The Lipschitz constant for the radial projection on real ℓ_p — implementation notes

J. J. Green

25 July 2013

Abstract

It can be shown that the that the Lipschitz constant for the radial projection on the space $L^p[0,1]$ (and the sequences spaces ℓ_p) can be calculated by a one-dimensional convex optimisation. In this note we discuss some details of the implementation of a efficient Newton-Raphson iteration for this optimisation.

In [5, Thms. 3, 5], Franchetti shows the the Lipschitz constant for the radial projection of the (real) Banach space $L^p[0,1]$ is given by

$$\max_{m\in[0,1]} \left(m^{p-1} + (1-m)^{p-1}\right)^{1/p} \left(m^{p'-1} + (1-m)^{p'-1}\right)^{1/p'} \tag{1}$$

where $1/p + 1/p' = 1$. He observes that the same value is found for the Lipschitz constant of the space $\ell_p^2(\mathbb{R})$ ([5, Sect 5.], [4]). Given the isometries

$$\ell_p^2(\mathbb{R}) \subseteq \ell_p^3(\mathbb{R}) \subseteq \cdots \subseteq \ell_p(\mathbb{R}) \subseteq L^p[0,1]$$

it follows that the same holds for finite-dimensional spaces $\ell_p^n(\mathbb{R})$ $(n = 2, 3, \ldots)$ and the sequence space $\ell_p(\mathbb{R})$, a fact which is useful in the calculation of the $\ell_p \to \ell_q$ operator-norms of matrices; see, for example, [3]. We denote this value by $k(p)$ and observe that one can substitute the conjugate p' in (1) to obtain the equivalent form

$$\max_{v\in[0,1]} \frac{\left(1 + v^{p-1}\right)^{1/p} \left(1 + v^{1/(p-1)}\right)^{(p-1)/p}}{1+v}$$

which we find more useful for our calculations.

Clearly it is the same to find the maximum of the pth power and then take the pth root, so let $x = p - 1$ and write

$$f(v) = \frac{\left(1 + v^x\right)\left(1 + v^{1/x}\right)^x}{(1+v)^{x+1}} = \frac{F(v)G(v)}{H(v)}. \tag{2}$$

An obvious strategy is to take the derivative of f and seek its root. We quickly find that

$$\frac{\mathrm{d}F}{\mathrm{d}v} = xv^{x-1}, \qquad \frac{\mathrm{d}G}{\mathrm{d}v} = v^{1/x-1}\left(1 + v^{1/x}\right)^{x-1}, \qquad \frac{\mathrm{d}H}{\mathrm{d}v} = (x+1)(1+v)^x$$

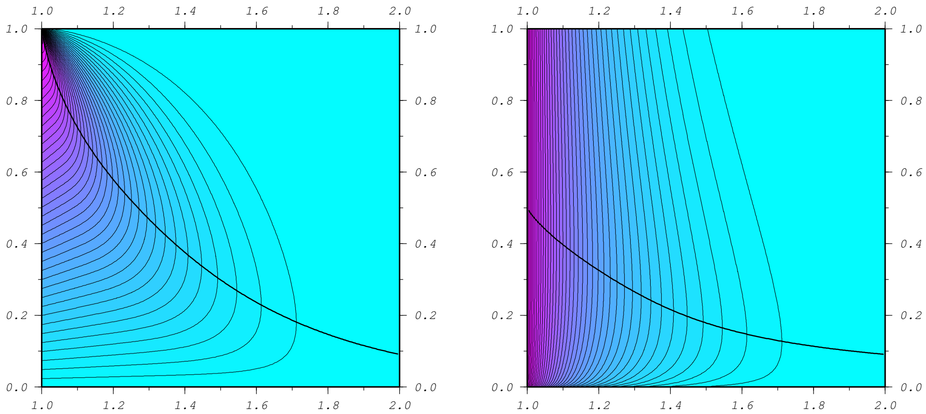

Figure 1: The functions $f(v, x)$ (left) and $f(ux, x)$ (right) as described in the text. In each case the horizontal ordinate is $p = 1 - x$ and the vertical is v, respectively u. The black line is the zero of the vertical derivative, so the location of the maxima of the functions for fixed p.

and that the derivative of f,

$$\frac{\mathrm{d}f}{\mathrm{d}v} = \frac{FG'H + F'GH - FGH'}{H^2},$$

which has the numerator $L(v)M(v)$, where

$$L(v) = \left(1 + v^x\right)^{x-1}\left(1 + v^{1/x}\right)^x v^{-1}$$

is non-zero, and $M(v)$ is

$$\left(1 + v^x\right)v^{1/x-1}(1 + v) + xv^{x-1}\left(1 + v^{1/x}\right)\left(1 + v\right) - (x + 1)\left(1 + v^x\right)\left(1 + v^{1/x}\right)$$

which can be rewritten as

$$M(v) = \left(1 + v^x\right)\left(v^{1/x} - v\right) + x\left(1 + v^{1/x}\right)\left(v^x - v\right). \tag{3}$$

One can find the solution of $M(v) = 0$ by bisection, substitute the result into (2) and take the pth root to get the required maximum. This works well for x which is not too close to zero or one (which is to say that p is not too close to one or two), but is rather inefficient computationally and difficult to analyse as x approaches zero. The reason for this is apparent when we plot $f(v, x)$ for a range of x as in the left frame of Figure 1. We see that there is a singularity as the zero-locus of M as x approaches zero.

It is for this reason that we introduce a new variable $u = v/x$ with respect to which the function f appears much better behaved, see the right-hand frame of Figure 1.

1 Newton-Raphson iteration

For x which is comfortably in the interior of $[0, 1]$, the solution to $M(v) = 0$ in u can be found rapidly by a Newton-Raphson iteration. Of course one needs the derivative of M and we record that this is

$$\big(x^3 - p(xv - v^{1/x})\big)v^{x-1} + \big(1 - p(xv - x^2v^x)\big)v^{1/x-1} - xp.$$

One also needs a good starting value for the iteration, and in our implementation we have used an order-nine Chebychev approximant of the zero locus of M determined by the bisection method, the coefficients actually calculated by the Chebfun package [1]. We find that this gives a starting value for the iteration which no further that 10^{-3} from the root over the applicable range of x. Moreover, the evaluation of the starting point from the Chebychev coefficients is extremely rapid using the recursive evaluation method of Clenshaw [2] which can be coded in a few lines of any language.

A more delicate matter is the *stopping* criterion for the iteration. In general one would seek to terminate when the difference between consecutive estimates of the zero (in other words $\delta u_i = -M(u_i)/M'(u_i)$) is around machine epsilon (ϵ_M) or a few multiples of it. The problem arises that, as x gets close to one, the derivative of M become very small, and when it drops below ϵ_M we obtain inaccurate values for δu_i. But this is a problem we can live with — a small derivative for M near the zero means that we do not need to be that close to the zero to obtain a sufficiently accurate estimate of the maximum of f; the problem is its own solution, if you will.

To quantify this idea, suppose that u is sufficiently close to the maximiser u_0 of f that it is well approximated by a quadratic,

$$f(u) \approx f(u_0) - \alpha(u - u_0)^2$$

so that $f(u_0) - f(u) < \epsilon$ if and only if $|u - u_0| < \sqrt{\epsilon/\alpha}$. But $f'(u) = -2\alpha(u - u_0)$ so $|u - u_0| = |f'(u)|/2\alpha$ and $f''(u) = f''(u_0) = 2\alpha$; thus $f(u_0) - f(u) < \epsilon$ if and only if

$$f'(u) < 2\sqrt{\epsilon\alpha} = \sqrt{2\epsilon\,|f''(u_0)|}.$$

Now, if $f'(u) = L(u)M(u)$ and $L(u)$ is nonzero, then the assumption that f is a quadratic implies already that L is a constant, so $L(u) = L(u_0)$, and likewise $M'(u) = M'(u_0)$, so that we can write the above condition as

$$M(u) < \sqrt{\left|\frac{2M'(u_0)}{L(u_0)}\right|} \times \sqrt{\epsilon}$$

The point is that we have (rather complicated) expressions for L and M and we can already evaluate u_0 by bisection, thus we can calculate the first factor on right-hand side of the above for any x and then look for a simple function of x which approximates it. Performing this calculation gives us the result shown in Figure 2 and one can see that $x(1 - x)$ is a reasonable order-of-magnitude approximant.

Using this stopping criterion (alongside the more usual one on δu_i mentioned above) and the Chebychev starting values leads to a Newton-Raphson iteration of at most three steps (more often two or one) in all of the tests performed at double precision (where machine epsilon is $\epsilon_\mathrm{M} = 2.22\times 10^{-16}$).

2 Asymptotics for p near one

The numerous degeneracies for the maximisation problem for $k(p)$ at $p = 1$ and 2 lead us to seek series expansions. We treat the $p = 1$ (so $x = 0$) case first.

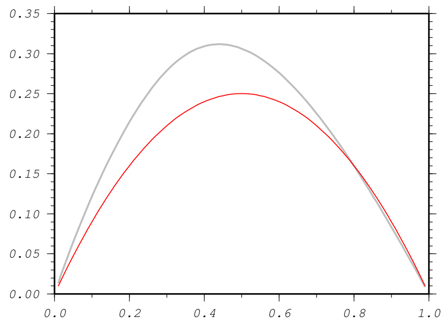

Figure 2: The quantity $\sqrt{|2M'(u_0)/L(u_0)|}$ in grey, and the function $x(1-x)$ used to approximate it in red. The horizontal ordinate is $x = p - 1$.

We first find a series expression for the root of $M(v)$, Equation 3, in terms of the variable u,

$$\big(1 + (ux)^x\big)\big((ux)^{1/x} - ux\big) + x\big(1 + (ux)^{1/x}\big)\big((ux)^x - ux\big) = 0.$$

We assume that for x sufficiently small we can neglect terms in $(ux)^{1/x}$ to obtain

$$ux\big(1 + (ux)^x\big) + x\big((ux)^x - ux\big) = (u-1)(ux)^x + u(x+1) = 0$$

and as $(ux)^x = 1 + x\log ux + \frac{1}{2}(x\log ux)^2 + \cdots$, we have

$$(u-1)(ux)^x + u(x+1) = -1 + u\big(2 + x + x\log ux + O\big(x^2\big)\big) - \big(x\log ux + O\big(x^2\big)\big)$$

and setting this to zero gives

$$u = \frac{1 + x\log ux + O\big(x^2\big)}{2 + x + x\log ux + O\big(x^2\big)}$$

and since neglecting terms in x gives $u = 1/2$, we take

$$u = \frac{1 + x\log\frac{x}{2}}{2 + x + x\log\frac{x}{2}} \tag{4}$$

as our approximation for the root of (3). Note that

$$\begin{aligned}
u &= \frac{1}{2}\left(1 + x\log\frac{x}{2}\right)\left(1 + \left(\frac{x}{2} + \frac{x}{2}\log\frac{x}{2}\right)\right)^{-1} \\
&= \frac{1}{2}\left(1 + x\log\frac{x}{2}\right)\left(1 - \left(\frac{x}{2} + \frac{x}{2}\log\frac{x}{2}\right) + O\big(x^2\big)\right) \\
&= \frac{1}{2}\left(1 + x\log\frac{x}{2} - \frac{x}{2} - \frac{x}{2}\log\frac{x}{2} + O\big(x^2\big)\right) \\
&= \frac{1}{2} - \frac{x}{4} + \frac{x}{4}\log\frac{x}{2} + O\big(x^2\big) \qquad (5)
\end{aligned}$$

Our strategy is to substitute this expression into the pth root of the terms on the right-hand side of (2), and so obtain a series for $f(ux)^{1/p}$.

First note that, using (5),

$$\begin{aligned}
\log ux &= \log x + \log\left(\frac{1}{2} - \frac{x}{4} + \frac{x}{4}\log\frac{x}{2} + O\big(x^2\big)\right) \\
&= \log x - \log 2 + \log\left(1 - \frac{x}{2} + \frac{x}{2}\log\frac{x}{2} + O\big(x^2\big)\right) \\
&= \log x - \log 2 - \frac{x}{2} + \frac{x}{2}\log\frac{x}{2} + O\big(x^2\big)
\end{aligned}$$

using the Taylor expansion for $\log(1+y)$ at $y = 0$, so

$$x\log ux = x\log\frac{x}{2} + O\big(x^2\big)$$

and then

$$\begin{aligned}
(ux)^x &= \exp\left(x\log ux\right) \\
&= 1 + x\log ux + \frac{1}{2}(x\log ux)^2 + O\big(x^3\big) \\
&= 1 + x\log\frac{x}{2} + O\big(x^2\big)
\end{aligned}$$

so

$$1 + (ux)^x = 2 + x\log\frac{x}{2} + O\big(x^2\big) = 2\left(1 + \frac{x}{2}\log\frac{x}{2} + O\big(x^2\big)\right).$$

Now, using the notation of (2),

$$F(ux)^{1/(x+1)} = 2^{1/(x+1)}\left(1 + \frac{x}{2}\log\frac{x}{2} + O\big(x^2\big)\right)^{1/(x+1)}$$

where we have

$$\begin{aligned}
&2^{1/(x+1)} \\
&= 2 \times 2^{-x/(x+1)} \\
&= 2\left(1 - \left(\frac{x}{x+1}\right)\log 2 + \frac{1}{2}\left(\frac{x}{x+1}\right)^2(\log 2)^2 - \cdots\right) \\
&= 2\left(1 - x\left(1 - x + O\big(x^2\big)\right)\log 2 + \frac{x^2}{2}\left(1 - x + O\big(x^2\big)\right)^2(\log 2)^2 - \cdots\right) \\
&= 2\left(1 - x\log 2\right) + O\big(x^2\big)
\end{aligned}$$

and

$$\begin{aligned}
&\left(1 + \frac{x}{2}\log\frac{x}{2} + O\big(x^2\big)\right)^{1/(x+1)} \\
&= 1 + \frac{1}{x+1}\left(\frac{x}{2}\log\frac{x}{2} + O\big(x^2\big)\right) + O\big(x^3\big) \\
&= 1 + \left(1 - x + O\big(x^2\big)\right)\left(\frac{x}{2}\log\frac{x}{2} + O\big(x^2\big)\right) + O\big(x^3\big) \\
&= 1 + \frac{x}{2}\log\frac{x}{2} + O\big(x^2\big)
\end{aligned}$$

which we combine to give

$$\begin{aligned}
F(ux)^{1/(x+1)} &= 2(1 - x\log 2)\left(1 + \frac{x}{2}\log\frac{x}{2}\right) + O\big(x^2\big) \\
&= 2 - 3x\log 2 + x\log x + O\big(x^2\big) \qquad (6)
\end{aligned}$$

The other terms in (2) are much easier to handle. We have

$$\begin{aligned}
G(ux)^{1/(x+1)} &= \Big(1 + (ux)^{1/x}\Big)^{x/(x+1)} \\
&= 1 + \frac{x}{x+1}(ux)^{1/x} + \frac{1}{2!}\left(\frac{x}{x+1}\right)\left(\frac{x}{x+1} - 1\right)(ux)^{1/x} + \cdots \\
&= 1 + x\big(1 - x + \cdots\big)(ux)^{1/x} + \frac{1}{2}\big(x - 2x^2 + \cdots\big)(ux)^{1/x} + \cdots \\
&= 1 + O\big(x^{1+1/x}\big) \\
&= 1 + O\big(x^2\big) \quad \text{for } x < 1,
\end{aligned} \tag{7}$$

in other words, G is asymptotically constant. Finally

$$\begin{aligned}
1/H(ux)^{1/(x+1)} &= (1 + ux)^{-1} \\
&= \big(1 + x/2 + O\big(x^2\big)\big)^{-1} \\
&= 1 - \big(x/2 + O\big(x^2\big)\big) + \big(x/2 + O\big(x^2\big)\big)^2 - \cdots \\
&= 1 - \frac{x}{2} + O\big(x^2\big).
\end{aligned} \tag{8}$$

Combining (6), (7) and (8) then gives

$$\begin{aligned}
f(ux)^{1/(x+1)} &= \big(2 - 3x\log 2 + x\log x + O\big(x^2\big)\big)\big(1 - x/2 + O\big(x^2\big)\big) \\
&= 2 - 3x\log 2 + x\log x - x + O\big(x^2\big) \\
&= 2 - x + x\log\frac{x}{8} + O\big(x^2\big).
\end{aligned} \tag{9}$$

It is this estimate which we use in our implementation for very small values of $x = p - 1$. Comparison of the estimate against the result obtained using the bisection method in high-precision arithmetic suggests that for $x < 10^{-9}$ the relative error of the estimate is less than ϵ_{M}.

3 Asymptotics for p near two

Our estimate for $k(p)$ for p near to two is much simpler than the corresponding result near one. We first note from Figure 1 that as $p - 1 = x \to 1$ the function $f(v)$ becomes flatter and flatter and it is easy to see that the limit is the constant one. Moreover, the zero locus (the position of the maximiser) seems to approach a value, denoted ξ, slightly below 0.1. Following an argument suggested by Stack Exchange Mathematics user Joriki (in question 144056) we can find the exact value of ξ.

Let $y = 1 - x = 2 - p > 0$, then it is straightforward to calculate

$$\begin{aligned}
(1 + v^x)(v^{1/x} - v) = v(v+1)\log(v)y& \\
+ v\log(v)\big((v+1) - \tfrac{1}{2}(v-1)(\log v)^2\big)y^2 + O\big(y^3\big)&
\end{aligned}$$

and

$$\begin{aligned}
(1 + v^{1/x})(v^x - v) = -v(v+1)\log(v)y& \\
+ \tfrac{1}{2}v(1-v)(\log v)^2 y^2 + O\big(y^3\big)&
\end{aligned}$$

and, combining these, we obtain a series expansion for (3) for p close to 2,

$$\begin{aligned}(1+v^x)(v^{1/x}-v) &+ x(1+v^{1/x})(v^x-v) \\ &= v\log(v)\big(2+2v+(1-v)\log(v)\big)y^2 + O\big(y^3\big).\end{aligned}$$

As $y \to 0$ (i.e., as $x \to 1$) the root of the left-hand side must converge to a root of the expression in parentheses on the right-hand side. Writing $y = (\alpha-1)/(\alpha+1)$, the latter becomes

$$\frac{2}{\alpha+1}\left(2\alpha + \log\frac{\alpha-1}{\alpha+1}\right),$$

and again, the root is the root of the parenthetic expression, namely

$$\log\frac{\alpha-1}{\alpha+1} = -2\alpha$$

leading quickly to $\coth(\alpha) = \alpha$. This $\alpha = 1.19967874\ldots$ is a quantity related the Laplace limit constant (see OEIS/A033259 and the references therein), and this then gives

$$\xi = \frac{\alpha-1}{\alpha+1} = 0.090776278\ldots$$

Motivated by the apparent flatness to Figure 1, we take the line $v = \xi$ as our approximation for the zero locus close to $x = 1$, and in this case $f(v)$ is, as a function of x, really rather simple. We use a computer algebra system (YACAS) to calculate its Taylor series at $y = x - 1 = 2 - p = 0$ and find that

$$\begin{aligned}k(2-y) = 1 &+ 0.21961441994532257538y^2 \\ &+ 0.21961441994532257538y^3 \\ &+ 0.13751213124566818941y^4\end{aligned} \tag{10}$$

to order 4. The coefficients of the y^2 and y^3 in this expression appear (numerically) to be $L^2/2$, where L is the Laplace limit constant.

Comparing the approximation (10) to the value obtained by bisection in high-precision arithmetic we find that it has a relative error of less that ϵ_{M} for $y < 10^{-3}$.

4 Accuracy and efficiency

The algorithm discussed in this document have been implemented in C and Matlab; and we here provide some details on the testing of the implementation. We generated three sets of values of $k(p)$ for 10,000 values of p uniformly distributed in the ranges $1 < p < 2$, $1 < p < 1+2\times10^{-9}$ and $2-2\times10^{-3} < p < 2$ respectively. Clearly the latter are used to inspect the series estimates described in the sections above. We find, in comparison to the corresponding value calculated by bisection in multiprecision arithmetic (to 50 significant figures) that the algorithm has a relative error of less that $1.5\epsilon_{\mathrm{M}}$ for all values tested (and the vast majority were less than ϵ_{M}). The plots of Figure 3 illustrate these results.

In terms of speed, we find an average time of 4µs is needed to evaluate $k(p)$ on a 2GHz Intel CPU.

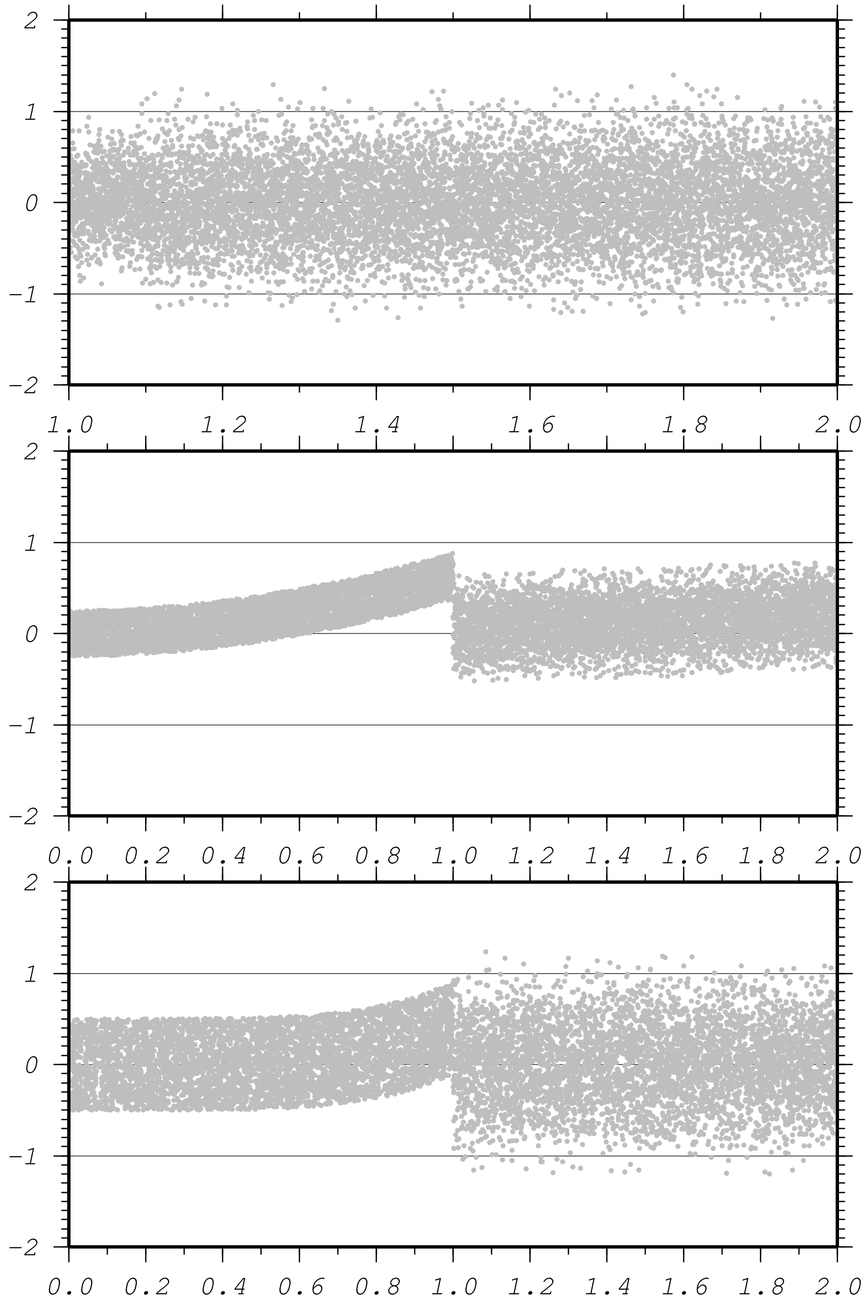

Figure 3: Accuracy of the implementation's calculation of $k(p)$ for various values of p. All plots have vertical ordinate which is the relative accuracy in multiples of machine epsilon $\epsilon_{\mathrm{M}} = 2.22{\times}10^{-16}$. The top plot shows the range $1 < p < 2$; the middle the range $0 < x' < 2$ where $x' = (p-1){\times}10^{-9}$; the bottom the range $0 < y' < 2$ where $y' = (2-p){\times}10^{-3}$ (so that $1.998 < p < 2$). The lower plots show clearly the transition from Newton-Raphson to the series estimate.

References

[1] Z. Battles and L. N. Trefethen. An extension of MATLAB to continuous functions and operators. *SIAM J. Sci. Comp.*, 25:1743–1770, 2004.

[2] C. W. Clenshaw. A note on the summation of Chebyshev series. *Math. Tab. Wash.*, 9:118–120, 1955.

[3] S. W. Drury. A counterexample to a conjecture of Matsaev. *Linear Algebra Appl.*, 435(2):323–329, July 2011.

[4] Carlo Franchetti. On the radial constant of normed spaces. *Cienc. Tecn,*, 3(1):1–9, 1979.

[5] Carlo Franchetti. The norm of the minimum projection onto hyperplanes in $L^p[0,1]$ and the radial constant. *Boll. Un. Mat. Ital.*, 4-B(7):803–821, 1990.