,\mathrm{d}x\right)\\
&= \frac{\mathrm{d}}{\mathrm{d}\mu}\Bigg((1-\lambda\mu^2)\int\limits_{0}^{1-r+\mu}\ln x\,\mathrm{d}x + 2\lambda\mu\int\limits_{0}^{1-r+\mu}x\ln x\,\mathrm{d}x - \lambda\int\limits_{0}^{1-r+\mu}x^2\ln x\,\mathrm{d}x\\
&\qquad + (1-\lambda\mu^2)\int\limits_{0}^{1+r-\mu}\ln x\,\mathrm{d}x - 2\lambda\mu\int\limits_{0}^{1+r-\mu}x\ln x\,\mathrm{d}x - \lambda\int\limits_{0}^{1+r-\mu}x^2\ln x\,\mathrm{d}x\Bigg)\\
&= -2\lambda\mu\int\limits_{0}^{1-r+\mu}\ln x\,\mathrm{d}x + (1-\lambda\mu^2)\ln(1-r+\mu) + 2\lambda\int\limits_{0}^{1-r+\mu}x\ln x\,\mathrm{d}x\\
&\quad + 2\lambda\mu(1-r+\mu)\ln(1-r+\mu) - \lambda(1-r+\mu)^2\ln(1-r+\mu)\\
&\quad - 2\lambda\mu\int\limits_{0}^{1+r-\mu}\ln x\,\mathrm{d}x - (1-\lambda\mu^2)\ln(1+r-\mu) - 2\lambda\int\limits_{0}^{1+r-\mu}x\ln x\,\mathrm{d}x\\
&\quad + 2\lambda\mu(1+r-\mu)\ln(1+r-\mu) + \lambda(1+r-\mu)^2\ln(1+r-\mu)\\
&= -2\lambda\mu\left[x\ln x - x\right]_0^{1-r+\mu} - 2\lambda\mu\left[x\ln x - x\right]_0^{1+r-\mu}\\
&\quad + 2\lambda\left[\tfrac{1}{2}x^2\ln x - \tfrac{1}{4}x^2\right]_0^{1-r+\mu} - 2\lambda\left[\tfrac{1}{2}x^2\ln x - \tfrac{1}{4}x^2\right]_0^{1+r-\mu}\\
&\quad + (1-\lambda\mu^2)\ln(1-r+\mu) - (1-\lambda\mu^2)\ln(1+r-\mu)\\
&\quad + 2\lambda\mu(1-r+\mu)\ln(1-r+\mu) + 2\lambda\mu(1+r-\mu)\ln(1+r-\mu)\\
&\quad - \lambda(1-r+\mu)^2\ln(1-r+\mu) + \lambda(1+r-\mu)^2\ln(1+r-\mu)\\
&= 2\lambda(r+\mu) + (1-\lambda\mu^2)\big(\ln(1-r+\mu) - \ln(1+r-\mu)\big)\ ,
\end{aligned}
\tag{43}
$$

i.e. (37) as claimed.

A.2. Objective function of the mode limit example

We aim at deriving an analytical expression for the objective function $E_{p,B_3}(\mu)$ used in the numerical test of the mode limit, Section 4.2.

Inserting (39) for γ into (12) for $-1 < p < 0$, we have

$$
\begin{aligned}
E_{p,B_3}(\mu) &= -\int\limits_{-2}^{-1}\frac{1}{6}(2+z)^3|z-\mu|^p\,\mathrm{d}z - \int\limits_{-1}^{0}\left(\frac{2}{3}-z^2-\frac{1}{2}z^3\right)|z-\mu|^p\,\mathrm{d}z\\
&\quad - \int\limits_{0}^{1}\left(\frac{2}{3}-z^2+\frac{1}{2}z^3\right)|z-\mu|^p\,\mathrm{d}z - \int\limits_{1}^{2}\frac{1}{6}(2-z)^3|z-\mu|^p\,\mathrm{d}z\\
&= -\int\limits_{-2-\mu}^{-1-\mu}\frac{1}{6}(2+\mu+z)^3|z|^p\,\mathrm{d}z - \int\limits_{-1-\mu}^{-\mu}\left(\frac{2}{3}-(\mu+z)^2-\frac{1}{2}(\mu+z)^3\right)|z|^p\,\mathrm{d}z\\
&\quad - \int\limits_{-\mu}^{1-\mu}\left(\frac{2}{3}-(\mu+z)^2+\frac{1}{2}(\mu+z)^3\right)|z|^p\,\mathrm{d}z - \int\limits_{1-\mu}^{2-\mu}\frac{1}{6}(2-\mu-z)^3|z|^p\,\mathrm{d}z\ .
\end{aligned}
\tag{44}
$$

Abbreviating

$$I_k(a,b) = \int_a^b z^k |z|^p \,\mathrm{d}z \qquad \text{for } k = 0,1,2,3 \tag{45}$$

we can further rewrite (44) to

$$\begin{aligned}
E_{p,B_3}(\mu) = &-\frac{1}{6}(2+\mu)^3 I_0(-2-\mu,-1-\mu) - \frac{1}{2}(2+\mu)^2 I_1(-2-\mu,-1-\mu)\\
&-\frac{1}{2}(2+\mu) I_2(-2-\mu,-1-\mu) - \frac{1}{6} I_3(-2-\mu,-1-\mu)\\
&-\left(\frac{2}{3}-\mu^2-\frac{1}{2}\mu^3\right) I_0(-1-\mu,-\mu) + \left(2\mu+\frac{3}{2}\mu^2\right) I_1(-1-\mu,-\mu)\\
&+\left(1+\frac{3}{2}\mu\right) I_2(-1-\mu,-\mu) + \frac{1}{2} I_3(-1-\mu,-\mu)\\
&-\left(\frac{2}{3}-\mu^2+\frac{1}{2}\mu^3\right) I_0(-\mu,1-\mu) + \left(2\mu-\frac{3}{2}\mu^2\right) I_1(-\mu,1-\mu)\\
&+\left(1-\frac{3}{2}\mu\right) I_2(-\mu,1-\mu) - \frac{1}{2} I_3(-\mu,1-\mu)\\
&-\frac{1}{6}(2-\mu)^3 I_0(1-\mu,2-\mu) + \frac{1}{2}(2-\mu)^2 I_1(1-\mu,2-\mu)\\
&-\frac{1}{2}(2-\mu) I_2(1-\mu,2-\mu) + \frac{1}{6} I_3(1-\mu,2-\mu)\ .
\end{aligned} \tag{46}$$

Noticing that

$$I_k(a,b) = \frac{1}{p+k+1}\left(|b|^{p+k+1}(\operatorname{sgn} b)^{k+1} - |a|^{p+k+1}(\operatorname{sgn} a)^{k+1}\right) = F_k(b) - F_k(a) \tag{47}$$

with

$$F_k(x) = \frac{1}{p+k+1}|x|^{p+k+1}(\operatorname{sgn} x)^{k+1}\ , \tag{48}$$

we finally obtain

$$\begin{aligned}
E_{p,B_3}(\mu) = &\left(-\frac{4}{3}+2\mu-\mu^2+\frac{1}{6}\mu^3\right) F_0(2-\mu) + \left(\frac{2}{3}-2\mu+2\mu^2-\frac{2}{3}\mu^3\right) F_0(1-\mu)\\
&+\mu^3 F_0(-\mu) + \left(-\frac{2}{3}-2\mu-2\mu^2-\frac{2}{3}\mu^3\right) F_0(-1-\mu)\\
&+\left(\frac{4}{3}+2\mu+\mu^2+\frac{1}{6}\mu^3\right) F_0(-2-\mu)\\
&+\left(2-2\mu+\frac{1}{2}\mu^2\right) F_1(2-\mu) + \left(-2+4\mu-2\mu^2\right) F_1(1-\mu) + 3\mu^2 F_1(-\mu)\\
&+\left(-2-4\mu-2\mu^2\right) F_1(-1-\mu) + \left(2+2\mu+\frac{1}{2}\mu^2\right) F_1(-2-\mu)\\
&+\left(-1+\frac{1}{2}\mu\right) F_2(2-\mu) + (2-2\mu)\, F_2(1-\mu) + 3\mu F_2(-\mu)\\
&+(-2-2\mu)\, F_2(-1-\mu) + \left(1+\frac{1}{2}\mu\right) F_2(-2-\mu)\\
&+\frac{1}{6}F_3(2-\mu) - \frac{2}{3}F_3(1-\mu) + F_3(-\mu) - \frac{2}{3}F_3(-1-\mu) + \frac{1}{6}F_3(-2-\mu)
\end{aligned} \tag{49}$$

which is the analytical expression we used to evaluate (41) numerically.

References

Afsari B (2011). "Riemannian L^p center of mass: existence, uniqueness, and convexity." *Proceedings of the American Mathematical Society*, **139**(2), 655–673.

Aftab K, Hartley R, Trumpf J (2015). "Generalized Weiszfeld algorithms for Lq optimization." *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **37**(4), 728–745.

Armatte M (2006). "Fréchet et la médiane: un moment dans une histoire de la robustesse." *Journal de la Société Française de Statistique*, **147**(2), 23–37.

Arnaudon M, Miclo L (2014). "Means in complete manifolds: uniqueness and approximation." *ESAIM: Probability and Statistics*, **18**, 185–206. doi:10.1051/ps/2013033.

Barral Souto J (1938). "El modo y otras medias, casos particulares de una misma expresión matemática." Report No. 3, Cuadernos de Trabajo, Instituto de Biometria, Universidad Nacional de Buenos Aires, Argentina.

Boscovich RJ, Maire C (1770). *Voyage astronomique et geographique, dans l'État de l'Eglise, entrepris par l'ordre et sous les auspices du Pape Benoit XIV*. Tilliard, Paris. Translated from the Latin original report of 1755.

Bullen PS (2003). *Handbook of Means and Their Inequalities*. Springer, Dordrecht. doi:10.1007/978-94-017-0399-4.

Cabanes Y, Barbaresco F, Arnaudon M, Bigot J (2019). "Unsupervised Machine Learning for Pathological Radar Clutter Clustering: the P-Mean-Shift Algorithm." In *C&ESAR 2019*. Rennes, France. URL https://hal.science/hal-02875430.

Cazals F, Delmas B, O'Donnell T (2020). "Fréchet mean and p-mean on the unit circle: characterization, decidability, and algorithm." *Technical Report hal-02484814*, HAL. URL https://inria.hal.science/hal-02484814.

Cazals F, Delmas B, O'Donnell T (2021). "Fréchet mean and p-mean on the unit circle: decidability, algorithm, and applications to clustering on the flat torus." In *SEA 2021 – 19th Symposium on Experimental Algorithms*. Sophia Antipolis, France. URL https://inria.hal.science/hal-03183028.

Ellis RL (1844). "On the Foundations of the Theories of Probabilities." *Transactions of the Cambridge Philosophical Society*, **8**(1), 204–219.

Fechner GT (1878). "Ueber den Ausgangswerth der kleinsten Abweichungssumme, dessen Bestimmung, Verwendung und Verallgemeinerung." *Abhandlungen der mathematisch-physischen Classe der Königlich Sächsischen Gesellschaft der Wissenschaften*, **11**, 1–76.

Fréchet M (1946). "Nouvelles définitions de la valeur moyenne et des valeurs equiprobables d'un nombre aléatoire." *Annales de l'Université de Lyon, Sciences, A*, pp. 5–26.

Fréchet M (1948a). "Les élements aléatoires de nature quelconque dans un espace distancié." *Annales de l'Institut Henri Poincaré*, **10**, 215–310.

Fréchet M (1948b). "Les valeurs typiques d'ordre nul ou infini d'un nombre aléatoire." *Revue de l'Institut International de Statistique*, **16**(1), 1–22.

Fréchet M (1948c). "Positions typiques d'un élément aléatoire de nature quelconque." *Annales scientifiques de l'É.N.S., 3e série*, **65**, 211–237.

Gabor D (1965). "Information theory in electron microscopy." *Laboratory Investigation*, **14**, 801–807.

Gauss CF (1809). *Theoria Motus Corporum Coelestium in Sectionibus Conicis Solem Ambientium.* Perthes & Besser, Hamburg.

Gini C, Galvani L (1929). "Di talune estensioni dei concetti di media ai caratteri qualitativi." *Metron*, **8**, 3–209.

Glaisher JWL (1872). "On the Law of Facility of Errors of Observations and on the Method of Least Squares." *Memoirs of the Royal Astronomical Society, Part II*, **39**, 75–124.

Griffin LD (2000). "Mean, Median and Mode Filtering of Images." *Proceedings of the Royal Society of London, Series A*, **456**(2004), 2995–3004. doi:10.1098/rspa.2000.0650.

Griffin LD, Lillholm M (2003). "Mode Estimation Using Pessimistic Scale Space Tracking." In LD Griffin, M Lillholm (eds.), *Scale-Space Methods in Computer Vision*, volume 2695 of *Lecture Notes in Computer Science*, pp. 266–280. Springer, Berlin. doi:10.1007/3-540-44935-3_19.

Guichard F, Morel JM (1997). "Partial Differential Equations and Image Iterative Filtering." In IS Duff, GA Watson (eds.), *The State of the Art in Numerical Analysis*, number 63 in IMA Conference Series (New Series), pp. 525–562. Clarendon Press, Oxford.

Gustin WS (1950). "Asymptotic Behavior of Mean Value Functions." *American Mathematical Monthly*, **57**, 541–544.

Huber PJ (1981). *Robust Statistics.* Wiley, New York.

Jackson D (1921). "Note on the Median of a Set of Numbers." *Bulletin of the American Mathematical Society*, **27**, 160–164.

Jordan C (1927). *Statistique Mathématique.* Gauthier-Villars, Paris.

Laplace PS (1774). "Mémoire sur la probabilité des causes par les événements." *Mémoires de l'Académie royale des Sciences de Paris (Savants étrangers)*, **VI**, 621pp. In: Oeuvres complètes de Laplace, vol. 8, 25–65, Gauthier-Villars, Paris (1891).

Laplace PS (1820). *Théorie Analytique des Probabilités.* Courcier, Paris. In: Oeuvres complètes de Laplace, vol. 7, Gauthier-Villars, Paris (1886).

Legendre AM (1805). *Nouvelles Méthodes pour la détermination des Orbites des Comètes.* Firmin Didot, Paris.

Lindenbaum M, Fischer M, Bruckstein A (1994). "On Gabor's contribution to image enhancement." *Pattern Recognition*, **27**, 1–8.

Livadiotis G (2017). "On the convergence and law of large numbers for the non-Euclidean $\mathcal{L}^p$-means." *Entropy*, **19**, 217:1–12. doi:10.3390/e19050217.

Pennec X (2020). "Advances in geometric statistics for manifold dimension reduction." In P Grohs, M Holler, A Weinmann (eds.), *Handbook of Variational Methods for Nonlinear Geometric Data*, pp. 339–359. Springer, Cham, Switzerland. doi:10.1007/978-3-030-31351-7.

Small CG (1990). "A survey of multidimensional medians." *International Statistical Review*, **58**(3), 263–277. doi:10.2307/1403809.

Torroba PL, Cap NL, Rabal HJ, Furlan WD (1994). "Fractional Order Mean in Image Processing." *Optical Engineering*, **33**(2), 528–534. doi:10.1117/12.152006.

Tukey JW (1971). *Exploratory Data Analysis.* Addison–Wesley, Menlo Park.

Weber A (1909). *Über den Standort der Industrien.* Mohr, Tübingen.

Weiszfeld E (1937). "Sur le point pour lequel la somme des distances de n points donnés est minimum." *Tôhoku Mathematics Journal*, **43**, 355–386.

Welk M (2019). "Multivariate Medians for Image and Shape Analysis." *Technical Report eess.IV:1911.00143*, arXiv.org. doi:10.48550/arXiv.1911.00143.

Welk M, Weickert J (2019). "PDE Evolutions for M-Smoothers: From Common Myths to Robust Numerics." In M Burger, J Lellmann, J Modersitzki (eds.), *Scale Space and Variational Methods in Computer Vision*, volume 11603 of *Lecture Notes in Computer Science*, pp. 263–248. Springer, Cham. doi:10.1007/978-3-030-22368-7_19.

Welk M, Weickert J (2021). "PDE Evolutions for M-Smoothers in One, Two, and Three Dimensions." *Journal of Mathematical Imaging and Vision*, **63**(2), 157–185. doi:10.1007/s10851-020-00986-1.

Affiliation:

Martin Welk
Institute of Biomedical Image Analysis
UMIT TIROL – Private University of Health Sciences and Health Technology
A-6060 Hall in Tirol, Austria
E-mail: martin.welk@umit-tirol.at
URL: https://www.umit-tirol.at/ibia

Austrian Journal of Statistics http://www.ajs.or.at/
published by the Austrian Society of Statistics http://www.osg.or.at/
Volume VV, Issue II *Submitted:* yyyy-mm-dd
doi:10.17713/ajs.v000.i00 *Accepted:* yyyy-mm-dd