Journal of Convex Analysis
Volume 24 (2017), No. 1, 199–212

Defining a Unique Median via Minimizing Families of Norms

Jeffrey Tsang*
Dept. of Mathematics and Statistics, University of Guelph, Guelph, Canada

Rajesh Pereira†
Dept. of Mathematics and Statistics, University of Guelph, Guelph, Canada

Received: November 4, 2014
Accepted: June 11, 2015

It is well-known that the median of an even number of datapoints is not unique; by any of many equivalent definitions, any point in the interval between the innermost points qualify. Recalling that the mean can be defined by a least squares approximation to the dataset, the median via least absolute differences, we consider minimizing the $\mathcal{L}_p$ norm from the dataset to the diagonal, and compute its limit as $p \to 1^+$ — the result is *not* the midpoint as typically used. We also construct a different family of strictly convex norms converging to $\mathcal{L}_1$ exhibiting a different limit-median.

1. Introduction

Let $\vec{x} \in \mathbb{R}^n$ be given as an ordered set of real numbers, ie $x_i \leq x_{i+1}, i = 1, \ldots, n-1$. Define the standard $\mathcal{L}_p$ family of norms as

$$\|\vec{x}\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}, \quad p \in [1, \infty)$$

with the limiting case of $p = \infty$ defined by $\|\vec{x}\|_\infty = \max_{i=1}^n |x_i|$. It was known to Gauss [2] that we can consider the *mean* of $\vec{x}$, as a list of n real numbers, to be the minimizer of the $\mathcal{L}_2$ distance to the diagonal:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \operatorname*{argmin}_{\mu \in \mathbb{R}} \left\| \vec{x} - \mu \vec{1} \right\|_2$$

where $\vec{1} = (1, 1, \ldots, 1) \in \mathbb{R}^n$, while Laplace preferred the *median*(s) of $\vec{x}$ [3], which minimizes the $\mathcal{L}_1$ distance to the diagonal:

$$\tilde{x} \begin{cases} = x_{\frac{n+1}{2}}, & n \text{ odd} \\ \in \left[x_{\frac{n}{2}}, x_{\frac{n}{2}+1} \right], & n \text{ even} \end{cases} \text{ satisfies } \min_{\mu \in \mathbb{R}} \left\| \vec{x} - \mu \vec{1} \right\|_1 .$$

*J. Tsang was supported by a Natural Sciences and Engineering Research Council of Canada Postgraduate Scholarship.
†R. Pereira was supported by a Natural Sciences and Engineering Research Council of Canada Discovery Grant.

ISSN 0944-6532 / $ 2.50

Extending this, we can consider the $\mathcal{L}_p$ minimizer to the diagonal, for arbitrary p, as a central point representing $\vec{x}$ as a dataset, induced by the $\mathcal{L}_p$ norm.

The reason the median is not unique when n is even, is due to $\mathcal{L}_1$ not being strictly convex; whenever $p > 1$ the minimizer is unique. It can be shown [1] that the $\mathcal{L}_p$ minimizer, call it $\mu(p)$, is continuous for $p \in (1, \infty]$. Then [1] notes that the limit of $\mu(p)$ as $p \to 1^+$ has some natural claim to being *the* median of an even set of data. If n is odd, it is proven that the limit is $x_{\frac{n+1}{2}}$, the unique median.

We find and prove here the value of $\lim_{p \to 1^+} \mu(p)$, for n even.

2. Theorem

We phrase this problem in terms of a continuously parametrized family of convex functions, which is strictly convex except in the boundary of the parameter space. The limit of the minimizer, as an implicit function of the parameter, can be found by minimizing the derivative with respect to the parameter:

Theorem 2.1. *Let $r > 0$, $\Gamma \subseteq \mathbb{R}$ connected, and $f(x;p) : \Gamma \times [0, r) \to \mathbb{R}$ be a family of functions with parameter p, which is continuous in x, p.*

Let the functions $f(x;p)$ be convex in x given p, for all p, and strictly convex for $p > 0$. Let $f(x;p)$ be differentiable in p in some neighbourhood of $p = 0$, and the derivative (w.r.t. p) $f'(x;p)$ be continuous in x.

Denote by $x \in [a, b]$ all solutions of $\operatorname{argmin}_{x\in\Gamma} f(x;0)$. Consider $\operatorname{argmin}_{x\in[a,b]} f'(x;0)$: denote its range of minima by $[a_1, b_1] \subseteq [a, b]$, which could be a unique point. Then

- *$f'(x;0)$ is convex for $x \in [a,b]$;*
- *$f'(x;p)$ is strictly convex for $x \in [a,b]$ given $p > 0$;*
- *$\liminf_{p\to 0^+} \operatorname{argmin}_{x\in\Gamma} f(x;p) \geq a_1$;*
- *$\limsup_{p\to 0^+} \operatorname{argmin}_{x\in\Gamma} f(x;p) \leq b_1$.*

In particular, if $a_1 = b_1 = x^*$, then we have that $\lim_{p\to 0^+} \operatorname{argmin}_{x\in\Gamma} f(x;p) = x^*$.

Proof. For the first claim, suppose that $f'(x;0)$ is not convex in $[a, b]$, so

$$\exists\, x_1, x_2 \in [a,b], t \in (0,1) \\ \quad \mid \quad f'\big(tx_1 + (1-t)x_2; 0\big) > tf'(x_1;0) + (1-t)f'(x_2;0).$$

Call $x^\dagger = tx_1 + (1-t)x_2$. Fix x, then by Taylor's theorem with Peano remainder term,

$$\exists\, \xi(p), \ \lim_{p\to 0} \xi(p) = 0 \quad \mid \quad f(x;p) = f(x;0) + pf'(x;0) + p\xi(p).$$

Apply it three times to the difference to get

$$\begin{aligned} & f(x^\dagger; p) - \big(tf(x_1; p) + (1-t)f(x_2; p)\big) \\ = & \big[f(x^\dagger; 0) + pf'(x^\dagger; 0) + p\xi(p)\big] - t\,[f(x_1; 0) + pf'(x_1; 0) + p\zeta(p)] \\ & \quad - (1-t)\,[f(x_2; 0) + pf'(x_2; 0) + p\varsigma(p)] \end{aligned}$$

for some error functions ξ, ζ, ς. By hypothesis, the first terms cancel out:

$$\begin{aligned} = p\,\big[&\big(f'(x^\dagger; 0) - (tf'(x_1; 0) + (1-t)f'(x_2; 0))\big) \\ &+\big(\xi(p) - (t\zeta(p) + (1-t)\varsigma(p))\big)\big]\,. \end{aligned}$$

The first half is positive and constant, again by hypothesis; the second half has a limiting value of 0 by Taylor's theorem. Therefore, there exists some $p > 0$ such that the difference is positive. But now

$$f(x^\dagger; p) - \big(tf(x_1; p) + (1-t)f(x_2; p)\big) > 0$$

contradicts that $f(x; p)$ was convex in the first place.

For the second claim, repeat with 0 replaced by any $q > 0$, and the requirement that

$$\begin{aligned} & \exists\, x_1, x_2 \in [a, b]\,, t \in (0, 1) \\ & \qquad \mid \quad f'(tx_1 + (1-t)x_2; q) \geq tf'(x_1; q) + (1-t)f'(x_2; q) \end{aligned}$$

instead. Recall $x^\dagger = tx_1 + (1-t)x_2$. Now the difference becomes

$$\begin{aligned} & f(x^\dagger; p) - \big(tf(x_1; p) + (1-t)f(x_2; p)\big) \\ = & \big[f(x^\dagger; q) - \big(tf(x_1; q) + (1-t)f(x_2; q)\big)\big] \\ & \quad + (p-q)\,\big[f'(x^\dagger; q) - \big(tf'(x_1; q) + (1-t)f'(x_2; q)\big)\big] \\ & \quad + (p-q)\,\big[\xi(p-q) - \big(t\zeta(p-q) + (1-t)\varsigma(p-q)\big)\big] \end{aligned}$$

for some error functions ξ, ζ, ς. Since $f(x; q)$ is *strictly* convex, the first term is positive; the second is nonnegative by hypothesis. With the third limiting to 0, *a fortiori* bounded, there exists some $p > q$ such that the difference is positive. Now

$$f(x^\dagger; p) - \big(tf(x_1; p) + (1-t)f(x_2; p)\big) > 0$$

contradicts $f(x; p)$ being convex.

We prove the last two claims together. Define $\mu(p) = \operatorname{argmin}_{x\in\Gamma} f(x; p)$ for $p > 0$, unique by strict convexity, and fix any $\phi \in [a_1, b_1]$. Let $\epsilon > 0$ be given. We wish to find $\delta > 0$ such that $0 < p < \delta \;\Rightarrow\; a_1 - \epsilon < \mu(p) < b_1 + \epsilon$.

First consider $\nu \in [a, b] \setminus (a_1 - \epsilon, b_1 + \epsilon)$. By Taylor's theorem we have

$$\exists\, \xi(p),\; \lim_{p\to 0} \xi(p) = 0 \quad \mid \quad f(u, p) = f(u; 0) + pf'(u; 0) + p\xi(p).$$

Apply it twice to the difference

$$\begin{aligned}
&f(\nu;p) - f(\phi;p) \\
&= \big(f(\nu;0) - f(\phi;0)\big) + p\big(f'(\nu;0) - f'(\phi;0)\big) + p\big(\xi(p) - \zeta(p)\big)
\end{aligned}$$

for some error functions ξ, ζ. By hypothesis $f(x;0)$ is constant in $[a,b]$ thus the first term is 0. Since ν is bounded away from $[a_1, b_1]$, that being the range of minima of the convex function $f'(x;0)$, the second term is bounded away from 0. Therefore let

$$\Omega_1(\epsilon) = \inf_{\nu \in [a,b] \setminus (a_1 - \epsilon, b_1 + \epsilon)} f'(\nu;0) - f'(\phi;0) > 0.$$

The third term has a limit of 0, thus

$$\begin{aligned}
\exists\, \delta_1 > 0 \quad | \quad 0 < p < \delta_1 \quad &\Rightarrow \quad |\xi(p) - \zeta(p)| < \Omega_1(\epsilon) \\
&\Rightarrow \quad f(\nu;p) - f(\phi;p) > 0
\end{aligned}$$

and by witness (ϕ itself) we have that the minimum $\mu(p) \notin [a,b] \setminus (a_1 - \epsilon, b_1 + \epsilon)$.

Now consider $\nu \in \Gamma \setminus [a,b]$. There are two cases for each side: either $a_1 - \epsilon < a$ or not (similarly, $b_1 + \epsilon > b$ or not, the proofs are completely analogous). Suppose first that $a_1 - \epsilon < a$. Recall the difference

$$\begin{aligned}
&f(\nu;p) - f(\phi;p) \\
&= \big(f(\nu;0) - f(\phi;0)\big) + p\big(f'(\nu;0) - f'(\phi;0)\big) + p\big(\xi(p) - \zeta(p)\big)
\end{aligned}$$

for $\nu \in \Gamma \setminus (a_1 - \epsilon, \infty)$. Since ν is bounded away from $[a,b]$, the range of minima of $f(x;0)$, we have that

$$\Omega_2(\epsilon) = \inf_{\nu \in \Gamma \setminus (a_1 - \epsilon, \infty)} f(\nu;0) - f(\phi;0) > 0.$$

Factoring out p, since the second term is bounded from below (convex function minus a constant), the third term convergent to 0,

$$\begin{aligned}
&\exists\, \delta_2 > 0 \quad | \quad 0 < p < \delta_2 \\
\Rightarrow \quad & p\left[\big(f'(\nu;0) - f'(\phi;0)\big) + \big(\xi(p) - \zeta(p)\big)\right] > -\Omega_2(\epsilon) \\
\Rightarrow \quad & f(\nu;p) - f(\phi;p) > 0
\end{aligned}$$

and by witness we have that the minimum $\mu(p) \notin \Gamma \setminus (a_1 - \epsilon, \infty)$.

Suppose instead that $a_1 - \epsilon \geq a$. Now since $a < a_1$, and is therefore outside the minima range of f', $f'(a;0)$ is bounded away from $f'(\phi;0)$. By continuity of f', for any $0 < \eta < f'(a;0) - f'(\phi;0)$ (pick one),

$$\exists\, \lambda > 0 \quad | \quad f'(\nu;0) - f'(\phi;0) > \eta \quad \forall \nu \in [a - \lambda, a].$$

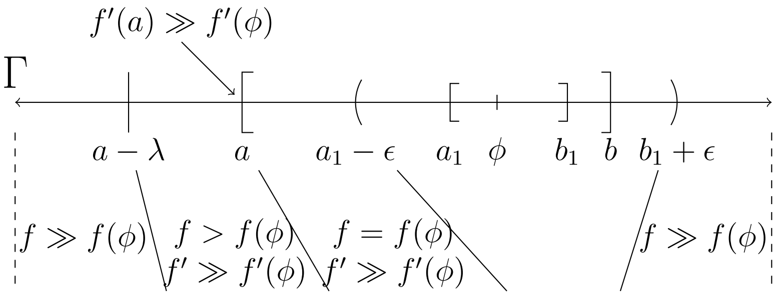

Figure 2.1: Summary of the 4 regions considered in the proof of Theorem 2.1, and the associated arguments such that $f(\nu, p) > f(\phi, p)$, hence $\mu(p)$ cannot be in each. $a_1 - \epsilon > a$ and $b_1 + \epsilon > b$ are both shown for completeness. f, f' refer to the values at $p = 0$, and $\gg$ denotes "bounded away from".

Consider $\nu \in [a - \lambda, a]$, the difference is again

$$\begin{aligned} &f(\nu; p) - f(\phi; p) \\ &= \big(f(\nu; 0) - f(\phi; 0)\big) + p\big(f'(\nu; 0) - f'(\phi; 0)\big) + p\big(\xi(p) - \zeta(p)\big), \end{aligned}$$

and this time the first term is positive by hypothesis, the second is bounded below by $\eta > 0$, and the third converges to 0 as always. Thus

$$\begin{aligned} \exists\, \delta_3 > 0 \quad | \quad 0 < p < \delta_3 \quad &\Rightarrow \quad |\xi(p) - \zeta(p)| < \eta \\ &\Rightarrow \quad f(\nu; p) - f(\phi; p) > 0 \end{aligned}$$

and by witness we have that the minimum $\mu(p) \notin [a - \lambda, a]$.

Next, notice that $a - \lambda < a$, and repeat the preceding to obtain $\delta_2 > 0$ such that $0 < p < \delta_2 \;\Rightarrow\; \mu(p) \notin \Gamma \setminus (a - \lambda, \infty)$. Finally, the analogous procedure finds $\delta_4 > 0$ such that $0 < p < \delta_4 \;\Rightarrow\; \mu(p) \notin \Gamma \setminus (-\infty, b)$.

In summation, if $0 < p < \delta = \min(\delta_1, \delta_2, \delta_3, \delta_4)$, then we have proven that

$$\begin{aligned} &\mu(p) \notin \big([a, b] \setminus (a_1 - \epsilon, b_1 + \epsilon)\big) \cup \big(\Gamma \setminus (a, \infty)\big) \cup \big(\Gamma \setminus (-\infty, b)\big) \\ \Rightarrow\;\; &\mu(p) \notin \big([a, b] \setminus (a_1 - \epsilon, b_1 + \epsilon)\big) \cup \big(\Gamma \setminus (a, b)\big) \\ \Rightarrow\;\; &\mu(p) \notin \Gamma \setminus (a_1 - \epsilon, b_1 + \epsilon) \\ \Rightarrow\;\; &\mu(p) \in (a_1 - \epsilon, b_1 + \epsilon) \end{aligned}$$

showing the limit inferior is at least a_1, and the limit superior at most b_1. □

Figure 2.1 shows a summary of the proof with a number line annotated with the regions of interest within Γ, and the arguments associated therewith.

Suppose now that we do not have a unique minimum for $f'(x; p)$. Notice that by the first two claims of Theorem 2.1, if f is additionally C^2 as well as C^1, then $f'(x; p) : [a_1, b_1] \times [0, r') \to \mathbb{R}$ satisfies the hypotheses of the theorem for some $0 < r' \leq r$, and we can recursively apply it to obtain $[a_2, b_2] \subseteq [a_1, b_1]$.

Theorem 2.2. *Let $f(x;p) : \Gamma \times [0,r) \to \mathbb{R}$ be as in Theorem 2.1. Suppose that $f(x;p)$ is enough times continuously differentiable in p as needed (if infinite, we require analyticity). Recursively apply Theorem 2.1 to $f^{(n+1)}(x;p)$ to obtain $[a_{n+1}, b_{n+1}] \subseteq [a_n, b_n]$. Then*

- *either the recursion terminates finitely with $a_n = b_n = x^*$ for some n, or*
- *the filtration $[a_1, b_1] \supseteq [a_2, b_2] \supseteq \cdots$ ends in a singleton, call the point x^*.*

Therefore, we have $\lim_{p\to 0^+} \mathrm{argmin}_{x\in\Gamma} f(x;p) = x^*$.

Proof. Suppose $\bigcap_{i=1}^{\infty} [a_n, b_n] = [\alpha, \beta]$, $\alpha < \beta$. Then we have that the entire Taylor series

$$\sum_{i=0}^{\infty} \frac{f^{(i)}(x;0)}{i!} p^i = f(x;p)$$

is termwise identically constant for the range $x \in [\alpha, \beta]$; equality by hypothesized analyticity. Therefore $f(x;p)$ is constant in $[\alpha, \beta]$ and is not strictly convex. $\square$

This is an effective algorithm to compute x^*: find the range of minima of successive derivatives of $f(x;p)$ at $p = 0$, within the range of minima of the preceding derivative, until a single point is reached, which is x^*.

3. The $\mathcal{L}_p$ limit median

We can apply the preceding theorem to the function family $\sigma(u,p) = \left\|\vec{x} - u\vec{1}\right\|_p^p$, taking $\vec{x}$ as fixed and suppressing all dependencies on it. The change in parameter space to $[1, \infty]$ is cosmetic. Direct examination of the function

$$\sigma(u,p) = \sum_{i=1}^{n} |x_i - u|^p$$

shows that it is C^∞ in both u and p unless $u = x_i$ for some i, then it is not differentiable in u, although still C^∞ in p. It is also obvious that $\sigma(u,1)$ is convex over the entire real line, and $\sigma(u,p)$ is strictly convex for $p > 1$.

Theorem 2.1 provides a constructive algorithm to determine the limit $\lim_{p\to 1^+} \mu(p)$. Minimizing the first derivative is sufficient.

Lemma 3.1. $\exists!\mu^*$ *minimizing* $\left.\frac{d\sigma}{dp}\right|_{p=1}$ *for* $u \in \left[x_{\frac{n}{2}}, x_{\frac{n}{2}+1}\right]$.

Proof. Use standard calculus to minimize $\sigma'(u,1)$ in the interval $\left[x_{\frac{n}{2}}, x_{\frac{n}{2}+1}\right]$:

$$\begin{aligned}
\frac{d}{du}\sigma'(u,1) &= \frac{d}{du}\left[\frac{d}{dp}\sum_{i=1}^{n}|x_i-u|^p\right|_{p=1} \\
&= \frac{d}{du}\sum_{i=1}^{n}|x_i-u|\log(|x_i-u|) \\
&= \sum_{i=1}^{n}\operatorname{sgn}(x_i-u)\log(|x_i-u|)+|x_i-u|\frac{\operatorname{sgn}(x_i-u)}{|x_i-u|} \\
&= \sum_{i=1}^{n}\operatorname{sgn}(x_i-u)\log(|x_i-u|)+\operatorname{sgn}(x_i-u)
\end{aligned}$$

The second term cancels out since $u \in \left(x_{\frac{n}{2}}, x_{\frac{n}{2}+1}\right)$, hence exactly $\frac{n}{2}$ terms have positive sign and the other half have negative sign. Now to minimize, set the derivative to 0:

$$\begin{aligned}
&\sum_{i=1}^{n}\operatorname{sgn}(x_i-u)\log(|x_i-u|)=0 \\
\Rightarrow \quad &\sum_{i=1}^{\frac{n}{2}}\log(|x_i-u|)=\sum_{i=\frac{n}{2}+1}^{n}\log(|x_i-u|) \\
\Rightarrow \quad &\prod_{i=1}^{\frac{n}{2}}(u-x_i)=\prod_{i=\frac{n}{2}+1}^{n}(x_i-u)
\end{aligned}$$

The left hand side is a polynomial, 0 at $x_{\frac{n}{2}}$, positive at $x_{\frac{n}{2}+1}$, and has no zeros beyond $x_{\frac{n}{2}}$, therefore is strictly increasing on $\left[x_{\frac{n}{2}}, x_{\frac{n}{2}+1}\right]$. Similarly, the right hand side is positive at $x_{\frac{n}{2}}$, 0 at $x_{\frac{n}{2}+1}$, and has no zeros before $x_{\frac{n}{2}+1}$, hence is strictly decreasing on $\left[x_{\frac{n}{2}}, x_{\frac{n}{2}+1}\right]$. Thus there is a unique solution, call it μ^*. $\square$

Theorem 2.1 proves therefore that $\lim_{p\to 1^+}\mu(p)=\mu^*$ and we have found the unique $\mathcal{L}_p$ family-induced limit median.

3.1. $n = 4$ case

It is illustrative to consider the closed form of the $\mathcal{L}_p$ limit median for the $n=4$ case (recall that $\vec{x}$ is in ascending order):

$$(\mu-x_1)(\mu-x_2)=(x_3-\mu)(x_4-\mu),$$
$$\mu=\frac{x_4x_3-x_2x_1}{x_4+x_3-x_2-x_1}$$

Several properties of note: that it is homogenous of degree 1, commutes with linear functions (including shifts), and while recovering the midpoint if $x_1 = x_2$, $x_3 = x_4$, in general it is not the midpoint. Inspection of the defining equation reveals that the degree of this polynomial is always odd; the solutions for $n = 6, 8$ (cubics) are not tractable, suggesting no simple closed form.

4. A constructed family of norms

Using the definition of a norm-minimizer to a diagonal, we can consider other parametrized families of norms that continuously go to $\mathcal{L}_1$. It has been asked in [1] whether the limit-median retrieved from the $\mathcal{L}_p$ is unique. Not surprisingly, the answer is negative, and we construct another family here to prove it.

We refer to the literature on Orlicz spaces [4] but only use a finite-dimensional space ($\mathbb{R}^n$) instead of a function space. We define the one-parameter Young function family for the Orlicz norms as

$$\theta(y;p) = \left(\sqrt{y^2+p^2} - \sqrt{p^2}\right)\left(\sqrt{1+p^2} + \sqrt{p^2}\right), \quad p \in \mathbb{R}^+$$

and the implicit function for the norm $\sigma = \left\|\vec{x} - u\vec{1}\right\|$ as

$$F(\vec{x}, u, p, \sigma) = \sum_{i=1}^{n} \theta\left(\frac{|x_i - u|}{\sigma}; p\right) = 1$$

with equality replacing the normal infimum since θ is strictly increasing.

The core of the Young function is $\sqrt{y^2+p^2}$, the rest being normalization such that $\theta(0;p) = 0$ and $\theta(1;p) = 1$. Note that this family does not require absolute values, and $\theta(y;0) = |y|$, which generates $\mathcal{L}_1$.

As an aside, let us compute the limit for the other side:

$$\begin{aligned}
\lim_{p\to\infty} \theta(y;p) &= \lim_{p\to\infty} \left(\sqrt{y^2+p^2} - \sqrt{p^2}\right)\left(\sqrt{1+p^2} + \sqrt{p^2}\right) \\
&= \lim_{p\to\infty} \frac{\sqrt{y^2+p^2} - \sqrt{p^2}}{\sqrt{1+p^2} - \sqrt{p^2}} \\
&= \lim_{h\to 0^+} \left(\sqrt{y^2 + \frac{1}{h^2}} - \sqrt{\frac{1}{h^2}}\right) \Bigg/ \left(\sqrt{1 + \frac{1}{h^2}} - \sqrt{\frac{1}{h^2}}\right) \\
&= \left(\lim_{h\to 0^+} \frac{\sqrt{h^2y^2+1} - \sqrt{1}}{h}\right) \Bigg/ \left(\lim_{h\to 0^+} \frac{\sqrt{h^2+1} - \sqrt{1}}{h}\right) \\
&= \left(\lim_{h\to 0^+} \frac{f(h) - f(0)}{h}\right) \Bigg/ \left(\lim_{h\to 0^+} \frac{g(h) - g(0)}{h}\right)
\end{aligned}$$

where $f(x) = \sqrt{1 + x^2y^2}$ and $g(x) = \sqrt{1 + x^2}$. Hence

$$\lim_{p\to\infty} \theta(y;p) = \frac{f'(0)}{g'(0)} = \left[\frac{\frac{1}{2}(2xy^2)(1+x^2y^2)^{-\frac{1}{2}}}{\frac{1}{2}(2x)(1+x^2)^{-\frac{1}{2}}}\right|_{x=0} = y^2$$

which is easily seen to generate $\mathcal{L}_2$. Therefore this construction manages to continuously connect between $\mathcal{L}_1$ and $\mathcal{L}_2$. With the dual, it thus connects $\mathcal{L}_2$ to $\mathcal{L}_\infty$ as well.

We treat $\vec{x}, u, p$ as independent variables, while σ depends on all 3. Recall that $\vec{x}$ represents our data and is held constant, and we make no further mention of it. Hence by the chain rule, we have (writing d for full derivatives and d for partial derivatives)

$$0 = \frac{\mathrm{d}F}{\mathrm{d}p} = \frac{dF}{dp} + \frac{dF}{d\sigma}\frac{d\sigma}{dp} \quad\Rightarrow\quad \frac{d\sigma}{dp} = -\frac{\frac{dF}{dp}}{\frac{dF}{d\sigma}}$$

as expected from the implicit function theorem. We can calculate the implicitly referenced terms.

$$\begin{aligned}
\frac{d\theta}{dy} &= \left(\frac{y}{\sqrt{y^2+p^2}}\right)\left(\sqrt{1+p^2}+\sqrt{p^2}\right), \\
\frac{d\theta}{dp} &= \left(\frac{p}{\sqrt{y^2+p^2}}-1\right)\left(\sqrt{1+p^2}+\sqrt{p^2}\right) \\
&\quad + \left(\sqrt{y^2+p^2}-\sqrt{p^2}\right)\left(\frac{p}{\sqrt{1+p^2}}+1\right),
\end{aligned}$$

Introduce the symbol $\psi_i = \dfrac{|x_i - \mu|}{\sigma}$ to shorten the writing a bit.

$$\begin{aligned}
\frac{dF}{dp} &= \sum_{i=1}^{n}\left(\frac{p}{\sqrt{\psi_i^2+p^2}}-1\right)\left(\sqrt{1+p^2}+\sqrt{p^2}\right) \\
&\quad + \left(\sqrt{\psi_i^2+p^2}-\sqrt{p^2}\right)\left(\frac{p}{\sqrt{1+p^2}}+1\right), \\
\frac{dF}{d\sigma} &= \sum_{i=1}^{n}\left(\frac{\psi_i}{\sqrt{\psi_i^2+p^2}}\right)\left(\sqrt{1+p^2}+\sqrt{p^2}\right)\left(-\frac{\psi_i}{\sigma}\right)
\end{aligned}$$

Define also the symbol $\Psi = \sum_{i=1}^{n} \psi_i = \sum_{i=1}^{n} \frac{|x_i-u|}{\sigma}$ as a shorthand. Now we can evaluate these derivatives at $p = 0$:

$$\left.\frac{dF}{dp}\right|_{p=0} = \sum_{i=1}^{n}(-1)(1) + \left(\sqrt{\psi_i^2}\right)(1) = \sum_{i=1}^{n}\psi_i - 1 = \Psi - n,$$
$$\left.\frac{dF}{d\sigma}\right|_{p=0} = \sum_{i=1}^{n}\left(\frac{\psi_i}{\sqrt{\psi_i^2}}\right)(1)\left(-\frac{\psi_i}{\sigma}\right) = \sum_{i=1}^{n}-\frac{\psi_i}{\sigma} = -\frac{\Psi}{\sigma}$$

and with these, $\frac{d\sigma}{dp}$. We note that for $p = 0$, that n, σ, Ψ are identical $\forall \mu \in \left[x_{\frac{n}{2}}, x_{\frac{n}{2}+1}\right]$.

$$\left.\frac{d\sigma}{dp}\right|_{p=0} = \left[-\left.\frac{\frac{dF}{dp}}{\frac{dF}{d\sigma}}\right|_{p=0}\right. = -\frac{\Psi - n}{-\frac{\Psi}{\sigma}} = \sigma\left(1 - \frac{n}{\Psi}\right)$$

which is unfortunately identically constant. Hence we calculate the second order terms. By implicitly differentiating the first equation we have

$$\begin{aligned}
0 = \frac{\mathrm{d}^2F}{\mathrm{d}p^2} &= \frac{d^2F}{dp^2} + \frac{d^2F}{dpd\sigma}\frac{d\sigma}{dp} + \frac{d}{dp}\left(\frac{dF}{d\sigma}\right)\cdot\frac{d\sigma}{dp} + \frac{dF}{d\sigma}\frac{d^2\sigma}{dp^2} \\
&= \frac{d^2F}{dp^2} + \left(2\frac{d^2F}{dpd\sigma} + \frac{d^2F}{d\sigma^2}\frac{d\sigma}{dp}\right)\cdot\frac{d\sigma}{dp} + \frac{dF}{d\sigma}\frac{d^2\sigma}{dp^2} \\
&= \frac{d^2F}{dp^2} + 2\frac{d^2F}{dpd\sigma}\frac{d\sigma}{dp} + \frac{d^2F}{d\sigma^2}\left(\frac{d\sigma}{dp}\right)^2 + \frac{dF}{d\sigma}\frac{d^2\sigma}{dp^2} \\
\Rightarrow \quad \frac{dF}{d\sigma}\frac{d^2\sigma}{dp^2} &= -\frac{d^2F}{dp^2} - 2\frac{d^2F}{dpd\sigma}\left(-\frac{\frac{dF}{dp}}{\frac{dF}{d\sigma}}\right) - \frac{d^2F}{d\sigma^2}\left(-\frac{\frac{dF}{dp}}{\frac{dF}{d\sigma}}\right)^2 \\
\frac{d^2\sigma}{dp^2} &= -\frac{\frac{d^2F}{dp^2}}{\frac{dF}{d\sigma}} + 2\frac{\frac{d^2F}{dpd\sigma}\frac{dF}{dp}}{\left(\frac{dF}{d\sigma}\right)^2} - \frac{\frac{d^2F}{d\sigma^2}\left(\frac{dF}{dp}\right)^2}{\left(\frac{dF}{d\sigma}\right)^3} \\
&= -\left(\frac{dF}{d\sigma}\right)^{-3}\left(\frac{d^2F}{dp^2}\left(\frac{dF}{d\sigma}\right)^2 - 2\frac{d^2F}{dpd\sigma}\frac{dF}{dp}\frac{dF}{d\sigma} + \frac{d^2F}{d\sigma^2}\left(\frac{dF}{dp}\right)^2\right)
\end{aligned}$$

using the equivalence of mixed partials as the function is twice differentiable.

$$
\begin{aligned}
\frac{d^2F}{dp^2} = & \sum_{i=1}^{n} \left(\frac{1}{\sqrt{\psi_i^2+p^2}} - \frac{p^2}{(\psi_i^2+p^2)^{\frac{3}{2}}} \right) \left(\sqrt{1+p^2} + \sqrt{p^2} \right) \\
& + 2 \left(\frac{p}{\sqrt{\psi_i^2+p^2}} - 1 \right) \left(\frac{p}{\sqrt{1+p^2}} + 1 \right) \\
& + \left(\sqrt{\psi_i^2+p^2} - \sqrt{p^2} \right) \left(\frac{1}{\sqrt{1+p^2}} - \frac{p^2}{(1+p^2)^{\frac{3}{2}}} \right), \\
\frac{d^2F}{dpd\sigma} = & \sum_{i=1}^{n} \left(-\frac{p\psi_i}{(\psi_i^2+p^2)^{\frac{3}{2}}} \right) \left(\sqrt{1+p^2} + \sqrt{p^2} \right) \left(-\frac{\psi_i}{\sigma} \right) \\
& + \left(\frac{\psi_i}{\sqrt{\psi_i^2+p^2}} \right) \left(\frac{p}{\sqrt{1+p^2}} + 1 \right) \left(-\frac{\psi_i}{\sigma} \right), \\
\frac{d^2F}{d\sigma^2} = & \sum_{i=1}^{n} \left(\frac{\psi_i}{\sigma\sqrt{\psi_i^2+p^2}} - \frac{\psi_i^3}{\sigma\,(\psi_i^2+p^2)^{\frac{3}{2}}} \right) \left(\sqrt{1+p^2} + \sqrt{p^2} \right) \left(-\frac{\psi_i}{\sigma} \right) \\
& + \left(\frac{\psi_i}{\sqrt{\psi_i^2+p^2}} \right) \left(\sqrt{1+p^2} + \sqrt{p^2} \right) \left(\frac{2\psi_i}{\sigma^2} \right)
\end{aligned}
$$

and at $p = 0$ we obtain:

$$
\begin{aligned}
\left.\frac{d^2F}{dp^2}\right|_{p=0} &= \sum_{i=1}^{n} \left(\frac{1}{\sqrt{\psi_i^2}} \right)(1) + 2(-1)(1) + \left(\sqrt{\psi_i^2} \right)(1) = \left(\sum_{i=1}^{n} \frac{1}{\psi_i} \right) + \Psi - 2n, \\
\left.\frac{d^2F}{dpd\sigma}\right|_{p=0} &= \sum_{i=1}^{n} (0)(1) \left(-\frac{\psi_i}{\sigma} \right) + \left(\frac{\psi_i}{\sqrt{\psi_i^2}} \right)(1) \left(-\frac{\psi_i}{\sigma} \right) = -\frac{\Psi}{\sigma}, \\
\left.\frac{d^2F}{d\sigma^2}\right|_{p=0} &= \sum_{i=1}^{n} \left(\frac{\psi_i}{\sigma\sqrt{\psi_i^2}} - \frac{\psi_i^3}{\sigma\,(\psi_i^2)^{\frac{3}{2}}} \right)(1) \left(-\frac{\psi_i}{\sigma} \right) \\
&\quad + \left(\frac{\psi_i}{\sqrt{\psi_i^2}} \right)(1) \left(\frac{2\psi_i}{\sigma^2} \right) = 0 + \frac{2\Psi}{\sigma^2}
\end{aligned}
$$

And we can now obtain $\dfrac{d^2\sigma}{dp^2}$:

$$
\begin{aligned}
\left.\frac{d^2\sigma}{dp^2}\right|_{p=0} &= \left[-\frac{\frac{d^2F}{dp^2}}{\frac{dF}{d\sigma}} + 2\frac{\frac{d^2F}{dpd\sigma}\frac{dF}{dp}}{\left(\frac{dF}{d\sigma}\right)^2} - \frac{\frac{d^2F}{d\sigma^2}\left(\frac{dF}{dp}\right)^2}{\left(\frac{dF}{d\sigma}\right)^3}\right|_{p=0} \\
&= -\frac{\Psi - 2n + \sum_{i=1}^n \frac{1}{\psi_i}}{-\frac{\Psi}{\sigma}} + 2\frac{\left(-\frac{\Psi}{\sigma}\right)(\Psi - n)}{\left(-\frac{\Psi}{\sigma}\right)^2} - \frac{\left(\frac{2\Psi}{\sigma^2}\right)(\Psi-n)^2}{\left(-\frac{\Psi}{\sigma}\right)^3} \\
&= \frac{\sigma}{\Psi}\left(\Psi - 2n + \sum_{i=1}^n \frac{1}{\psi_i} - 2(\Psi - n) + \frac{2}{\Psi}(\Psi - n)^2\right) \\
&= \frac{\sigma}{\Psi}\left(\Psi - 4n + \frac{2n^2}{\Psi} + \sum_{i=1}^n \frac{1}{\psi_i}\right).
\end{aligned}
$$

We find a term that is no longer identically constant in the median-range, namely $\sum_{i=1}^n \frac{1}{\psi_i}$; therefore all we have to do is to minimize it.

Lemma 4.1. $\exists!\mu^*$ *minimizing* $\left.\frac{d^2\sigma}{dp^2}\right|_{p=0}$ *for* $u \in \left[x_{\frac{n}{2}}, x_{\frac{n}{2}+1}\right]$.

Proof. With standard calculus, this is straightforward:

$$
\begin{aligned}
0 &= \frac{d}{du}\left[\left.\frac{d^2\sigma}{dp^2}\right|_{p=0}\right. = \frac{d}{du}\frac{\sigma}{\Psi}\left(\Psi - 4n + \frac{2n^2}{\Psi} + \sum_{i=1}^n \frac{\sigma}{|x_i - u|}\right) \\
&= \frac{\sigma^2}{\Psi}\frac{d}{du}\sum_{i=1}^n \frac{1}{|x_i - u|} = -\frac{\sigma^2}{\Psi}\sum_{i=1}^n \operatorname{sgn}(x_i - u)\,|x_i - u|^{-2} \\
\Rightarrow \quad & \sum_{i=1}^{\frac{n}{2}} \frac{1}{(x_i - u)^2} = \sum_{i=\frac{n}{2}+1}^{n} \frac{1}{(x_i - u)^2}.
\end{aligned}
$$

Again we see there is a unique solution since LHS is strictly decreasing, infinite at $x_{\frac{n}{2}}$ and finite at $x_{\frac{n}{2}+1}$, and RHS is strictly increasing, finite at $x_{\frac{n}{2}}$ and infinite at $x_{\frac{n}{2}+1}$. □

One way of interpreting the equation above is that it is the "Lagrangian point" which balances the "gravitational" inverse-square attraction to each datapoint. Viewed as a differential equation, it is an unstable equilibrium point.

5. Comparison of unit balls

We present in Figure 5.1 a comparison of the unit balls for both the $\mathcal{L}_p$ and constructed families, in the positive quadrant of the plane.

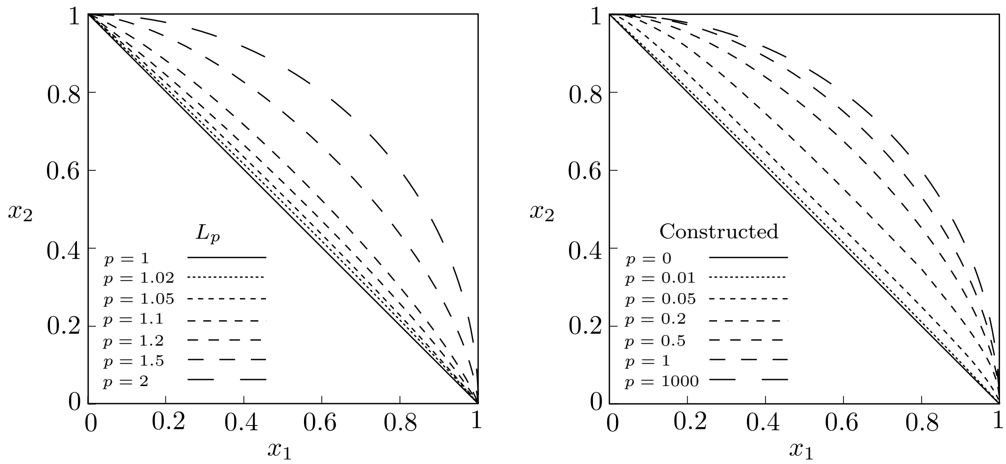

Figure 5.1: Boundaries of the unit balls for both norm families in the positive quadrant of the plane, for various values of the parameters. p in the families have different meaning; that notwithstanding, as p increases, unit balls in both families increase in size.

An immediate illustration is that for the constructed family, the $p \to \infty$ limit is indeed $\mathcal{L}_2$. The difference between them is hard to tell, but the constructed family has a "flatter" profile in the middle of the segment.

This would cause the minimizers to fall closer to the middle too, verified as that limit balances inverse squares, which decreases faster than the logarithmic $\mathcal{L}_p$ equation. Therefore the points far from the median-range have less of an effect on its limit-median.

6. Discussion and conclusions

The limit of the $\mathcal{L}_p$ norm minimizers was found, and was not the midpoint as commonly used. The naturalness of the $\mathcal{L}_p$ family may lend some credence to this alternate definition of median. However, we note two substantial objections:

Firstly, the limit-median is not unique, depending on the choice of family of norms used to define it. Given the relative prominence of $\mathcal{L}_1$ and $\mathcal{L}_2$ compared to $\mathcal{L}_p$ for $1 < p < 2$, the constructed family, which does interpolate between $\mathcal{L}_1$ and $\mathcal{L}_2$, could well rebut the presumptive position of the $\mathcal{L}_p$ family as "most" natural.

Secondly, the actual choice of definition to break the ties for medians is unlikely to matter. In reasonable usage, datasets typically number at least into the tens, if not thousands, and the range encompassed by the middle two elements is probably tiny. In that case, any heuristic, including the simple to calculate midpoint, would give essentially the same answer. If the median-range is large, this conversely argues against using the median as a summary statistic in the first place, the data being at least bimodal.

Looking at the problem from the opposite side, we can then ask *which* functions of the datapoints could concievably be the limit of a norm family minimizer? We are of course, interested in producing a family that converges to the midpoint, if possible. However, the qualitative results from the two families considered seem to contradict its existence: it looks intuitive that any strictly convex norm must depend on all the datapoints, not just the innermost — in which case the limit would still have some dependence also.

References

[1] A. Al-Salman, M. Hajja: Towards a well-defined median, J. Math. Inequal. 1(1) (2007) 23–30.

[2] C. F. Gauss: Theoria Combinationis Observationum Erroribus Minimis Obnoxiae, H. Dietrich, Göttingen (1823).

[3] P. S. Laplace: Sur quelques Points du Système du Monde, Mémoires de l'Académie Royale des Sciences de Paris, Paris (1789).

[4] M. M. Rao, Z. D. Ren: Theory of Orlicz Spaces, Marcel Dekker, New York (1991).