Available online at www.sciencedirect.com

ScienceDirect

Journal of Complexity 23 (2007) 918–925

Journal of COMPLEXITY

www.elsevier.com/locate/jco

Deterministic constructions of compressed sensing matrices☆

Ronald A. DeVore

Department of Mathematics, University of South Carolina, Columbia, SC 29208, USA

Received 8 January 2007; accepted 16 April 2007

With high esteem to Professor Henryk Wozniakowski on the occasion of his 60th birthday

Available online 4 May 2007

Abstract

Compressed sensing is a new area of signal processing. Its goal is to minimize the number of samples that need to be taken from a signal for faithful reconstruction. The performance of compressed sensing on signal classes is directly related to Gelfand widths. Similar to the deeper constructions of optimal subspaces in Gelfand widths, most sampling algorithms are based on randomization. However, for possible circuit implementation, it is important to understand what can be done with purely deterministic sampling. In this note, we show how to construct sampling matrices using finite fields. One such construction gives cyclic matrices which are interesting for circuit implementation. While the guaranteed performance of these deterministic constructions is not comparable to the random constructions, these matrices have the best known performance for purely deterministic constructions.

Keywords: Compressed sensing; Sampling; Widths; Deterministic construction

1. Introduction

Compressed sensing (CS) offers an alternative to the classical Shannon theory for sampling signals. The Shannon theory models signals as bandlimited and encodes them through their time samples. The Shannon approach is problematic for broadband signals since the high sampling rates cannot be implemented in circuitry. In CS one replaces the bandlimited model of signals by the assumption that the signal is sparse or compressible with respect to some basis or dictionary of wave forms and enlarges the concept of sample to include the application of any linear functional.

☆ This research was conducted while the author was the visiting Texas Instrument Professor at Rice University.

E-mail address: devore@math.sc.edu.

0885-064X/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.jco.2007.04.002

Much of the methodology of CS traces back to early work on Gelfand widths and information based complexity (IBC); see [6,5,4] for a discussion of these connections.

This paper will be concerned with the discrete CS problem where we are given a discrete signal which is a vector $x \in \mathbb{R}^N$ with N large and we wish to capture x by linear information. This means that we are allowed to sample x by inner products $v \cdot x$ of x with vectors v. We are interested in seeing how well we can do given a budget $n < N$ in the number of samples we are allowed to take. This should be contrasted to the usual paradigm in compression, where one represents the signal with respect to some basis, computes all of its coefficients, but then retains only a small number (in our case n) of the largest of these coefficients to obtain compression. Here we want to see if we can avoid computing all of these coefficients and merely take a compressed number of samples to begin with.

If we choose n sampling vectors then our sampling can be represented by an $n \times N$ matrix Φ (called a CS matrix) whose rows are the vectors v that have been chosen for the sampling. Thus, the information we extract from x through Φ is the vector $y = \Phi x$ which lies in the lower dimensional space $\mathbb{R}^n$. The question becomes: What are good sampling matrices Φ?

To give this question a precise formulation, we need to specify several ingredients. First, what will we allow as decoders of y. That is how will we recover x or an approximation $\bar{x}$ to x from y. Here we will be very general and consider any mapping Δ from $\mathbb{R}^n \to \mathbb{R}^N$ as a potential decoder. The mapping Δ will generally be nonlinear—in contrast to Φ which is assumed to be linear. The problem of having practical, numerically implementable decoders is an important one and to a large extent separates CS from the earlier work on widths and IBC. However, this will not be the concern of this paper. Given that the dimensions n, N of our problem are fixed, we let $\mathscr{A}_{n,N}$ denote the set of all encoding–decoding pairs (Φ, Δ) where Φ is an $n \times N$ matrix and Δ maps $\mathbb{R}^n \to \mathbb{R}^N$.

A second ingredient is how we shall measure distortion. The vector $\bar{x} := \Delta(\Phi x)$ will in general not be the same as x. We can measure the distortion $x - \bar{x}$ in any norm on $\mathbb{R}^N$. The typical choices are the ℓ_p^N norms:

$$\|x\|_{\ell_p^N} := \begin{cases} \left(\sum_{j=1}^N |x_j|^p\right)^{1/p}, & 0 < p < \infty, \\ \max_{j=1,\dots,N} |x_j|, & p = \infty. \end{cases} \tag{1.1}$$

There are several ways in which we can measure performance of a CS matrix (see [4]). In this paper, we shall restrict our attention to only one method which relates to Gelfand widths. Given a vector $x \in \mathbb{R}^N$, the performance of the encoding–decoding pair (Φ, Δ) in the metric of ℓ_p^N is given by

$$E(x, \Phi, \Delta)_{\ell_p^N} := \|x - \Delta(\Phi x)\|_{\ell_p^N}. \tag{1.2}$$

Rather than measure the performance on each individual x, we shall measure performance on a class K. If K is a bounded set contained in $\mathbb{R}^N$, the error of this encoding–decoding on K is given by

$$E(K, \Phi, \Delta)_{\ell_p^N} := \sup_{x \in K} E(x, \Phi, \Delta)_{\ell_p^N}. \tag{1.3}$$

Thus, the error of the class K is determined by the largest error on K. The best possible performance of an encoder–decoder is given by

$$E_{n,N}(K)_{\ell_p^N} := \inf_{(\Phi,\Delta) \in \mathscr{A}_{n,N}} E(K, \Phi, \Delta)_{\ell_p^N}. \tag{1.4}$$

We say that an encoder–decoder pair $(\Phi, \Delta) \in \mathcal{A}_{n,N}$ is *near optimal* on K with constant M, if

$$E(K, \Phi, \Delta)_{\ell_p^N} \leqslant M E_{n,N}(K)_{\ell_p^N}. \tag{1.5}$$

If $M = 1$ we say the pair is *optimal*. This is the so-called min–max way of measuring optimality prevalent in approximation theory, information based complexity, and statistics.

Given a set K, the optimal performance $E_{n,N}(K)_{\ell_p^N}$ of CS is directly connected with the Gelfand widths of the set K. If K is a compact set in ℓ_p^N, and n is a positive integer, then the Gelfand width of K is by definition

$$d^n(K)_{\ell_p^N} := \inf_Y \sup\{\|x\|_{\ell_p^N} : x \in K \cap Y\}, \tag{1.6}$$

where the infimum is taken over all subspaces Y of X with *codimension* $\leqslant n$. If $K = -K$ and $K + K \subset C_0 K$, for some constant C_0, then

$$d^n(K)_{\ell_p^N} \leqslant E_{n,N}(K)_{\ell_p^N} \leqslant C_0 d^n(K)_{\ell_p^N}, \quad 1 \leqslant n \leqslant N. \tag{1.7}$$

In other words, finding the best performance of encoding–decoding on K is equivalent to finding its Gelfand width. The relation between these two problems is the following. If (Φ, Δ) is an encoding–decoding pair for CS on K, then the null space Y of Φ is a space of codimension n which is a candidate for Gelfand widths. Conversely, given any space Y for Gelfand widths then any basis for its orthogonal complement gives a CS matrix Φ for CS on K. Using these correspondences, one easily proves (1.7) (see [4]).

The Gelfand widths of the unit balls $K = U(\ell_q^N)$ in ℓ_p^N are known up to multiplicative constants. We highlight only one of these results for the Gelfand width of $U(\ell_1^N)$ in ℓ_2^N which is the deepest result in this field. It states that there exist absolute constants C_1, C_2 such that

$$C_1 \sqrt{\frac{\log(N/n)}{n}} \leqslant d^n(U(\ell_1^N))_{\ell_2^N} \leqslant C_2 \sqrt{\frac{\log(N/n)}{n}}. \tag{1.8}$$

The upper estimate in (1.8) was proved by Kashin [8] save for the correct power of the logarithm. Later Garneev and Gluskin proved the upper and lower bounds in (1.8) (see [7]). The upper bound is proved via random constructions and there remains to this date no deterministic proof of the upper bound in (1.8). In CS, their constructions correspond to random matrices whose entries are independent realizations of a Gaussian or Bernouli random variable.

Our interest in this paper centers around deterministic constructions of matrices Φ for CS. We ask how close we can get to the Gelfand width of classes with such constructions. We shall give constructions of matrices Φ using finite fields which are related to the use of finite fields to prove results on Kolmogorov widths as given in [2]. A related construction using number theory was given by Maiorov [11] (see also [10] for another deterministic construction). Our constructions will not give optimal or near optimal performance, as will be explained later. However, their performance is the best known to the author for deterministic constructions. We shall also consider modifications of this construction so that the resulting matrices Φ are circulant (each row of Φ is a certain shift of the previous row with wrapping). The importance of circulant matrices is that they can be more readily implemented in circuits.

An outline of our paper is the following. In the next section, we discuss the restricted isometry property (RIP) introduced by Candès and Tao [3] and how this property guarantees upper bounds for the performance of CS matrices on classes. The following section, gives our construction of CS matrices and the proof that they satisfy a RIP. The final section gives some concluding remarks.

2. Some simple results about CS matrices

How can we decide if a given matrix Φ is good for CS? Candès and Tao [3] have introduced a condition on matrices which they call the restricted isometry property and show that whenever a matrix Φ satisfies this property, we can obtain estimates for its performance on sets $K = U(\ell_q^N)$. For the remainder of this paper, $\|\cdot\|$ will always denote an ℓ_2 norm. All other norms will be subscripted.

If $k \geqslant 1$ is an integer, we denote by Σ_k the set of all vectors $x \in \mathbb{R}^N$ such that at most k of the coordinates of x are nonzero. In other words, Σ_k is the union of all the k-dimensional spaces X_T, $\#(T) = k$, where $T \subset \{1, \dots, N\}$ and X_T is the linear space of all $x \in \mathbb{R}^N$ which vanish outside of T. Given any vector $x \in \mathbb{R}^N$, we define

$$\sigma_k(x)_{\ell_p^N} := \inf_{z \in \Sigma_k} \|x - z\|_{\ell_p^N}, \tag{2.1}$$

which is the error of k term approximation to x in ℓ_p^N.

Following Candès and Tao, we say that Φ has the RIP of order k and constant $\delta \in (0, 1)$ if

$$(1 - \delta)\|x\|^2 \leqslant \|\Phi x\|^2 \leqslant (1 + \delta)\|x\|^2, \quad x \in \Sigma_k. \tag{2.2}$$

Notice that $\Phi x \in \mathbb{R}^n$ so that $\|\Phi x\|$ is the ℓ_2^n norm.

To get a better understanding of this property, consider the $n \times \#(T)$ matrices Φ_T formed by the columns of Φ with indices from T. Then (2.2) is equivalent to showing that the Grammian matrices

$$A_T := \Phi_T^t \Phi_T, \quad \#(T) = k, \tag{2.3}$$

are bounded and boundedly invertible on ℓ_2 with bounds as in (2.2), uniform for all T such that $\#(T) = k$. The matrix A_T is symmetric and nonnegatively definite, so this is equivalent to each of these matrices having their eigenvalues in $[1 - \delta, 1 + \delta]$.

The importance of the RIP is seen from the following theorem of Candès and Tao [3] (reinterpreted in [4]). If the $n \times N$ matrix Φ satisfies RIP of order $3k$ for some $\delta \in (0, 1)$, then there is a decoder Δ such that for any vector $x \in \mathbb{R}^N$, we have

$$\|x - \Delta(\Phi x)\|_{\ell_2^N} \leqslant C \frac{\sigma_k(x)_{\ell_1^N}}{\sqrt{k}}. \tag{2.4}$$

This means that the bigger the value of k for which we can verify the RIP then the better guarantee we have on the performance of Φ. As an example, let us return to the case of the set $K = U(\ell_1^N)$. If an $n \times N$ matrix Φ has the RIP of order k then (2.4) shows that

$$d^n(U(\ell_1^N))_{\ell_2^N} \leqslant E_{n,N}(U(\ell_1^N))_{\ell_2^N} \leqslant C/\sqrt{k}. \tag{2.5}$$

To get the optimal result we want Φ to satisfy RIP of order $k = n/\log(N/n)$. Matrices of this type can be constructed using random variables such as Gaussian or Bernouli as their entries (see [1] for example). However, there are no deterministic constructions for k of this size. In the next section, we shall give a deterministic construction of matrices Φ which satisfy RIP for a more modest range of k.

3. Deterministic constructions of CS matrices

We shall give a deterministic construction of matrices which satisfy the RIP. The vehicle for this construction are finite fields F. For simplicity of this exposition, we shall consider only the case that F has prime order and hence is the field of integers modulo p. The results we prove can be established for other finite fields as well. Given F, we consider the set $F \times F$ of ordered pairs. Note that this set has $n := p^2$ elements. Given any integer $0 < r < p$, we let $\mathbb{P}_r$ denote the set of polynomials of degree $\leqslant r$ on F. There are $N := p^{r+1}$ such polynomials. Any polynomial $Q \in \mathbb{P}_r$ can be represented as $Q(x) = a_0 + a_1 x + \cdots + a_r x^r$ where the coefficients $a_0, \ldots, a_r$ are in F. If we consider this polynomial as a mapping of F to F then its graph $\mathcal{G}(Q)$ is the set of ordered pairs $(x, Q(x))$, $x \in F$. This graph is a subset of $F \times F$.

We order the elements of $F \times F$ lexicographically as $(0, 0), (0, 1), \ldots, (p-1, p-1)$. For any $Q \in \mathbb{P}_r$, we denote by v_Q the vector indexed on $F \times F$ which takes the value one at any ordered pair from the graph of Q and takes the value zero otherwise. Note that there are exactly p ones in v_Q; one in the first p entries, one in the next p entries, and so on.

Theorem 3.1. *Let Φ_0 be the $n \times N$ matrix with columns v_Q, $Q \in \mathbb{P}_r$, with these columns ordered lexicographically with respect to the coefficients of the polynomials. Then, the matrix $\Phi := \frac{1}{\sqrt{p}}\Phi_0$ satisfies the RIP with $\delta = (k-1)r/p$ for any $k < p/r + 1$.*

Proof. Let T be any subset of column indices with $\#(T) = k$ and let Φ_T be the matrix created from Φ by selecting these columns. The Grammian matrix $A_T := \Phi_T^t \Phi_T$ has entries $\frac{1}{p} v_Q \cdot v_R$ with $Q, R \in \mathbb{P}_r$. The diagonal entries of A_T are all one. For any $Q, R \in \mathbb{P}_r$ with $Q \neq R$, there are at most r values of $x \in F$ such that $Q(x) = R(x)$. So any off diagonal entry of A_T is $\leqslant r/p$. It follows that the off diagonal entries in any row or column of A_T have sum $\leqslant (k-1)r/p = \delta < 1$ whenever $k < p/r + 1$. Hence we can write

$$A_T = I + B_T, \tag{3.1}$$

where $\|B_T\| \leqslant \delta$ where the norm is taken on either of ℓ_1 or ℓ_∞. By interpolation of operators, the norm of B_T is $\leqslant \delta$ as an operator from ℓ_2 to ℓ_2. It follows that the spectral norm of A_T is $\leqslant 1 + \delta$ and that of its inverse is $\leqslant (1-\delta)^{-1}$. This verifies (2.2) and proves the lemma. □

Notice that since $n = p^2$ and $N = p^{r+1}$, $\log(N/n) = (r-1)\log p = (r-1)\log(n)/2$, we have constructed matrices that satisfy RIP for the range $k - 1 < p/r < \sqrt{n}\log n/(2\log(N/n))$.

Our next goal is to modify the above construction to obtain circulant matrices $\Phi = (\phi_{i,j})$. A circulant matrix has the property that

$$\phi_{i+1, j+\ell} = \phi_{i,j}, \tag{3.2}$$

where $\ell := N/n$ and the arithmetic on indices is done modulo N. Hence a circulant matrix is determined by its first ℓ columns. Once these columns have been specified, all other entries are determined by imposing condition (3.2). Each other column will be a cyclic shift of one of the first ℓ columns.

As in the previous theorem, our construction will use the vectors v_Q, $Q \in \mathbb{P}_r$, to generate the first ℓ columns. However, now we must be more selective in which polynomials we shall choose for these columns. Let us observe how we fill out the matrix from its first ℓ columns. The next block of ℓ columns is each gotten by a cyclic shift. For example each column with index $m + \ell$

with $m \in \{1, \dots, \ell\}$ is obtained by taking the entries in column m and shifting them down one while the last entry in the mth column is moved to the top position. We continue in this fashion to the next block of ℓ columns and so forth. There will be $n = p^2$ such blocks. Consider the jth block, $0 \leqslant j \leqslant n-1$. We can write $j = a + bp$ with $a, b \in \{0, \dots, p-1\}$. Each column in this block will be a cyclic shift of the corresponding column v_Q from the first block. Recall that we index the rows of Φ by $(x, y) \in F \times F$. The entry in the (x, y) position of v_Q will now occupy the position (x', y') where $y' = y + j = y + a$ modulo p and $x' = x + b$ modulo p or $x' = x + b + 1$ modulo p. Since the ones in v_Q occur precisely in the positions $(x, Q(x))$ the new ones in the corresponding column of block j will occur either in position (x', y') where $y' = Q(x) + a$ modulo p and $x' = x + b$ modulo p or $x' = x + b + 1$ modulo p.

To describe the set of polynomials we shall use for the columns, we define the equivalence relation that two polynomials P, Q of degree r over F are equivalent (written $P \equiv Q$) if there exist $a, b \in F$ such that

$$P(x) = Q(x + a) + b, \quad \forall x \in F. \tag{3.3}$$

Let us see what the structure of such an equivalence class is. For this, we use the simple lemma.

Lemma 3.2. *If f is any function on F for which there exist $a, b \in F$, not both zero, such that $f(x) = f(x + a) + b$ for all $x \in F$, then f is a linear function.*

Proof. It follows that $f(a) = f(0) - b$ and more generally $f(ka) = f(0) - kb$, for each $k \in F$. If $a \neq 0$, then ka, $k = 1, \dots, p$ exhaust F and so $f(x) = f(0) - a^{-1}bx$ for all $x \in F$ so that f is linear. If $a = 0$, then $f(x) = f(x) + b$ and hence $b = 0$ as well. □

Let us now consider the equivalence classes. One equivalence class consists of all the constant functions; there are p functions in this equivalence class. For each $P(x) = \alpha x$ with $\alpha \neq 0$, its equivalence class will consist of all linear functions of the form $\alpha x + b$, $b \in F$; there are again p functions in each of these equivalence classes. Finally if P is a polynomial which is not linear, then its equivalence class will consist of the p^2 polynomials $P(x + a) + b$ corresponding to the p^2 choices of a, b (see Lemma 3.2).

Let Λ_r consist of a set of representatives from each of the equivalence classes which do not consist of linear polynomials. That is we choose one representative from each of these equivalence classes except that we never take polynomials of degree $\leqslant 1$. Let us see what the cardinality of Λ_r is. There are p^{r+1} polynomials of degree $\leqslant r$ and p^2 linear polynomials. So there are $p^{r+1} - p^2$ polynomials which are not linear. They are divided into sets of size p^2 (the equivalence classes). Hence, $\ell := \#(\Lambda_r) = p^{r-1} - 1$. Now, there are $n = p^2$ cyclic shifts so $N = p^{r+1} - p^2$.

In going further in this section, let Φ_0 denote the circulant matrix whose first ℓ columns are the v_Q, $Q \in \Lambda_r$ written in lexicographic order. Our next lemma bounds the inner products of any two columns of Φ_0.

Lemma 3.3. *For any two columns $v \neq w$ from the matrix Φ_0, we have*

$$|v \cdot w| \leqslant 4r. \tag{3.4}$$

Proof. Each of the columns v, w of Φ_0 can be described as a cyclic shift of vectors v_Q, v_R with $Q, R \in \Lambda_r$. As we have observed above, there are integers a_0, b_0 (depending only on v) such that any one in column v occurs at a position (x', y') if and only if $x' = x + b_0 + \varepsilon_0$ and $y' = Q(x) + a_0$

with $x \in F$ and $\varepsilon_0 \in \{0, 1\}$. Similarly, a one occurs in column w at position (x'', y'') if and only if $x'' = \bar{x} + b_1 + \varepsilon_1$ and $y'' = R(\bar{x}) + a_1$ with $\bar{x} \in F$ and $\varepsilon_1 \in \{0, 1\}$. The inner product $v \cdot w$ counts the number of row positions for which there is a one in each of these two columns. That is the number of solutions to $x + b_0 + \varepsilon_0 = \bar{x} + b_1 + \varepsilon_1$ and $Q(x) + a_0 = R(\bar{x}) + a_1$ with $x, \bar{x} \in F$ and $\varepsilon_0, \varepsilon_1 \in \{0, 1\}$.

Consider first the case when $Q \neq R$. We fix one of the four possibilities for $\varepsilon_0, \varepsilon_1$. These equations mean that $\bar{x} = x + b$ and $R(x + b) = Q(x) + a$ with $b = b_0 - b_1 + \varepsilon_0 - \varepsilon_1$ and $a = a_0 - a_1$. Since $R \neq Q$, we know that $R(\cdot + b)$ is not identical to $Q(\cdot) + a$ because these R and Q are not equivalent. In this case the only possible x which can satisfy the above are the zeros of the nonzero polynomial $R(\cdot + b) - Q(\cdot) - a$. Thus there are at most r such x because this latter polynomial has degree $\leqslant r$. Since there are four possibilities for $(\varepsilon_0, \varepsilon_1)$, we have $|v \cdot w| \leqslant 4r$ as desired.

Now consider the case when $R = Q$ and any one of the four possible values for $(\varepsilon_0, \varepsilon_1)$. Similar to the case just handled, we have that $\bar{x} = x + b$ and $Q(x + b) - a = Q(x)$. We are interested in the number of x for which this can happen. As long as these two polynomials are not identical this can happen at most r times. But we know that they can only be identical if Q is linear (see Lemma 3.2) and we know linear polynomials are not in Λ_r. Thus, even in the case $Q = R$ we also have that $|v \cdot w|$ is at most $4r$. □

Theorem 3.4. *The cyclic matrix* $\Phi := \frac{1}{\sqrt{p}}\Phi_0$ *has the RIP* (2.2) *with* $\delta = 4(k-1)r/p$ *whenever* $k - 1 < p/4r$.

Proof. The proof is the same as that of Theorem 3.1. □

Notice that since $n = p^2$ and $N = p^{r+1} - p^2$, $\log(N/n) < (r-1)\log p = (r-1)\log(n)/2$, we have constructed matrices that satisfy RIP for the range $k-1 < p/(4r) < \sqrt{n}\log n/(8\log(N/n))$.

4. Concluding remarks

The matrices of our two theorems satisfy RIP of order k for $k \leqslant C\sqrt{n}\log n/\log(N/n)$ which is the largest range of k that is known to the author for deterministic constructions. However, it falls far short of the range $k \leqslant Cn/\log(N/n)$ known for probabilistic constructions. The fact is that we know from probabilistic constructions that there exist $n \times N$ matrices Φ with entries $\pm 1/\sqrt{n}$ that satisfy RIP for the larger range $k \leqslant Cn/\log(N/n)$. We just cannot explicitly describe one of these matrices when N and n are large. It is therefore very interesting to try to obtain a larger range of k with deterministic methods and to understand if there are any essential limitations to deterministic methods.

Let us point out some of the deficiencies in our approach. First, we begin by asking what are good compressed sensing matrices. The restricted isometry property is just a sufficient condition to guarantee that a matrix Φ has good performance on classes. Two matrices can has exactly the same performance on classes and yet one will satisfy RIP and the other not. So there may be a more direct avenue to constructing good CS matrices by not going through RIP.

The RIP is a condition on the spectral norm of the matrices $A_T = \Phi_T^t\Phi_T$. We have bounded the spectral norm by bounding the ℓ_1 and ℓ_∞ norms (which are much easier to handle than the spectral norm) and then using interpolation. The bounds we have gotten on k appear to be the best we could expect to get by this approach. Indeed, with an eye toward results on distribution of scalar products of unit vectors (see [9, Lemma 4.1, Chapter 14]), it seems that we could not

improve much on the bounds we gave for diagonal dominance. Of course, the spectral norm of a matrix can be much smaller than the ℓ_1, ℓ_∞ norms. Thus it may be that estimating the spectral norm directly may be the way to go to obtain stronger results than ours.

Acknowledgments

The author thanks the Electrical and Computer Engineering Department at Rice, in particular Professor Rich Baraniuk, for their great hospitality. This research was supported by the Office of Naval Research Contracts ONR-N00014-03-1-0051, ONR/DEPSCoR N00014-03-1-0675, and ONR/DEPSCoR N00014-05-1-0715; and the National Science Foundation Grant DMS-354707.

References

[1] R. Baraniuk, M. Davenport, R. DeVore, M. Wakin, The Johnson–Lindenstrauss meets compressed sensing, Constr. Approx., to appear.
[2] C. de Boor, R. DeVore, K. Hoellig, Mixed norm n-widths, Proc. Amer. Math. Soc. 80 (1980) 577–583.
[3] E. Candès, T. Tao, Decoding by linear programming, IEEE Trans. Inform. Theory 51 (2005) 4203–4215.
[4] A. Cohen, W. Dahmen, R. DeVore, Compressed sensing and best k-term approximation, submitted for publication.
[5] R. DeVore, Optimal Computation, vol. I, Proceedings of ICM 2006, Madrid, European Mathematical Society Publishing House, 2007, to appear.
[6] D. Donoho, Compressed sensing, IEEE Trans. Inform. Theory 52 (2006) 1289–1306.
[7] E.D. Gluskin, Norms of random matrices and widths of finite-dimensional sets, Math. USSR Sb. 48 (1984) 173–182.
[8] B. Kashin, The widths of certain finite dimensional sets and classes of smooth functions, Izvestia 41 (1977) 334–351.
[9] G.G. Lorentz, M. von Golitschek, Yu. Makovoz, Constructive Approximation: Advanced Problems, Springer Grundlehren, vol. 304, Springer, Berlin, Heidelberg, 1996.
[10] V. Maiorov, Trigonometric widths of Sobolev classes in the space L_q, Math. Zametki 40 (1986) 161–173.
[11] V. Maiorov, Linear diameters of Sobolev classes, Soviet Dokl. 43 (1991) 1127–1130.