A note on the geometry of the multiresolution criterion

Thoralf Mildenberger *

October 13, 2006

Abstract

Several recent developments in nonparametric regression are based on the concept of data approximation: They aim at finding the simplest model that is an adequate approximation to the data. Approximations are regarded as adequate iff the residuals 'look like noise'. This is usually checked with the so-called multiresolution criterion. We show that this criterion is related to a special norm (the 'multiresolution norm'), and point out some important differences between this norm and the p-norms often used to measure the size of residuals. We also treat an important approximation problem with regard to this norm that can be solved using linear programming. Finally, we give sharp upper and lower bounds for the multiresolution norm in terms of p-norms.

1 Introduction

Some recent developments in nonparametric statistics are based on the new concept of *data approximation* proposed in Davies [3]. Much of the new methodology, including the *multiresolution criterion* we consider here, was introduced by Davies and Kovac [6] in connection with the taut-string procedure. Further results and applications to other problems are presented in [5], [4], [10], [7] and [8]. Bernholt and Hofmeister [1] give a fast algorithm to check the criterion; their point of view is also geometric, but completely different from the one adopted in this note.

Consider the nonparametric regression model

$$y(t) = f(t) + \varepsilon(t)$$

for $t = 1, \ldots, N$, where $\varepsilon(1), \ldots, \varepsilon(N)$ is iid normal distributed noise with (known) variance σ^2. Within the data approximation framework, one aims at minimizing a measure of complexity of the regression function f subject to the condition that f is an *adequate approximation* to the data. Complexity may be measured in terms of smoothness (e.g. total variation of a derivative), but also by discrete criteria (e.g. the number of modes or

*Department of Statistics, University of Dortmund. E-mail: mildenbe@statistik.uni-dortmund.de

jumps of the function). An approximation f is regarded as adequate iff the residuals "look like" white noise, i.e. are not too large and not too regular. This is formalized by the *multiresolution criterion*:

$$\max_{I\in\mathcal{I}} \frac{1}{\sqrt{|I|}} \left| \sum_{t\in I} r(t) \right| \leq C \tag{1}$$

where $r(t) = f(t) - y(t)$ denotes the residuals, and $\mathcal{I}$ is the system of all intervals in $\{1,\ldots,N\}$, although smaller systems may be used. The maximum in (1) is large when there are single large residuals or long intervals consisting mainly of residuals with the same sign. The constant C generally depends on N and the noise level σ. For large N, one may use the asymptotically justified choice $C = \sigma\sqrt{\tau \log(N)}$ for $\tau > 2$. Then, if the residuals are Gaussian white noise, (1) holds a.s. as N tends to infinity, cf. [6]. For small N, one can choose a suitable quantile of the left hand side by simulations. In the present paper, we are not concerned with the choice of C, nor will we consider any asymptotics or probabilities.

Our aim is to present a geometric interpretation of (1) that provides new insight into statistical procedures based on the multiresolution criterion. All the calculations are elementary and mainly serve the purpose of illustrating how the criterion works. Nevertheless, some of the results may be useful for theoretical comparisons to other widely used measures of fit or construction of algorithms.

It is well known that the set of all $r \in \mathbb{R}^N$ for which (1) is fulfilled can be characterized by a finite number of linear inequalities (cf. [7], ch. 2.1). However, to our knowledge, it has never been explicitly stated that this set is indeed a ball in a norm derived from the multiresolution criterion. We introduce the *multiresolution norm* in section 2 and state some elementary properties. The usefulness of the criterion as a formalized residual plot is mainly due to the fact that this norm does not share some important invariance properties that e.g. the p-norms *do* possess. We investigate this in the third section. In section 4, we show how best approximations from subspaces with respect to the multiresolution norm can be characterized as solutions of linear programming problems. The fifth section contains sharp upper and lower bounds for the multiresolution norm in terms of p-norms.

2 Multiresolution conditions as a norm

Definition 1. Let $N \in \mathbb{N}$ and $\mathcal{I} = \{I | I = \{i, i+1, \ldots, l\}, 1 \leq i \leq l \leq N\}$. The *multiresolution norm* is defined by the mapping

$$\begin{aligned} \|\cdot\|_{\mathsf{MR}} &: \mathbb{R}^N \longrightarrow \mathbb{R}_+ \\ \|(x_1,\ldots,x_n)\|_{\mathsf{MR}} &:= \max_{I\in\mathcal{I}} \frac{1}{\sqrt{|I|}} \left| \sum_{t\in I} x_t \right|. \end{aligned}$$

The following proposition states that this mapping is indeed a norm.

Proposition 1. 1. For $N \in \mathbb{N}$, $\|\cdot\|_{\mathsf{MR}}$ is a norm.

2. For $N \geq 2$, $\|\cdot\|_{\mathsf{MR}}$ does not define an inner product.

3. For $N \geq 2$, $\|\cdot\|_{\mathsf{MR}}$ is not strictly convex.

Proof. 1. The nonnegativity of $\|\cdot\|_{\mathsf{MR}}$ is obvious. Since $\mathcal{I}$ contains all singletons of $\{1, \ldots, N\}$, all components of a vector x have to be zero for $\|x\|_{\mathsf{MR}}$ to be zero.
For $x \in \mathbb{R}^N$ and $\alpha \in \mathbb{R}$, we have

$$\|\alpha x\|_{\mathsf{MR}} = \max_{I \in \mathcal{I}} \frac{1}{\sqrt{|I|}} \left| \sum_{t \in I} \alpha x_t \right| = |\alpha| \max_{I \in \mathcal{I}} \frac{1}{\sqrt{|I|}} \left| \sum_{t \in I} x_t \right| = |\alpha| \|x\|_{\mathsf{MR}}.$$

For any $x, y \in \mathbb{R}^N$, we get

$$\begin{aligned} \|x + y\|_{\mathsf{MR}} &= \max_{I \in \mathcal{I}} \frac{1}{\sqrt{|I|}} \left| \sum_{t \in I} (x_t + y_t) \right| \leq \max_{I \in \mathcal{I}} \frac{1}{\sqrt{|I|}} \left(\left| \sum_{t \in I} x_t \right| + \left| \sum_{t \in I} y_t \right| \right) \\ &\leq \max_{I \in \mathcal{I}} \frac{1}{\sqrt{|I|}} \left| \sum_{t \in I} x_t \right| + \max_{I \in \mathcal{I}} \frac{1}{\sqrt{|I|}} \left| \sum_{t \in I} y_t \right| = \|x\|_{\mathsf{MR}} + \|y\|_{\mathsf{MR}}. \end{aligned}$$

2. It is well known that a normed space $(X, \|\cdot\|)$ is an inner product space if and only if the *parallelogram identity* holds, i.e. iff

$$\|x + y\|^2 + \|x - y\|^2 = 2\|x\|^2 + 2\|y\|^2 \tag{2}$$

for any $x, y \in X$ (cf. [12], ch. I.5). For $N > 2$ consider $x = (1, 0, 0, \ldots, 0)$ and $y = (0, -1, 0, \ldots, 0)$. Then we have

$$\begin{aligned} &\|x + y\|_{\mathsf{MR}}^2 + \|x - y\|_{\mathsf{MR}}^2 = \|(1, -1, 0, \ldots, 0)\|_{\mathsf{MR}}^2 + \|(1, 1, 0, \ldots, 0)\|_{\mathsf{MR}}^2 \\ &= \left(\max\{1, 1, 0/\sqrt{2}, 0, \ldots\} \right)^2 + \left(\max\{1, 1, 2/\sqrt{2}, 0, \ldots\} \right)^2 = 1 + 2 = 3, \end{aligned}$$

but

$$2\|x\|_{\mathsf{MR}}^2 + 2\|y\|_{\mathsf{MR}}^2 = 2\|(1, 0, 0, \ldots, 0)\|_{\mathsf{MR}}^2 + 2\|(0, -1, 0, \ldots, 0)\|_{\mathsf{MR}}^2 = 2 + 2 = 4,$$

thus (2) does not hold.

3. Strict convexity means that the boundary of the unit ball contains no straight line segments, i.e. for any $x, y \in \mathbb{R}^N$ with $\|x\| = \|y\| = 1$, $\|\frac{1}{2}(x + y)\| = 1$ implies $x = y$. If we take $x = (-1, 0, \ldots, 0)$ and $y = (-1, 1, 0, \ldots, 0)$, then $\|x\|_{\mathsf{MR}} = \|y\|_{\mathsf{MR}} = 1$, but

$$\|\frac{1}{2}(x + y)\|_{\mathsf{MR}} = \|(-1, \frac{1}{2}, 0, \ldots, 0)\|_{\mathsf{MR}} = 1,$$

thus the multiresolution norm is not strictly convex.

□

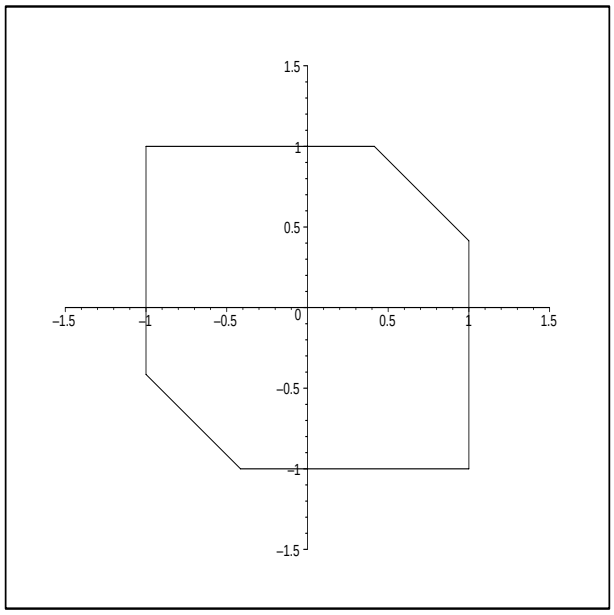

Figure 1: The multiresolution norm unit ball for $N = 2$

We also note that, for other systems of intervals, we always get a semi-norm, but the mapping is not positive definite in general. Our definition of the multiresolution norm can be rewritten as

$$\|x\|_{\mathsf{MR}} = \|Tx\|_\infty$$

using the maximum norm and a suitable matrix T. For other choices of $\mathcal{I}$ the matrix would be different. Positive definiteness holds iff T is injective. A sufficient, but not necessary condition is that the interval system contains all singletons. This is true e.g. for the dyadic scheme proposed in [6]. The matrix T is related to the noise detection criteria proposed as generalizations of the multiresolution criterion in chapter 5 of [10].

Figure 1 shows the unit ball for $N = 2$. Note that, in order to lie inside a ball, the components of a vector may be larger if the signs differ.

Since $\|\cdot\|_{\mathsf{MR}}$ is a norm, checking the multiresolution criterion means calculating a norm of the residuals; (1) then translates to:

$$\|r\|_{\mathsf{MR}} \leq C. \tag{3}$$

3 Lack of invariance

The size of residuals is often measured in other norms, the most widely used being p-norms defined by

$$\|(x_1, \ldots, x_N)\|_p = \left(\sum_{t=1}^{N} |x_t|^p \right)^{1/p} \quad \text{for } (1 \leq p < \infty)$$

and

$$\|(x_1, \ldots, x_N)\|_\infty = \max\{|x_1|, \ldots, |x_N|\}.$$

In section 4, we will give exact bounds for the multiresolution norm in terms of p-norms, but we first point out some important differences. By the definitions above, it is clear that for $1 \leq p < \infty$ the p-norm of a vector depends only on the absolute values of the components, and hence is invariant under

1. changes of sign in one or several components, and

2. permutation of the components.

To see that the multiresolution norm is not invariant under these transformations we calculate the multiresolution norms of vectors consisting of components that differ in sign only. These vectors can also be considered as permutations of each other. We have

$$\|(1, -1, 1)\|_{\mathsf{MR}} = 1,$$

but

$$\|(1, 1, -1)\|_{\mathsf{MR}} = \sqrt{2}.$$

Note that changing the signs of some components of a vector may alter the multiresolution norm, while changing all signs simultaneously does not, since this corresponds to multiplication with -1.

In the following, we use $|x| = (|x_1|, \ldots, |x_N|)$ as abbreviation for a vector that is obtained from another vector x by replacing all components by their absolute values. We have the following lemma:

Lemma 1. For $x \in \mathbb{R}^N$,

$$\|x\|_{\mathsf{MR}} \leq \||x|\|_{\mathsf{MR}},$$

and strict inequality is possible.

Proof. We have $\max_{I\in\mathcal{I}} \frac{1}{\sqrt{|I|}} \left|\sum_{t\in I} x_t\right| \leq \max_{I\in\mathcal{I}} \frac{1}{\sqrt{|I|}} \sum_{t\in I} |x_t|$. An example for strict inequality is $\|(1, -1)\|_{\mathsf{MR}} = 1 < \sqrt{2} = \|(1, 1)\|_{\mathsf{MR}}$. □

The usefulness of the multiresolution criterion is mainly due to its ability to detect long runs of residuals with the same sign which suggest that the signal has been systematically over- or underestimated on some interval. This dependence of the multiresolution norm on the sign pattern of a vector shows up in particular when considering vectors that consist of components with the same absolute size:

Proposition 2. Consider the set of all $x = (x_1, \ldots, x_N)$ with $|x_1| = \cdots = |x_N| =: m > 0$.

1. $\|x\|_{\mathsf{MR}}$ is maximal when all components have the same sign. Then $\|x\|_{\mathsf{MR}} = \sqrt{N}m$.

2. $\|x\|_{\mathsf{MR}}$ is minimal iff the components have alternating signs. Then $\|x\|_{\mathsf{MR}} = m$.

3. $\|x\|_{\mathsf{MR}} \geq \sqrt{\ell} m$, where ℓ is the length of the longest run of components with the same sign.

Proof. 1. Obvious from Lemma 1.

2. First suppose the signs are alternating. For intervals containing an even number of points all terms in the summation cancel out. For odd interval length, one term will remain. In this case we have $\frac{1}{\sqrt{|I|}} \left|\sum_{t\in I} x_t\right| = m/\sqrt{|I|}$. This is maximized for single-point-intervals, so we have $\|x\|_{\mathsf{MR}} = m$.

   Now suppose that at least two neighboring components, say x_i and x_{i+1}, have the same sign. Then $\|x\|_{\mathsf{MR}} \geq \frac{1}{\sqrt{2}}|x_i + x_{i+1}| = \frac{2}{\sqrt{2}} m = \sqrt{2} m > m$.

3. Consider an interval I consisting of components with the same sign. Then $\frac{1}{\sqrt{|I|}} |\sum_{t\in I} x_t| = \sqrt{|I|} m$.

□

Furthermore, the p-norms have a *decomposition property*:

$$\|(x_1, \ldots, x_k, x_{k+1}, \ldots, x_N)\|_p^p = \|(x_1, \ldots, x_k, 0, \ldots, 0)\|_p^p + \|(0, \ldots, 0, x_{k+1}, \ldots, x_N)\|_p^p$$

for $1 \leq p < \infty$ and

$$\|(x_1, \ldots, x_k, x_{k+1}, \ldots, x_N)\|_\infty = \max\{\|(x_1, \ldots, x_k, 0, \ldots, 0)\|_\infty, \|(0, \ldots, 0, x_{k+1}, \ldots, x_N)\|_\infty\},$$

i.e., after splitting up a vector into two subvectors, the norm of the whole vector can still be reconstructed from the norms of the two subvectors. A simple induction argument shows that this property still holds for partitions into more than two subvectors. This property is related to the so-called *reduction principle* in several complexity-penalized M-estimators for nonparametric regression, which allows for separate approximations of the signal on different subintervals. See [9] for a recent account and fast algorithms.

However, the multiresolution norm does not fulfil this property, as is easily seen: $\|(1,1)\|_{\mathsf{MR}} = \sqrt{2}$ and $\|(1,-1)\|_{\mathsf{MR}} = 1$, but the one-component subvectors have $\|(1,0)\|_{\mathsf{MR}} = \|(0,1)\|_{\mathsf{MR}} = \|(0,-1)\|_{\mathsf{MR}} = 1$. Thus there is no function that maps the norms of the subvectors to the norm of the original vector. We have the following inequality:

Proposition 3. Let $\mathcal{I}_1$ and $\mathcal{I}_2$ denote the sets of all subintervals of $\{1, \ldots, k\}$ and $\{k+1, \ldots, N\}$, respectively. Then

$$\max\left\{\max_{I\in\mathcal{I}_1} \frac{1}{\sqrt{|I|}} \left|\sum_{t\in I} x_t\right|, \max_{I\in\mathcal{I}_2} \frac{1}{\sqrt{|I|}} \left|\sum_{t\in I} x_t\right|\right\} \leq \|x\|_{\mathsf{MR}}.$$

Proof. The inequality is obvious since $\mathcal{I}_1 \cup \mathcal{I}_2 \subseteq \mathcal{I}$. Equality holds e.g. for $k = 1$ and $\|(1,-1)\|_{\mathsf{MR}}$. □

4 Best approximations via linear programming

It is well known that minimizing certain measures of complexity (like the total variation) over all signals that fulfil (1) can be considered as a linear programming problem (cf. [7], ch. 2.1). In the framework presented here, this means minimizing over a ball in the multiresolution norm with radius C. We consider a different, but related problem: Given a set $A \subset \mathbb{R}^N$ of candidates, the problem now is to decide whether A contains an adequate approximation to the data $y \in \mathbb{R}^N$, i.e. whether there exists $x \in A$ with $\|x - y\|_{\mathsf{MR}} \leq C$. This can be achieved by calculating a best approximation x_0 to y with respect to the multiresolution norm, i.e. a solution of

$$\|x_0 - y\|_{\mathsf{MR}} = \min_{x \in A} \|x - y\|_{\mathsf{MR}},$$

provided that a best approximation exists. This is guaranteed e.g. if A is compact or a linear subspace (cf. chapter 1 in [11]). If the distance of x_0 to y is larger than C, then no adequate approximation exists in A. One might then consider a larger set A' of candidate approximations. Proceeding in this way, one can – at least theoretically – construct algorithms to find an adequate approximation of minimal complexity if the measure of complexity is discrete and direct minimization over the multiresolution ball is not possible. To do this, one must be able to calculate the set of candidate approximations for a given value of the complexity measure, and existence of a best approximation must be guaranteed.

In the following, we will treat the case where A is a linear subspace of $\mathbb{R}^N$. There always exists at least one best approximation, and the set of all best approximations is convex (cf. [11], Theorem 2.2). Since the multiresolution norm is not strictly convex (cf. part 3 of Proposition 1), this set may consist of more than one element. Consider for example $y = (0, 1)$ and the subspace $A := \{(x, 0) | x \in \mathbb{R}\}$. Then

$$\|(x, 0) - (0, 1)\|_{\mathsf{MR}} = \max\{|x|, 1, |x - 1|/\sqrt{2}\}$$

and the minimum value of 1 is attained by all $(x, 0)$ with $1 - \sqrt{2} < x < 1$.

We now characterize the set

$$A^0 := \operatorname*{\mathsf{argmin}}_{x \in A} \|x - y\|_{\mathsf{MR}} \tag{4}$$

for given data $y \in \mathbb{R}^N$ and a linear subspace $A \subseteq \mathbb{R}^N$ as the solution of a linear programming problem. Let $\mathsf{dim} A = k$ and $\mathcal{A}$ be a matrix the columns of which form a basis of A. With $\lambda = (\lambda_1, \ldots, \lambda_k) \in \mathbb{R}^N$, instead of (4) we can equivalently solve (with $A^0 = \mathcal{A}(A^*)$)

$$A^* = \operatorname*{\mathsf{argmin}}_{\lambda \in \mathbb{R}^k} \|\mathcal{A}\lambda - y\|_{\mathsf{MR}}. \tag{5}$$

Combining the fact that the multiresolution ball is given by a set of linear inequalities (cf. [7], ch. 2.1) with the standard technique to solve maximum-norm approximation problems (cf. [2], p.293), we get:

Proposition 4. Let $y, \mathcal{A}, A^*$ be given as above. A^* is then the set of solutions of the linear programming problem

$$s \longrightarrow \min!$$

subject to

$$s \geq 0 \tag{6}$$

$$\sqrt{|I|}s + \sum_{t \in I} (\mathcal{A}\lambda)_t \geq \sum_{t \in I} y_t \quad \forall I \in \mathcal{I} \tag{7}$$

$$-\sqrt{|I|}s + \sum_{t \in I} (\mathcal{A}\lambda)_t \leq \sum_{t \in I} y_t \quad \forall I \in \mathcal{I} \tag{8}$$

with $\lambda \in \mathbb{R}^k$ and $s \in \mathbb{R}$.

Proof. We define

$$A_s := \{\lambda \mid \|\mathcal{A}\lambda - y\|_{\mathsf{MR}} \leq s\}. \tag{9}$$

It is now clear that $A^* = A_s$ for the smallest possible $s \geq 0$ such that $A_s \neq \emptyset$. The minimum is attained since at least one best approximation exists. This yields the target function and the side condition (6).

$$\|\mathcal{A}\lambda - y\|_{\mathsf{MR}} \leq s$$

translates into (7) and (8) by elementary manipulations. □

The problem of calculating best approximations with respect to the multiresolution norm can now be solved by standard techniques like the simplex-algorithm.

5 Bounds for other norms

Finally, after emphasizing some important differences between the multiresolution norm and the widely used p-norms, we will show how they are exactly related. Since all norms in $\mathbb{R}^N$ are equivalent, it is possible to give upper and lower bounds:

Theorem 1. For $x \in \mathbb{R}^N$:

$$\frac{1}{N^{1/p}}\|x\|_p \leq \|x\|_{\mathsf{MR}} \leq \|x\|_p \qquad (1 \leq p \leq 2) \tag{10}$$

$$\frac{1}{N^{1/p}}\|x\|_p \leq \|x\|_{\mathsf{MR}} \leq N^{1/2-1/p}\|x\|_p \qquad (2 < p < \infty) \tag{11}$$

$$\|x\|_\infty \leq \|x\|_{\mathsf{MR}} \leq \sqrt{N}\|x\|_\infty \tag{12}$$

and all bounds are sharp.

Proof. 1. Lower bounds: The case $p = \infty$ is trivial, since the multiresolution norm is the maximum over a larger set.
Now consider $p \in [1, \infty)$. Suppose there exists $x \in \mathbb{R}^N$ such that $\|x\|_p =: m$ and

$$\|x\|_{\mathsf{MR}} < \frac{m}{N^{1/p}}.$$

Since $\|x\|_\infty \le \|x\|_{\mathsf{MR}}$, we get

$$|x_t| < \frac{m}{N^{1/p}}$$

for all $t \in \{1, \dots, N\}$. This implies

$$\|x\|_p = \left(\sum_{t=1}^{N} |x_t|^p \right)^{1/p} < \left(N \frac{m^p}{N} \right)^{1/p} = m,$$

which is a contradiction to $\|x\|_p = m$.
To see that the bounds cannot be improved, take vectors of the form $x = (a, -a, a, -a, \dots)$ for $a \in \mathbb{R}$. Then $\|x\|_\infty = \|x\|_{\mathsf{MR}} = a$, and $\|x\|_p = N^{1/p} a$.

2. Upper bounds: Consider again the case $p = \infty$. Let $\|x\|_\infty =: m$ be fixed. We now construct a vector with the largest possible multiresolution norm. Because of Lemma 1, we may restrict to vectors with nonnegative components. The sum over any interval I becomes larger if we increase the absolute size of the components. The multiresolution norm then is maximized by replacing every component by the largest possible value m. Then

$$\|(m, \dots, m)\|_{\mathsf{MR}} = \max_{I \in \mathcal{I}} \frac{|I|}{\sqrt{|I|}} m = \max_{I \in \mathcal{I}} \sqrt{|I|} m = \sqrt{N} m = \sqrt{N} \|(m, \dots, m)\|_\infty.$$

For $p \in [1, \infty)$, we use the Hölder inequality. For $I \in \mathcal{I}$ we get:

$$\frac{1}{\sqrt{|I|}} \left| \sum_{t \in I} x_t \right| \le \frac{1}{\sqrt{|I|}} \sum_{t \in I} |x_t \cdot 1| \tag{13}$$

$$\le \frac{1}{\sqrt{|I|}} \left(\sum_{t \in I} |x_t|^p \right)^{1/p} \left(\sum_{t \in I} 1 \right)^{1-1/p} \tag{14}$$

$$\le \|x\|_p |I|^{1/2-1/p} \tag{15}$$

For $p \in [1, 2]$, we have $|I|^{1/2-1/p} \le 1$, and (15) leads to the upper bound in (10). To see that this bound is sharp, take the unit vectors $x = (1, 0, \dots, 0)$.
For $p > 2$, we get $|I|^{1/2-1/p} \le N^{1/2-1/p}$, so (15) gives the upper bound in (11). Equality holds again for vectors of the form $x = (m, \dots, m)$ with $\|x\|_{\mathsf{MR}} = \sqrt{N} m$ and $\|x\|_p = N^{1/p} m$.

□

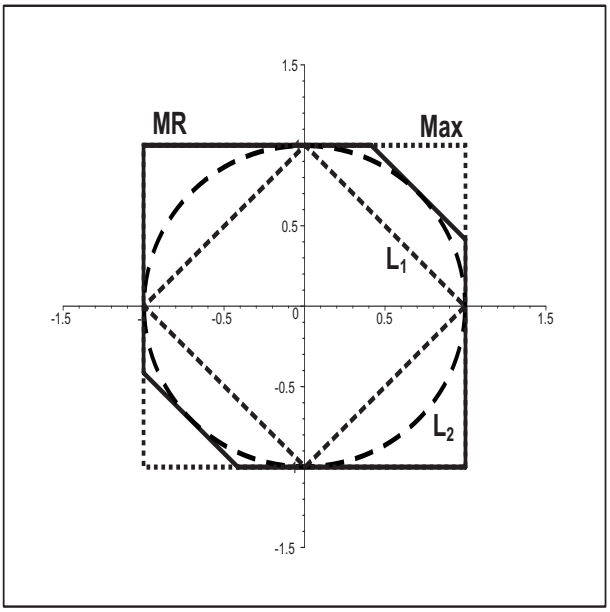

Figure 2: Multiresolution, 1-, 2- and ∞-norm unit balls for $N = 2$

Figure 2 shows the unit balls with respect to the 1-, 2-, ∞- and multiresolution norm for $N = 2$. Note that the 2-norm ball touches the multiresolution ball in $\pm(1/\sqrt{2}, 1/\sqrt{2})$, and p-norm unit balls for smaller p are entirely contained in the multiresolution unit ball. This illustrates the fact that the upper bounds in (10) and (11) are different, depending on whether $p \leq 2$ or $p > 2$.

Acknowledgements

This work has been supported by the Collaborative Research Center 'Reduction of Complexity in Multivariate Data Structures' (SFB 475) of the German Research Foundation (DFG). The author also wants to thank Prof. Dr. Ursula Gather for helpful remarks.

References

[1] T. Bernholt and T. Hofmeister. An algorithm for a generalized maximum subsequence problem. In J.R. Correa, A. Hevia, and M. Kiwi, editors, *Latin 2006: Theoretical Informatics*, volume 3887 of *Lecture notes in Computer Science*, pages 178–189, Berlin, Heidelberg, 2006. Springer Verlag.

[2] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[3] P. L. Davies. Data features. *Statistica Neerlandica*, 49(2):185–245, 1995.

[4] P. L. Davies. Approximating data and Statistical procedures – I. Approximating data. Technical Report 7/2003, SFB 475, University of Dortmund, Dortmund, Germany, 2003.

[5] P.L. Davies. Statistical procedures and robust statistics. *Estadistica*, 54:1–27, 2002.

[6] P.L. Davies and A. Kovac. Local extremes, runs, strings and multiresolution (with discussion and rejoinder). *Annals of Statistics*, 29:1–65, 2001.

[7] P.L. Davies and A. Kovac. Quantifying the cost of simultaneous non-parametric approximation of several samples. *Preprint*, 2004.

[8] P.L. Davies and M. Meise. Approximating data with weighted smoothing splines. *Preprint*, 2006.

[9] F. Friedrich, A. Kempe, V. Liebscher, and G. Winkler. Complexity Penalised M-Estimators: Fast Computation. *Preprint*, 2005.

[10] A. Majidi. Glatte nichtparametrische Regression unter formerhaltenden Bedingungen. *Dissertation, Universität Essen*, 2003.

[11] M. J. D. Powell. *Approximation theory and methods*. Cambridge University Press, 1981.

[12] K. Yosida. *Functional Analysis*. Springer Verlag, Berlin, Göttingen, Heidelberg, 1965.