On the size of incoherent systems ⋆

J.L. Nelson [a,*], V.N. Temlyakov [b]

[a] *Department of Mathematics, University of South Carolina, Columbia, South Carolina, 29208, USA*

[b] *Department of Mathematics, University of South Carolina, Columbia, South Carolina, 29208, USA*

Abstract

This paper concerns special redundant systems, namely, incoherent systems or systems with small coherence parameter. Simple greedy-type algorithms perform well on these systems. For example, known results show that the smaller the coherence parameter, the better the performance of the Orthogonal Greedy Algorithm. These systems with small coherence parameter are also useful in the construction of compressed sensing matrices. Therefore, it is very desirable to build dictionaries with small coherence parameter.

We discuss the following problem for both $\mathbb{R}^n$ and $\mathbb{C}^n$: How large can a system with coherence parameter not exceeding a fixed number μ be? We obtain upper and lower bounds for the maximal cardinality of such systems. Although the results herein are either known or simple corollaries of known results, our objective is to demonstrate how fundamental results from different areas of mathematics–linear algebra, probability, and number theory–collaborate on this important approximation theoretic problem.

Key words: coherence, Gramm matrix, compressed sensing

1991 MSC: 41A65, 41A25, 41A46, 46B20

Preprint submitted to Elsevier *5 August 2008*

---

1 Introduction

The past decade has seen great successes in studying nonlinear approximation, motivated by numerous applications (see the surveys [5,14,15]). The fundamental question of nonlinear approximation is how to devise good constructive methods (algorithms) of nonlinear approximation. This problem has two levels of nonlinearity. The first level of nonlinearity is m-term approximation with regard to a basis. In this problem one can use the unique function expansion with regard to a given basis to build an approximant. Nonlinearity enters by looking for m-term approximants with terms (i.e. basis elements in approximant) allowed to depend on a given function. The second level of nonlinearity replaces a basis with a more general system which is not necessarily minimal (for example, a redundant system or a dictionary). This setting is much more complicated than the first one (the basis case), however, the importance of redundant systems arises in both theoretical questions and in practical applications (see for instance [16,11,8]).

In this paper we study special redundant systems, namely, incoherent systems or systems with small coherence parameter. We consider systems which have become popular in signal processing. Let $\mathcal{D} = \{g^k\}_{k=1}^N$ be a normalized ($\|g^k\| =$

[⋆] This research was supported by the National Science Foundation Grant DMS 0554832.

[∗] Corresponding author.
Email addresses: `nelsonjl@math.sc.edu` (J.L. Nelson), `temlyak@math.sc.edu` (V.N. Temlyakov).

1, $k = 1, \dots, N$) system of vectors in $\mathbb{R}^n$ or $\mathbb{C}^n$ equipped with the Euclidean norm. Denote

$$\mu(\mathcal{D}) := \sup_{k \neq l} |\langle g^k, g^l \rangle|,$$

the coherence parameter of a system $\mathcal{D}$. For an orthonormal basis $\mathcal{B}$ we have $\mu(\mathcal{B}) = 0$. It is clear that the smaller the $\mu(\mathcal{D})$ the more the $\mathcal{D}$ resembles an orthonormal system. However, in the case $\mu(\mathcal{D}) > 0$ the $\mathcal{D}$ can be a redundant system. We call $\mathcal{D}$ a dictionary if it spans the whole space. Dictionaries with small coherence parameter $\mu(\mathcal{D})$ are useful in signal processing because simple greedy-type algorithms perform well on them. Denote

$$\sigma_m(f, \mathcal{D}) := \inf_{\{c_k\}, \Lambda : \#\Lambda \le m} \left\| f - \sum_{k \in \Lambda} c_k g^k \right\|,$$

the best m-term approximation of f with regard to $\mathcal{D}$. It was proved in [10] that for the residual f_m^o of the Orthogonal Greedy Algorithm (Orthogonal Matching Pursuit) one has

$$\|f_m^o\| \le 8m^{1/2}\sigma_m(f, \mathcal{D}) \quad \text{for} \quad m < 1/(32\mu(\mathcal{D})). \tag{1}$$

The bound (1) was improved for small m in [9]:

$$\|f_{[m \log m]}^o\| \le 24\sigma_m(f, \mathcal{D}) \quad \text{for} \quad m \le 0.05\mu(\mathcal{D})^{-2/3}. \tag{2}$$

The above results show that the smaller the coherence parameter $\mu(\mathcal{D})$ the better performance of the Orthogonal Greedy Algorithm. Therefore, it is very desirable to build dictionaries with small coherence parameter $\mu(\mathcal{D})$.

It is known that the concept of Restricted Isometry Property (RIP) of the matrix $\Phi := \Phi(\mathcal{D})$ formed by column vectors g^k of a dictionary $\mathcal{D}$ plays an important role in Compressed Sensing (see [3,2,6]). Denote Φ_T a matrix consisting of columns of Φ with indices from T. The RIP has been introduced in [3]: $\delta_S < 1$ is the S-restricted isometry constant of Φ if it is the smallest quantity such that

$$(1-\delta_S)\|c\|_2^2 \le \|\Phi_T c\|_2^2 \le (1+\delta_S)\|c\|_2^2$$

for all subsets T with $|T| \le S$ and all coefficient sequences $\{c_j\}_{j\in T}$. It is known (see, for instance, [9, Lemma 2.1]) that $\Phi(\mathcal{D})$ has RIP for $S \le \delta/\mu(\mathcal{D})$ with the RIP constant δ. Thus, dictionaries $\mathcal{D}$ with small coherence parameter $\mu(\mathcal{D})$ provide compressed sensing matrices $\Phi(\mathcal{D})$ with RIP property (see [7] for detailed discussion).

In this paper we discuss the following problem for both $\mathbb{R}^n$ and $\mathbb{C}^n$. How large can a system $\mathcal{D}$ with $\mu(\mathcal{D}) \le \mu$ be? In other words, we want to estimate the characteristics

$$N(n,\mu) := \sup\{N : \exists \mathcal{D} \quad \text{such that} \quad \#\mathcal{D} \ge N, \mu(\mathcal{D}) \le \mu\}.$$

It is known (see a discussion in Section 2) that for a system $\mathcal{D}$ with $\#\mathcal{D} \ge 2n$ one has $\mu(\mathcal{D}) \ge (2n)^{-1/2}$. Thus, a natural range for μ is $[(2n)^{-1/2}, 1]$. We establish in this paper that $N(n, Cn^{-1/2})$ has polynomial in n growth. In Section 2 we derive from N. Alon's result the bound

$$N(n,\mu) \le \exp(C_1 n\mu^2 \ln(1/\mu)). \tag{3}$$

In Section 3, using probabilistic technique, we complement (3) by the following lower bound (for $\mathbb{R}^n$)

$$N(n,\mu) \ge C_2 \exp(n\mu^2/2).$$

A very interesting and difficult problem is to provide an explicit (deterministic) construction of a large system with small coherence. We address this problem in Section 4. In Section 4 we use Weil's sums to construct a system $\mathcal{D}$ in $\mathbb{C}^n$ with $\mu(\mathcal{D}) \le \mu$ of cardinality of order $\exp(\mu n^{1/2} \ln n)$. We note that B.S. Kashin [12] constructed a system $\mathcal{D}$ in $\mathbb{R}^n$ with $\mu(\mathcal{D}) \le \mu$ of cardinality of order $\exp(C\mu n^{1/2} \ln n)$ using symbols of Legendre. Similar results have been obtained by R. DeVore [7]. He used the finite fields technique.

We note that results of the paper are either known or simple corollaries of known results. The objective of the paper is to demonstrate how fundamental results from different areas of mathematics–linear algebra, probability, number theory–can be used in studying an important problem from approximation theory. Also, we want to attract attention to the problem of finding the right order of growth of the quantity $N(n, \mu)$.

2 Upper bounds

We begin this section with an argument that establishes that for big dictionaries $\mathcal{D}$ ($\#(\mathcal{D}) \ge 2n$) the coherence parameter $\mu(\mathcal{D})$ is always bounded from below by $c_0 n^{-1/2}$, $c_0 > 0$ is an absolute constant. This argument is an elementary linear algebra argument that works for both $\mathbb{R}^n$ and $\mathbb{C}^n$. The argument is well known (see [17] and references therein). We present this proof here for completeness.

Let $\mathcal{D} := \{g^j\}_{j=1}^N$ be a normalized system of vectors in $\mathbb{R}^n$, $g^j = (g_1^j, \dots, g_n^j)^T$. Denote by

$$\Phi := [g^1, \dots, g^N]$$

a $n \times N$ matrix formed by column vectors $\{g^j\}$. Consider the transposed matrix Φ^T that is formed by the row vectors $(g_1^j, \dots, g_n^j)$, $j = 1, \dots, N$, or by the column vectors $h_i := (g_i^1, \dots, g_i^N)^T$, $i = 1, \dots, n$. Then the Gramm matrix G of the system $\{g^j\}_{j=1}^N$ can be written as

$$G = \Phi^T \Phi.$$

It is well known and easy to understand that $\operatorname{rank} G \le n$ (indeed, the columns of G are linear combinations of n columns h_i, $i = 1, \dots, n$). Therefore, the positive (nonnegative) definite symmetric matrix G has at most n nonzero eigenvalues $\lambda_k > 0$. By normalization assumption $\|g^j\| = 1$, $j = 1, \dots, N$, we obtain from the property of traces of matrices that

$$\sum_k \lambda_k = N.$$

By Hilbert-Schmidt theory for the Singular Value Decompositions we get

$$\sum_{i,j=1}^N (\langle g^i, g^j \rangle)^2 = \sum_k \lambda_k^2. \tag{4}$$

By Cauchy inequality

$$\sum_k \lambda_k^2 \ge n^{-1} \left(\sum_k \lambda_k \right)^2 = N^2/n.$$

Therefore, (4) gives

$$N + (N^2 - N)\mu(\mathcal{D})^2 \ge \sum_{i,j=1}^N (\langle g^i, g^j \rangle)^2 \ge N^2/n$$

and

$$\mu(\mathcal{D}) \ge \left(\frac{N-n}{n(N-1)} \right)^{1/2}. \tag{5}$$

In particular, (5) implies for $N \geq 2n$ that $\mu \geq (2n)^{-1/2}$.

The lower bound (5) has been derived from the property $\operatorname{rank} G \leq n$ of the Gramm matrix of the system $\mathcal{D}$. We now use a fundamental result of N. Alon [1] to derive an upper bound for $N(n,\mu)$ from the property $\operatorname{rank} G \leq n$.

Theorem 1 *(N. Alon) Let* $A := \|a_{i,j}\|_{i,j=1}^N$ *be a square matrix of the form* $a_{i,i} = 1$, $i = 1, \dots, N$; $|a_{i,j}| \leq \varepsilon < 1$, $i \neq j$. *Then*

$$\min(N, (\ln N)(\varepsilon^2 \ln(1/\varepsilon))^{-1}) \leq C_1 \operatorname{rank} A \tag{6}$$

with an absolute constant C_1.

We apply this theorem with $A = G$ and $\varepsilon = \mu$. For $N \geq C_2 n \ln(n+1)$ with large enough C_2 the inequality $\mu \geq (2n)^{-1/2}$ implies that

$$N \geq (\ln N)(\varepsilon^2 \ln(1/\varepsilon))^{-1}.$$

Therefore, (6) gives the inequality

$$(\ln N)(\mu^2 \ln(1/\mu))^{-1} \leq C_1 n$$

and

$$N \leq \exp(C_1 n \mu^2 \ln(1/\mu)). \tag{7}$$

In particular, in the case $\mu = C_3 n^{-1/2}$ inequality (7) gives the polynomial bound $N \leq n^{C_4}$.

3 Lower bounds via Probabilistic approach

In this section we prove existence of large systems with small coherence. The proof is based on Hoeffding's inequality.

Theorem 2 *(Hoeffding's inequality) Let* ξ_i *be real random variables on* (X,Σ,ρ) *such that* $|\xi_i - E\xi_i| \le b_i$, $i=1,\dots,m$, *almost surely. Consider a new random variable* ζ *on* (X^m,Σ^m,ρ^m) *defined as*

$$\zeta(\omega) := \sum_{i=1}^{m} \xi_i(\omega_i), \quad \omega = (\omega_1,\dots,\omega_m).$$

Then for $t>0$

$$\rho^m\{\omega : |\zeta(\omega) - E\zeta| \ge mt\} \le 2\exp\left(-\frac{m^2t^2}{2\|b\|_2^2}\right). \tag{8}$$

Theorem 2 implies the following inequality in the case of complex random variables. If $z = a+ib$ then $|z|\ge t$ implies that either $|a| \ge 2^{-1/2}t$ or $|b| \ge 2^{-1/2}t$. Therefore, in the complex case we have the following inequality instead of (8)

$$\rho^m\{\omega : |\zeta(\omega) - E\zeta| \ge mt\} \le 4\exp\left(-\frac{m^2t^2}{4\|b\|_2^2}\right). \tag{9}$$

Let us begin with a construction in the $\mathbb{C}^n$. Consider random variables $\xi_k = e^{i2\pi kx}$, $x\in[0,1]$, $k=-N,\dots,-1,1,\dots,N$. Let ρ be the Lebesgue measure on $[0,1]$. Then by (9) with $t=\mu$ we get for each k

$$\rho^n\{(x_1,\dots,x_n) : \left|n^{-1}\sum_{j=1}^{n} e^{i2\pi kx_j}\right| > \mu\} \le 4\exp(-n\mu^2/4). \tag{10}$$

Therefore, for any N satisfying

$$N < 8^{-1} \exp(n\mu^2/4)$$

there exists a set of points $y_1, \dots, y_n$ such that for all $|k| \le N$, $k \neq 0$ we have

$$\left| n^{-1} \sum_{j=1}^{n} e^{i2\pi k y_j} \right| \le \mu.$$

We now define a system $\mathcal{D}_N = \{g^l\}_{l=1}^N$ by

$$g^l := n^{-1/2} (e^{i2\pi l y_1}, \dots, e^{i2\pi l y_n})^T.$$

It is a normalized system with the property

$$|\langle g^l, g^m \rangle| = \left| n^{-1} \sum_{j=1}^{n} e^{i2\pi (l-m) y_j} \right| \le \mu$$

provided $l \neq m$. Thus, we have built a normalized system of N vectors with coherence $\le \mu$ with N of the order of $\exp(n\mu^2/4)$ that is very close to the corresponding upper bound (see (7)) $N \le \exp(C_1 n\mu^2 \ln(1/\mu))$.

We proceed to a construction in $\mathbb{R}^n$. This construction is similar to the above one with the exponential functions $e^{i2\pi kx}$ replaced by the Walsh functions $w_k(x)$. We recall the definition of the Walsh system of functions (see, for instance, [13]). Let

$$r_k(x) := \operatorname{sign} \sin(2^k \pi x), \quad x \in [0,1], \quad k = 1, 2, \dots,$$

be the Rademacher system. We define $w_m(x)$ for $m = 1, 2, \dots$ in the following way. Let

$$m = \sum_{j=0}^{l} a_j 2^j, \quad a_j = 0, 1, \quad j = 0, 1, \dots, l.$$

Denote $J_m := \{j : a_j = 1\}$. Define

$$w_m(x) := \prod_{j \in J_m} r_{j+1}(x).$$

Then $\{w_m(x)\}_{m=1}^{\infty}$ forms an orthonormal system on $[0,1]$. It has the following property convenient for us. Let

$$m = \sum_{j=0}^{l} a_j 2^j, \quad k = \sum_{j=0}^{l} b_j 2^j, \quad s(m,k) := \sum_{j=0}^{l} |a_j - b_j| 2^j.$$

Then for all x, except maybe dyadic rationals, one has

$$w_m(x) w_k(x) = w_{s(m,k)}(x).$$

It is clear that $s(m,k) \le m + k$.

Consider real random variables $\xi_k = w_k(x)$, $x \in [0,1]$, $k = 1, \dots, 2N$. Let ρ be the Lebesgue measure on $[0,1]$. Then by Hoeffding's inequality, (8) with $t = \mu$, we get for each $k = 1, \dots, 2N$,

$$\rho^n\{(x_1, \dots, x_n) : \left| n^{-1} \sum_{i=1}^{n} w_k(x_i) \right| > \mu\} \le 2\exp(-n\mu^2/2).$$

Therefore, for any $N < 4^{-1}\exp(n\mu^2/2)$ there exists a set of points $y_1, \dots, y_n$ that are not dyadic rationals such that for all $k \in [1, 2N]$ we have

$$\left| n^{-1} \sum_{i=1}^{n} w_k(x_i) \right| \le \mu.$$

Consider the following system $W_N = \{g^l\}_{l=1}^{N}$ in $\mathbb{R}^n$

$$g^l := n^{-1/2}(w_l(y_1), \dots, w_l(y_n))^T.$$

It is a normalized system satisfying for $l \neq m$

$$|\langle g^l, g^m\rangle| = \left|n^{-1}\sum_{i=1}^{n} w_l(y_i)w_m(y_i)\right| = \left|n^{-1}\sum_{i=1}^{n} w_{s(l,m)}(y_i)\right| \leq \mu.$$

Thus, the above system W_N with $N = [4^{-1}\exp(n\mu^2/2)] - 1$ provides an example of a large system with coherence parameter $\leq \mu$.

4 Lower bounds. Deterministic construction

In this section we present a deterministic construction of large systems with small coherence. The construction is based on the following variant of the A. Weil theorem (see [4]).

Theorem 3 *Let $r \geq 2$ be a natural number and let $p > r$ be a prime number. Denote for $a := (a_1, \dots, a_r)$, a_j are integers,*

$$F(a, u) := a_r u^r + \ldots + a_1 u.$$

Then for $a \neq (0, \dots, 0)$ mod p one has

$$|S(a)| \leq (r-1)p^{1/2}, \quad S(a) := \sum_{u=1}^{p} e^{2\pi i F(a,u)/p}. \tag{11}$$

Clearly, in a particular case $r = 2$ the inequality (11) gives the classical result for the magnitude of Gaussian sums. Consider the following set $W(r, p)$ of vectors in $\mathbb{C}^p$:

$$v^a := p^{-1/2}(e^{2\pi i F(a,1)/p}, \dots, e^{2\pi i F(a,p)/p})^T$$

for $a_j \in [1,p]$, $j = 1,\ldots,r$. It is clear that this is a set of normalized vectors. The size of this set is p^r. Now, it remains to find the magnitude of the coherence parameter. If we consider v^a and $v^{a'}$ where $a \neq a'$, we may use the above Theorem 3 to bound the required inner product directly:

$$\begin{aligned} |\langle v^a, v^{a'}\rangle| &= p^{-1}\left|\sum_{u=1}^{p} e^{2\pi i(F(a,u)-F(a',u))/p}\right| \\ &= p^{-1}\left|\sum_{u=1}^{p} e^{2\pi i F(a-a',u)/p}\right| \le (r-1)p^{-1/2}. \end{aligned} \tag{12}$$

For given n and $\mu \ge (2/n)^{1/2}$ we set p to be the biggest prime not exceeding n. Then $n/2 \le p \le n$. We specify r to be the biggest natural number such that $(r-1)p^{-1/2} \le \mu$. Then by (12) $\mu(W(r,p)) \le \mu$. For the cardinality of the $W(r,p)$ we have

$$|W(r,p)| = p^r = e^{r\ln p} \ge e^{\mu p^{1/2}\ln p}.$$

Acknowledgements

The authors are thankful to professors B. Kashin and K. Oskolkov for useful discussions.

References

[1] N. Alon, Problems and results in extremal combinatorics, I, Discrete Math. 273 (2003) 31-53.

[2] E. Candes, Compressive sampling, ICM Proceedings, Madrid 3 (2006) 1433–1452.

[3] E. Candes and T. Tao, Decoding by linear programming, IEEE Trans. Inform. Theory 51 (2005) 4203–4215.

[4] L. Carlitz and S. Uchiyama, Bounds for exponential sums, Duke Math. J. 24 (1957) 37–41.

[5] R.A. DeVore, Nonlinear approximation, Acta Numerica (1998) 51–150.

[6] R.A. DeVore, Optimal computation, ICM Proceedings, Madrid 1 (2006) 187–215.

[7] R.A. DeVore, Deterministic Constructions of Compressed Sensing Matrices, Manuscript (2007) 1–9.

[8] D.L. Donoho, Sparse Components of Images and Optimal Atomic Decompositions, Constr. Approx. 17 (2001) 353–382.

[9] D. Donoho, M. Elad, and V.N. Temlyakov, On the Lebesgue type inequalities for greedy approximation, J. Approximation Theory 147 (2007) 185–195.

[10] A.C. Gilbert, S. Muthukrishnan, and M.J. Strauss, Approximation of functions over redundant dictionaries using coherence, The 14th Annual ACM-SIAM Symposium on Discrete Algorithms (2003).

[11] P.J. Huber, Projection Pursuit, The Annals of Statistics 13 (1985) 435–475.

[12] B.S. Kashin, On widths of octahedrons, Uspekhi Matem. Nauk 30 (1975) 251–252 (in Russian).

[13] B.S. Kashin and A.A. Saakyan, Orthogonal Series, American Math. Soc., Providence, R.I.(1989).

[14] V.N. Temlyakov, Nonlinear methods of approximation, Found. Comput. Math. 3 (2003) 33–107.

[15] V.N. Temlyakov, Greedy Approximation, Acta Numerica (2008).

[16] E. Schmidt, Zur Theorie der linearen und nichtlinearen Integralgleichungen. I, Math. Annalen 63 (1906-1907) 433–476.

[17] T. Stromberg and R. Heath Jr., Grassmannian frames with applications to coding and communications, Appl. Comput. Harm. Anal. 14 (2003) 257–275.