On the Impossibility of Uniform Sparse Reconstruction using Greedy Methods

Holger Rauhut*

August 9, 2007

Abstract

It has previously shown that a trigonometric polynomial having at most M non-vanishing coefficients can be recovered from $N = \mathcal{O}(M \log(D))$ random samples by the greedy methods thresholding and orthogonal matching pursuit with high probability. In this note we show that these results cannot be made uniform in the sense that a *single* (random) sampling set cannot guarantee recovery of *all* such M-sparse trigonometric polynomials simultaneously with high probability using the two greedy methods.

Key Words: random sampling, trigonometric polynomials, greedy algorithms, orthogonal matching pursuit, thresholding, sparse recovery, compressed sensing.
AMS Subject classification: 94A20, 42A05.

1 Introduction

The basic goal of the sparse recovery (compressed sensing) problem is to exactly reconstruct a sparse signal $c \in \mathbb{C}^D$ with at most M nonzero components, $M \ll D$, from

$$N = \mathcal{O}(M \log^n(D)) \tag{1.1}$$

nonadaptive linear measurements of c [7, 2, 13]. These measurements are given by the vector $\Phi c \in \mathbb{C}^N$, where Φ is an $N \times D$ matrix.

Basically two approaches for the reconstruction techniques have been proposed: ℓ_1-minimization (Basis Pursuit) [2, 4, 5, 7] and greedy methods such as simple thresholding and orthogonal matching pursuit (OMP) [10, 16]. Both types of methods are able to reconstruct a sufficiently sparse signal c exactly with high probability if the measurement matrix Φ is a random Gaussian or Bernoulli matrix [7, 4, 1, 9]. Moreover, if Φ is partial random Fourier matrix then there are rigorous results of the same type for ℓ_1-minimization and thresholding, while for OMP the claim is supported by partial theoretical results and vast numerical experiments in [10, 12].

Despite recent progress on efficient solvers for ℓ_1-minimization, usually greedy algorithms are still considered faster than Basis Pursuit. In particular, it is hard to beat simple thresholding in terms of computation speed. However, ℓ_1-minimization has the

*Numerical Harmonic Analysis Group, Department of Mathematics, University of Vienna, Nordbergstrasse 15, A-1090 Vienna. E-mail: `holger.rauhut@univie.ac.at`.
H. R. acknowledges financial support by the European Union's Human Potential Programme under contract MEIF-CT 2006-022811.

advantage that recovery holds uniform in the sense that a single (random) measurement matrix can guarantee exact reconstruction simultaneously of *all* sparse signals (in the range of (1.1)) [3, 7, 12, 14]. In contrast, it is known that for the Gaussian ensemble and in the range of (1.1) thresholding and OMP cannot recover *all* signals – even supported on one fixed set T – with high probability using the same matrix [8], and this phenomenon easily extends to the Bernoulli ensemble. In this note we show that also for partial Fourier matrices thresholding and OMP do not guarantee uniform recovery if N is only linear in the sparsity M (i.e. in the range of (1.1)). More precisely, if $N \le CM^2$ then there exists an M-sparse signal depending on the (randomly chosen) matrix Φ such that exact reconstruction fails for thresholding with high probability, while the corresponding statement for OMP holds under the condition $N \le CM^{3/2}$. We note, however, that a recent variant called regularized orthogonal matching pursuit does ensure uniform recovery [11].

Acknowledgement. I wish to express my gratitude to Roman Vershynin who suggested the counterexample worked out in this note.

2 Notation and Previous results

For some finite subset $\Gamma \subset \mathbb{Z}^d$, $d \in \mathbb{N}$, we let Π_Γ denote the space of all trigonometric polynomials in dimension d whose coefficients are supported on Γ. An element f of Π_Γ is of the form $f(x) = \sum_{k\in\Gamma} c_k e^{2\pi i k\cdot x}$, $x \in [0,1]^d$, with Fourier coefficients $c_k \in \mathbb{C}$. The dimension of Π_Γ will be denoted by $D := |\Gamma|$.

A trigonometric polynomial is called M-sparse if at most M coefficients c_k are non-zero and the set of all M-sparse trigonometric polynomials in Π_Γ is denoted by $\Pi_\Gamma(M)$. The goal is to reconstruct such a sparse trigonometric polynomial $f \in \Pi_\Gamma(M)$ from sample values $f(x_1), \ldots, f(x_N)$, where the number N of sampling points $x_1, \ldots, x_N \in [0, 2\pi]^d$ is small compared to the dimension D.

Given the sampling set $X = (x_1, \ldots, x_N)$ we denote by $\mathcal{F}_X$ the $N \times D$ matrix (recall that $D = |\Gamma|$) with entries

$$(\mathcal{F}_X)_{j,k} \;=\; e^{2\pi i k\cdot x_j}, \quad 1 \le j \le N,\, k \in \Gamma. \tag{2.1}$$

Then clearly $f(x_j) = (\mathcal{F}_X c)_j$ if c is the vector of Fourier coefficients of f. Let ϕ_k denote the k-th column of $\mathcal{F}_X$, i.e.,

$$\phi_k \;=\; \begin{pmatrix} e^{2\pi i k\cdot x_1} \\ \vdots \\ e^{2\pi i k\cdot x_N} \end{pmatrix},$$

then $\mathcal{F}_X = (\phi_{k_1}|\phi_{k_2}|\ldots|\phi_{k_D})$. By $\mathcal{F}_{TX}$ we denote the restriction of $\mathcal{F}_X$ to the columns indexed by a subset $T \subset \Gamma$.

We use the following two probability models for $x_1, \ldots, x_N$:

(1) The sampling points $x_1, \ldots, x_N$ are independent and uniformly distributed random variable on the cube $[0,1]^d$.

(2) The sampling points $x_1, \ldots, x_N$ are independent and uniformly distributed on the grid $\frac{1}{m}\mathbb{Z}_m^d = \{0, \frac{1}{m}, \ldots, \frac{m-1}{m}\}^d$ for some $m \in \mathbb{N}, m \ge 2$.

In both cases the matrix $\mathcal{F}_X$ is a random partial Fourier matrix. More precisely, in the first case it is a non-equispaced Fourier matrix, and in the second case it is a submatrix consisting of random rows of the discrete Fourier matrix on $\mathbb{Z}_m^d$.

So far mainly ℓ_1-minimization (Basis Pursuit) [2, 4, 5, 8, 13, 12] and greedy algorithms [9, 10, 12] were proposed as methods for reconstructing the vector c of Fourier coefficients of a sparse trigonometric polynomial and hence the polynomial itself. Denote $y = (f(x_j))_{j=1}^N$ the vector of sample values. Basis Pursuit (BP) consists in determining the solution of the minimization problem

$$\min \|c\|_1 = \sum_{k \in \Gamma} |c_k| \quad \text{subject to} \quad \mathcal{F}_X c = y = (f(x_j))_{j=1}^N.$$

This problem can be solved with convex optimization techniques. The following reconstruction theorem has been shown recently [12, 14, 15, 4].

Theorem 2.1. *Let the sampling set $X = (x_1, \ldots, x_N)$ be chosen according to the probability model (1) or (2), and suppose*

$$\frac{N}{\log(N)} \geq CM \log^2(M) \log(D) \log(\epsilon^{-1}). \tag{2.2}$$

Then with probability at least $1 - \epsilon$ Basis Pursuit reconstructs every M-sparse trigonometric polynomial.

Since $M, N \leq D$, relation (2.2) is satisfied in particular if $N \geq CM \log^4(D) \log(\epsilon^{-1})$. Neglecting the log-factors the required number of samples N scales linear in the sparsity M.

Presumably the simplest greedy algorithm is thresholding. It consists of the following steps:

1. Determine the set $T \subset \Gamma$ corresponding to the M largest correlations $|\langle y, \phi_k \rangle|, k \in \Gamma$.
2. Compute the coefficients corresponding to the orthogonal projection of y onto the linear span of $\{\phi_k, k \in T\}$, i.e., the non-zero coefficients of c are determined as the minimizer of the least spares problem $\min \|\mathcal{F}_{TX} c - y\|_2$.

Thresholding is much faster in practice than Basis Pursuit. In [10] the following reconstruction theorems have been shown.

Theorem 2.2. *Let $f \in \Pi_\Gamma(M)$ with Fourier coefficients c. Define its dynamic range by*

$$R := \frac{\max_{k \in T} |c_k|}{\min_{k \in T} |c_k|}.$$

Choose the sampling points $x_1, \ldots, x_N$ according to the probability model (1) or (2). If for some $\epsilon > 0$

$$N \geq CMR^2 \log(4D/\epsilon) \tag{2.3}$$

then with probability at least $1 - \epsilon$ thresholding recovers f exactly. The constant C is no larger than 17.89.

Apart from the dependence on the dynamic range R there is a subtle difference to the recovery theorem for BP above. Recovery holds only for the given sparse polynomial with high probability, while Theorem 2.1 states that a single sampling set can recover *all* sparse polynomials with high probability. One can remove this drawback for thresholding; however, at the cost of dramatically increasing the required number of samples:

Theorem 2.3. *Choose the sampling set $X = (x_1, \ldots, x_N)$ according to the probability models (1) or (2), and assume that*

$$N \geq CM^2 R^2 \log(D/\epsilon).$$

Then with probability at least $1 - \epsilon$ thresholding recovers all M-sparse trigonometric polynomials for which the dynamic range of their coefficients is at most R.

Explicit (and small) constants can be found in [10]. Below we will show that the quadratic dependence of the number of samples N on the sparsity M cannot be improved if one requires uniformity, i.e., recovery of all sparse trigonometric polynomials from a single sampling set.

Orthogonal Matching Pursuit (OMP) is an iterative greedy algorithm, which adds a new element of the support T in each step by maximizing the correlation of the current residual with the remaining columns ϕ_k. Formally, it consists of the following steps.

- Initialize: Set current residual $r_0 := y$ and support set $T_0 := \emptyset$.
- Iterate until a stopping criterion is met (iteration counter s):
  - Determine $k_s := \text{argmax}_{k \in T} |\langle r_{s-1}, \phi_k \rangle|$ and set $T_s := T_{s-1} \cup \{k_s\}$.
  - Update the residual by $r_s := y - P_{T_s} y$ where P_{T_s} denotes the orthogonal projection of y onto the span of $\{\phi_k, k \in T_s\}$.
- Set $T = T_s$; the non-zero coefficients of c are given by $P_T y = \sum_{k \in T} c_k \phi_k$.

More details on the implementation of this algorithm can be found in [10]. In practice, it is slower than thresholding but usually faster than BP. Moreover, numerical tests indicate a higher recovery rate than thresholding and a similar (average) rate as BP (despite the theoretical differences this paper is concerned with).

Due to stochastic dependency issues it seems difficult to analyze fully the performance of OMP, but at least we can say something about the first step [10].

Theorem 2.4. *Let $f \in \Pi_\Gamma(M)$ with coefficients supported on T. Choose random sampling points $x_1, \ldots, x_N$ according to one of our two probability models. If*

$$N \geq CM \log(D/\epsilon)$$

then with probability at least $1 - \epsilon$ OMP selects an element of the true support T in the first iteration.

Compared to the corresponding result for thresholding (Theorem 2.2) the dependence on the dynamic range is removed, and numerical experiments in [10] indicate that this remains true also for the further iterations. However, again the above statement is non-uniform, in the sense that a single sampling set X is good only for the given polynomial but not necessarily for all sparse polynomials simultaneously. As for thresholding one can remove this drawback at the cost of dramatically increasing the required number of sampling points, and in this case we have a statement for the full application of OMP.

Theorem 2.5. *Let $X = (x_1, \ldots, x_N)$ be chosen according to the probability model (1) or (2). Assume that*

$$N \geq CM^2 \log(D/\epsilon).$$

Then with probability at least $1 - \epsilon$ OMP recovers every M-sparse trigonometric polynomial in M steps.

We will show below that the above statement is close to optimal if one requires uniformity. The number of required samples N increases at least faster than $M^{3/2}$, so Theorem 2.4 cannot be made uniform.

3 Main Results

The following result shows that Theorem 2.3 cannot be significantly improved if one requires uniformity. For simplicity we restrict Γ to a particular set of basic frequencies, although the statement holds also for more general sets (see also Remark 4).

Theorem 3.1. *Let $\Gamma = \mathbb{Z}_m^d$ with $m \geq 2$. Let $X = (x_1, \ldots, x_N)$ be randomly chosen according to the probability model (1) or (2). Suppose $M \leq |\Gamma|/4 = m^d/4$ and*

$$N \leq \frac{1}{4\sigma} M^2 - \frac{7}{2}(M-1) \tag{3.1}$$

for some $\sigma > 2$. Then with probability exceeding

$$1 - \frac{4}{M} - \frac{1}{(\sigma-1)^2}$$

there exists an M-sparse trigonometric polynomial (depending on X) which thresholding fails to reconstruct.

The proof actually provides an explicit trigonometric polynomial (depending on X), which thresholding fails to reconstruct with the stated probability.

A similar statement holds also for OMP.

Theorem 3.2. *Let $\Gamma = \mathbb{Z}_m^d$ with $m \geq 2$. Let $X = (x_1, \ldots, x_N)$ be randomly chosen according to the probability model (1) or (2). Suppose $M \leq |\Gamma|/4 = m^d/4$ and*

$$N \leq \frac{\tau}{5} M^{3/2} - \frac{7}{2}(M-1) \tag{3.2}$$

for some $\tau < 1$. Then with probability exceeding $1 - \frac{4}{M} - \tau^2$ there exists an M-sparse trigonometric polynomial (depending on X) which OMP fails to reconstruct in M steps.

The exponent 3/2 in (3.2) is probably not optimal, and we expect that (3.2) can be improved to $N \leq C_n M^{2-1/n}$ for any $n \geq 2$ with a constant C_n depending on n. However, this seems to make the already technical proof even more tedious. Since the present statement already shows that the recovery Theorem 2.4 for OMP cannot be made uniform without spoiling the linear dependence $N = \mathcal{O}(M)$, we did not further pursue this issue here.

We note that the above theorem does not exclude the possibility that the pathological M-sparse trigonometric polynomial is reconstructed after $M+1$ or more steps. The proof of the theorem actually shows that OMP selects a wrong element in the first step. However, if the reconstructed support set T' contains the true support set T (which is possible if OMP does more than M steps) and if T' is not unreasonably large (say $|T'| \leq 2|T|$) then it will usually happen that the coefficients on $T' \setminus T$ will be set to zero so that nevertheless the correct polynomial is recovered. This scenario was actually observed in numerical experiments. Nevertheless we expect that a version of Theorem 2.4 holds, which shows an impossibility of uniform recovery even when OMP is allowed to perform more than M steps.

4 Proofs

We develop the proofs of both Theorems 2.2 and 2.4 in parallel.

First note that by linear algebra necessarily $N \geq 2M$ if some method (in particular OMP and thresholding) is able to recover all M-sparse trigonometric polynomials, see also [6, Lemma 3.1]. Indeed, if $N < 2M$ and $|T| = 2M$ then $\mathcal{F}_{TX}$ is not invertible, i.e., there exists c supported on T with $\mathcal{F}_{TX} c = 0$. Split T into T_1 and T_2 with $|T_1| = |T_2| = M$ and $T = T_1 \cup T_2$. Denote by c_i the vector that coincides with c on T_i and is zero outside ($i = 1, 2$), i.e., both c_1 and c_2 are M-sparse. Then $c = c_1 + c_2$, but $\mathcal{F}_X c = \mathcal{F}_X(c_1 + c_2) = 0$, i.e., $F_X(c_1) = \mathcal{F}_X(-c_2)$. Hence, both c_1 and $-c_2$ provide the same measurements and no reconstruction method can distinguish between both M-sparse vectors. Hence, from now on we assume $N \geq 2M$ in addition to (3.1) or (3.2).

Both for Thresholding and OMP we choose the M-sparse trigonometric polynomial as follows. Let $T \subset \Gamma$ of size M be the support of its coefficients; and let $\ell \in \Gamma \setminus T$. Later we give some restrictions on T and ℓ, which are not essential however. Then we choose the non-zero components of c as

$$c_k := \langle \phi_\ell, \phi_k \rangle, \quad k \in T.$$

Clearly, c_k depends on the sampling set X. Let

$$y = \mathcal{F}_X c = \sum_{k \in T} \langle \phi_\ell, \phi_k \rangle \phi_k \tag{4.1}$$

be the corresponding vector of sample values.

Thresholding fails to select the correct support T if

$$|\langle y, \phi_\ell \rangle| > \min_{k \in T} |\langle y, \phi_k \rangle|.$$

Hence, fixing some $k \in T$ and $\alpha > 0$ the probability that thresholding succeeds can be estimated from above by

$$\mathbb{P}(|\langle y, \phi_k \rangle| > |\langle y, \phi_\ell \rangle|) \leq \mathbb{P}(|\langle y, \phi_k \rangle| > \alpha) + \mathbb{P}(|\langle y, \phi_\ell \rangle| < \alpha).$$

Now we consider OMP. It selects a wrong element in the first step, and consequently cannot recover c in M steps, if

$$|\langle y, \phi_\ell \rangle| > \max_{k \in T} |\langle y, \phi_k \rangle|.$$

Similarly as above, the probability that OMP does not fail in the first step can be upper bounded by

$$\begin{aligned} \mathbb{P}(\max_{k \in T} |\langle y, \phi_k \rangle| > |\langle y, \phi_\ell \rangle|) &\leq \mathbb{P}(\max_{k \in T} |\langle y, \phi_k \rangle| > \alpha) + \mathbb{P}(|\langle y, \phi_\ell \rangle| < \alpha) \\ &\leq \sum_{k \in T} \mathbb{P}(|\langle y, \phi_k \rangle| > \alpha) + \mathbb{P}(|\langle y, \phi_\ell \rangle| < \alpha). \end{aligned} \tag{4.2}$$

Hence, both OMP and thresholding require an analysis of $\mathbb{P}(|\langle y, \phi_\ell \rangle| < \alpha)$ and $\mathbb{P}(|\langle y, \phi_k \rangle| > \alpha)$. Assuming $\alpha \leq \frac{1}{2}\mathbb{E}|\langle y, \phi_\ell \rangle|$ we can estimate the first term as

$$\begin{aligned} \mathbb{P}(|\langle y, \phi_\ell \rangle| < \alpha) &= \mathbb{P}(\mathbb{E}|\langle y, \phi_\ell \rangle| - |\langle y, \phi_\ell \rangle| > \mathbb{E}|\langle y, \phi_\ell \rangle| - \alpha) \\ &\leq \frac{\mathbb{E}\left[\mathbb{E}|\langle y, \phi_\ell \rangle| - |\langle y, \phi_\ell \rangle|\right]^2}{(\mathbb{E}|\langle y, \phi_\ell \rangle| - \alpha)^2} \leq 4 \frac{\mathbb{E}\left[|\langle y, \phi_\ell \rangle|^2\right] - (\mathbb{E}|\langle y, \phi_\ell \rangle|)^2}{(\mathbb{E}|\langle y, \phi_\ell|)^2}. \end{aligned} \tag{4.3}$$

Similarly, assuming $\alpha^2 \geq \sigma \mathbb{E}[|\langle y, \phi_k \rangle|^2]$ for $\sigma > 1$ we can estimate

$$\begin{aligned} \mathbb{P}(|\langle y, \phi_k \rangle| > \alpha) &= \mathbb{P}(|\langle y, \phi_k \rangle|^2 - \mathbb{E}[|\langle y, \phi_k \rangle|^2] > \alpha^2 - \mathbb{E}[|\langle y, \phi_k \rangle|^2]) \\ &\leq \frac{\mathbb{E}\left[|\langle y, \phi_k \rangle|^2 - \mathbb{E}[|\langle y, \phi_k \rangle|^2]\right]^2}{(\alpha^2 - \mathbb{E}[|\langle y, \phi_k \rangle|^2])^2} \\ &\leq \frac{\mathbb{E}[|\langle y, \phi_k \rangle|^4] - (\mathbb{E}[|\langle y, \phi_k \rangle|^2])^2}{(\sigma - 1)^2 (\mathbb{E}[|\langle y, \phi_k \rangle|^2])^2}. \end{aligned} \tag{4.4}$$

An α satisfying the two assumed conditions exists if and only if

$$(\mathbb{E}|\langle y, \phi_\ell \rangle|)^2 \geq 4\sigma \mathbb{E}[|\langle y, \phi_k \rangle|^2].$$

It remains to compute the expectations $\mathbb{E}|\langle y, \phi_\ell \rangle|$, $\mathbb{E}[|\langle y, \phi_\ell \rangle|^2]$, $\mathbb{E}[|\langle y, \phi_k \rangle|^2]$ and $\mathbb{E}[|\langle y, \phi_k \rangle|^4]$.

Lemma 4.1. *For $\ell \notin T$ and y given by (4.1) it holds $\mathbb{E}|\langle y, \phi_\ell\rangle| = MN$.*

Proof. We have

$$\mathbb{E}|\langle y, \phi_\ell\rangle| = \mathbb{E}\sum_{k\in T} |\langle \phi_\ell, \phi_k\rangle|^2 = \sum_{k\in T} \mathbb{E}|\langle \phi_\ell, \phi_k\rangle|^2.$$

Further,

$$\begin{aligned}\mathbb{E}|\langle \phi_\ell, \phi_k\rangle|^2 &= \mathbb{E}\sum_{j=1}^{N}\sum_{j'=1}^{N} e^{2\pi i(\ell-k)\cdot x_j} e^{-2\pi i(\ell-k)x_{j'}} \\ &= \sum_{j,j'=1}^{N} \mathbb{E}[\exp(2\pi i(\ell-k)\cdot(x_j - x_{j'}))]\end{aligned}$$

Since the x_j are independent and uniformly distributed on $[0,1]^d$ or on the grid $\frac{1}{m}\mathbb{Z}_m^d$, and since $\ell \neq k$ we have

$$\mathbb{E}[\exp(2\pi i(\ell-k)\cdot(x_j - x_{j'}))] = \delta_{j,j'}.$$

Hence, $\mathbb{E}|\langle \phi_\ell, \phi_k\rangle|^2 = N$ and $\mathbb{E}[|\langle y, \phi_k\rangle|] = MN$. ∎

Lemma 4.2. *It holds $\mathbb{E}[|\langle y, \phi_\ell\rangle|^2] = N^2M(M+1) - MN$.*

Proof. We have

$$\mathbb{E}[|\langle y, \phi_\ell\rangle|^2] = \mathbb{E}\left[\sum_{k\in T} |\langle \phi_\ell, \phi_k\rangle|^2\right]^2 = \sum_{k,k'\in T} \mathbb{E}\left[|\langle \phi_\ell, \phi_k\rangle|^2 |\langle \phi_\ell, \phi_{k'}\rangle|^2\right].$$

Furthermore,

$$\mathbb{E}\left[|\langle \phi_\ell, \phi_k\rangle|^2 |\langle \phi_\ell, \phi_{k'}\rangle|^2\right] = \sum_{j_1,j_2,j_3,j_4=1}^{N} \mathbb{E}\left[e^{2\pi i((\ell-k)\cdot(x_{j_1}-x_{j_2}) + (\ell-k')\cdot(x_{j_3}-x_{j_4}))}\right].$$

If $k = k'$ then the expectation in the above sum equals 1 if and only if $\{j_1, j_3\} = \{j_2, j_4\}$, and vanishes otherwise. Hence, for $j_1, j_2, j_3, j_4 \in \{1, \ldots, N\}$ this happens $N + 2N(N-1) = N(2N-1)$ times. If $k \neq k'$ then the expectation equals 1 if and only if $j_1 = j_2$ and $j_3 = j_4$ and vanishes otherwise. This happens N^2 times for $j_1, \ldots, j_4 \in \{1, \ldots, N\}$. Combining everything we obtain

$$\mathbb{E}[|\langle y, \phi_\ell|^2] = MN(2N-1) + M(M-1)N^2 = N^2M(M+1) - MN.$$

∎

Lemma 4.3. *For $k_0 \in T$ and y given by (4.1) we have*

$$\mathbb{E}[|\langle y, \phi_{k_0}\rangle|^2] = (M-1)(M-2)N + 3(M-1)N^2 + N^3.$$

Proof. By definition of y

$$\begin{aligned}\mathbb{E}[|\langle y, \phi_{k_0}\rangle|^2] &= \mathbb{E}\left[\Big|\sum_{k\in T}\langle \phi_\ell, \phi_k\rangle\langle \phi_k, \phi_{k_0}\rangle\Big|^2\right] \\ &= \sum_{k,k'\in T} \mathbb{E}\left[\langle \phi_\ell, \phi_k\rangle\langle \phi_k, \phi_{k_0}\rangle\langle \phi_{k'}, \phi_\ell\rangle\langle \phi_{k_0}, \phi_{k'}\rangle\right].\end{aligned}$$

The expectation in the previous expression equals

$$\begin{aligned}
&\mathbb{E}\left[\langle \phi_\ell, \phi_k\rangle \langle \phi_k, \phi_{k_0}\rangle \langle \phi_{k'}, \phi_\ell\rangle \langle \phi_{k_0}, \phi_{k'}\rangle\right] \\
&= \sum_{j_1,j_2,j_3,j_4=1}^{N} \mathbb{E}\left[e^{2\pi i[(\ell-k)x_{j_1}+(k-k_0)\cdot x_{j_2}-(\ell-k')\cdot x_{j_3}-(k'-k_0)\cdot x_{j_4}]}\right].
\end{aligned}$$

If $k = k'$ then the expectation in the above sum equals 1 if and only if $j_1 = j_3$ and either $k_0 = k = k'$ or $j_2 = j_4$. Hence, if $k_0 = k = k'$ this leaves N^3 possibilities for $j_1, \ldots, j_4 \in \{1, \ldots, N\}$, while for $k_0 \neq k$ there are N^2 possibilities. Further, if $k \neq k'$ then the expectation equals 1 if and only if $j_1 = j_2 = j_3 = j_4$ (giving N possibilities) or $k' = k_0$ and $j_1 = j_2 = j_3$ (giving $(M-1)N^2$ possibilities for k and $j_1, \ldots, j_4$) or $k = k_0$ and $j_1 = j_3 = j_4$ (resulting in $(M-1)N^2$ possibilities for k' and $j_1, \ldots, j_4$). Altogether we obtain the stated expression for $\mathbb{E}[|\langle y, \phi_{k_0}\rangle|^2]$. ∎

Lemma 4.4. *Suppose $\ell_n \equiv 1 \mod 4$ for some component ℓ_n of $\ell \in \mathbb{Z}^d$ and $k \equiv 0 \mod 4$ for all $k \in T$. Then for $k_0 \in T$ and y given by (4.1) it holds*

$$\begin{aligned}
&\mathbb{E}[|\langle y, \phi_{k_0}\rangle|^4] \\
&\leq NM^4 + N^2(2M^4 + 41M^3) + N^3(12M^3 + 94M^2 + 20M) \\
&+ N^4(14M^2 + 36M + 6) + N^5(4M + 25) + 2N^6.
\end{aligned}$$

The proof of this lemma is elementary but quite tedious, and postponed to the next section.

Equipped with these auxiliary results we turn to the proofs of our main results. Choose T and ℓ as in Lemma 4.4. Since $M \leq |\Gamma|/4$ this is certainly possible for our choice of Γ. By Lemma 4.3

$$\begin{aligned}
\mathbb{E}[|\langle y, \phi_k\rangle|^2]^2 &= N^6 + N^5(6M-6) + N^4(11M^2 - 24M + 13) + N^3(6M^3 - 24M^2 + 30M - 12) \\
&+ N^2(M^4 - 6M^3 + 13M^2 - 12M + 4).
\end{aligned}$$

Then a calculation shows that for $k \in T$

$$\begin{aligned}
&\mathbb{E}[|\langle y, \phi_k\rangle|^4] - (\mathbb{E}[|\langle y, \phi_k\rangle|^2])^2 \\
&\leq N^6 - N^5(2M - 31) + N^4(3M^2 + 60M - 7) + N^3(6M^3 + 118M^2 - 10M + 12) \\
&+ N^2(M^4 + 47M^3 + -13M^2 + 12M - 4) + NM^4.
\end{aligned}$$

Since $N \geq 2M$ it is straightforward to verify that

$$\mathbb{E}[|\langle y, \phi_k\rangle|^2]^2 \geq \mathbb{E}[|\langle y, \phi_k\rangle|^4] - \mathbb{E}[|\langle y, \phi_k\rangle|^2]^2 \tag{4.5}$$

for all $M \geq 20$ (say). Hence, under condition $\alpha^2 \geq \sigma \mathbb{E}[|\langle y, \phi_k\rangle|^2]$ by (4.4) we have

$$\mathbb{P}(|\langle y, \phi_k\rangle| > \alpha) \leq \frac{\mathbb{E}[|\langle y, \phi_k\rangle|^4] - (\mathbb{E}[|\langle y, \phi_k\rangle|^2])^2}{(\sigma-1)^2(\mathbb{E}[|\langle y, \phi_k\rangle|^2])^2} \leq \frac{1}{(\sigma-1)^2}.$$

Further, by (4.3) and Lemmas 4.1 and 4.2 we get

$$\mathbb{P}(|\langle y, \phi_\ell\rangle| < \alpha) \leq 4\frac{\mathbb{E}\left[|\langle y, \phi_\ell\rangle|^2\right] - (\mathbb{E}|\langle y, \phi_\ell\rangle|)^2}{(\mathbb{E}|\langle y, \phi_\ell|)^2} = \frac{4(N-1)}{MN} \leq \frac{4}{M}$$

under condition $\alpha \leq \frac{1}{2}\mathbb{E}|\langle y, \phi_\ell\rangle|$. Summarizing, if

$$(\mathbb{E}|\langle y, \phi_\ell\rangle|)^2 \geq 4\sigma \mathbb{E}[|\langle y, \phi_k\rangle|^2] \tag{4.6}$$

then the probability that thresholding succeeds can be upper bounded by

$$\mathbb{P}(\min_{k\in T} |\langle y, \phi_k\rangle| > |\langle y, \phi_\ell\rangle|) \leq \frac{4}{M} + \frac{1}{(\sigma-1)^2}.$$

By Lemmas 4.1 and 4.3 the condition (4.6) is equivalent to

$$M^2N^2 \geq 4\sigma((M-1)(M-2)N + 3(M-1)N^2 + N^3)).$$

Since $N \geq 2M$ this is satisfied under (3.1). Moreover, for $M < 20$ the maximal N satisfying (3.1) is less than 1 and then the statement of the theorem becomes trivial. Hence, for all valid M, N (4.5) is satisfied and the proof of Theorem 3.1 is finished.

By (4.2) the probability that OMP succeeds in M steps can be estimated from above by

$$\mathbb{P}(\max_{k\in T} |\langle y, \phi_k\rangle| > |\langle y, \phi_\ell\rangle|) \leq \frac{4}{M} + \frac{M}{(\sigma-1)^2}.$$

We choose σ such that $\frac{M}{(\sigma-1)^2} = \tau^2 < 1$, i.e., $\sigma = \sqrt{M}\tau^{-1}+1$. As above (4.6) is satisfied under condition (3.1), which now reads

$$N \leq \frac{1}{4\sqrt{M}\tau^{-1}+1}M^2 - \frac{7}{2}(M-1).$$

This is certainly satisfied if

$$N \leq \frac{\tau}{5}M^{3/2} - \frac{7}{2}(M-1). \tag{4.7}$$

Then the probability that OMP succeeds in M steps can be upperestimated by

$$\frac{4}{M} + \tau^2.$$

Moreover, if $M < 20$ then as above the minimal N satisfying (4.7) is less than 1, and hence, again we can omit the condition $M \geq 20$ ensuring (4.5). This completes the proof of Theorem 3.2.

Remark.

- The proof shows that the conditions on Γ and M can be slightly relaxed in Theorems 3.1 and 3.2. We only have to require the existence of ℓ and T as in Lemma 4.4. In particular, also $\Gamma = \{-q, -q+1, \ldots, q\}^d$ is a valid choice corresponding to spaces of trigonometric polynomials of degree at most q.
- In order to improve Theorems 3.1 and 3.2 with respect to the probability estimate and the exponent 3/2 in (3.2) one might work with higher moments rather than only the 4th moment as in Lemma 4.4. However, computing higher moments will be even more tedious than the proof of Lemma 4.4.

5 Proof of Lemma 4.4

We have

$$\begin{aligned}
\mathbb{E}[|\langle y, \phi_{k_0}\rangle|^4] &= \mathbb{E}\left[\Big|\sum_{k\in T}\langle \phi_\ell, \phi_k\rangle\langle \phi_k, \phi_{k_0}\rangle\Big|^4\right] \\
&= \sum_{k_1,k_2,k_3,k_4\in T} \mathbb{E}\left[\langle \phi_\ell, \phi_{k_1}\rangle\langle \phi_{k_2}, \phi_\ell\rangle\langle \phi_\ell, \phi_{k_3}\rangle\langle \phi_{k_4}, \phi_\ell\rangle \times\right. \\
&\qquad \left.\times\langle \phi_{k_1}, \phi_{k_0}\rangle\langle \phi_{k_0}, \phi_{k_2}\rangle\langle \phi_{k_3}, \phi_{k_0}\rangle\langle \phi_{k_0}, \phi_{k_4}\rangle\right] \\
&= \sum_{k_1,k_2,k_3,k_4} E(k_1,k_2,k_3,k_4),
\end{aligned}$$

where

$$E(k_1,k_2,k_3,k_4) = \sum_{j_1,\ldots,j_8=1}^{N} \mathbb{E}\left[e^{2\pi i\left(\sum_{i=1}^4(-1)^i\left[(\ell-k_i)\cdot x_{j_i}+(k_i-k_0)\cdot x_{j_{4+i}}\right]\right)}\right].$$

Given $(j_1,\ldots,j_8)$ we let $\mathcal{A} = (A_1,\ldots,A_k)$ be a partition of $\{1,\ldots,8\}$ into r disjoint subsets (called blocks) such that $i,i' \in \{1,\ldots,8\}$ is contained in the same block $A \in \mathcal{A}$ if and only if $j_i = j_{i'} =: j_A$. By $B(r)$ we denote the collection of all such partitions of $\{1,\ldots,8\}$ into r blocks. Then by independence the previous sum can be written as

$$\begin{aligned}
&E(k_1,k_2,k_3,k_4) \\
&= \sum_{r=1}^{8}\sum_{\mathcal{A}\in B(r)}\frac{N!}{(N-r)!}\prod_{A\in\mathcal{A}}\mathbb{E}\left[e^{2\pi i\left(\sum_{i\in A, i\le 4}(-1)^i(\ell-k_i)\cdot x_{j_A}+\sum_{i\in A,i\ge 5}(-1)^i(k_{i-4}-k_0)\cdot x_{j_A}\right)}\right] \\
&= \sum_{r=1}^{8}\sum_{\mathcal{A}\in B(r)}\frac{N!}{(N-r)!}\prod_{A\in\mathcal{A}}\delta_0\left(\sum_{i\in A,i\le 4}(-1)^i(\ell-k_i)+\sum_{i\in A,i\ge 5}(-1)^i(k_{i-4}-k_0)\right).
\end{aligned}$$

The product in the previous expression contributes to the sum if and only if

$$\sum_{i\in A,i\le 4}(-1)^i(\ell-k_i)+\sum_{i\in A,i\ge 5}(-1)^i(k_{i-4}-k_0) = 0 \quad \text{for all } A\in\mathcal{A}. \tag{5.1}$$

The condition $k \equiv 0 \mod 4$ for all $k \in T$ implies that the second sum above is always 0 modulo 4. Thus, if the sum above vanishes then necessarily also the first sum has to be $0 \mod 4$. But due to the condition $\ell_n \equiv 1 \mod 4$ this can only happen if ℓ cancels completely, which in turn implies that either the pairs $\{1,2\}$ and $\{3,4\}$ or the pairs $\{1,4\}$ and $\{2,3\}$ are each contained in the same block of the partition $\mathcal{A}$. Note that this means as well that the partition $\mathcal{A}$ contains at most 6 blocks and we can write

$$\mathbb{E}[|\langle y, \phi_{k_0}\rangle|^4] = \sum_{r=1}^{6}\frac{N!}{(N-r)!}D(r)$$

where $D(r) = \sum_{\mathcal{A}\in B(r)} C_r(\mathcal{A})$ and

$$C_r(\mathcal{A}) = \#\{(k_1,k_2,k_3,k_4)\in T^4 : (5.1) \text{ is satisfied}\}.$$

Hence, we need to determine the coefficients $D(r)$, $r = 1, \dots, 6$. For $r = 6$ we have to consider only the partitions

$$\mathcal{A}_1 = \{\{1,2\},\{3,4\},\{5\},\{6\},\{7\},\{8\}\} \quad \text{and} \quad \mathcal{A}_2 = \{\{1,4\},\{2,3\},\{5\},\{6\},\{7\},\{8\}\}.$$

But (5.1) then means $k_i = k_0$ for all $i = 1, \dots, 4$, and we conclude that $D(6) = 2$. Now consider $r = 1$. Then the only partition consists of the block $A = \{1, \dots, 8\}$ and (5.1) is satisfied for all possible choices of $(k_1, k_2, k_3, k_4) \in T^4$. Hence $D(1) = |T|^4 = M^4$.

Now consider $r = 5$. We have the following possible partitions:

- $\mathcal{A} = \{\{1,2,3,4\},\{5\},\{6\},\{7\},\{8\}\}$. As above (5.1) necessarily requires $k_1 = k_2 = k_3 = k_4 = k_0$, hence, $C_5(\mathcal{A}) = 1$.
- We take the two partitions $\{\{1,2\},\{3,4\}\}$ and $\{\{1,4\},\{2,3\}\}$ of the numbers $\{1, \dots, 4\}$ and then add one of the numbers $5, \dots, 8$ to one of these blocks and let the other 3 blocks of $\mathcal{A}$ consist of the remaining numbers each. This gives an overall number of 16 possible partitions, and as above (5.1) requires $k_1 = k_2 = k_3 = k_4 = k_0$. Hence, $C_5(\mathcal{A}) = 1$ for each of those 16 partitions.
- We take again the two partitions $\{\{1,2\},\{3,4\}\}$ and $\{\{1,4\},\{2,3\}\}$. Then we form a partition of $\{5, \dots, 8\}$ into 2 blocks of 1 elements and 1 block of 2 elements and combine these into a partition of $\{1, \dots, 8\}$. In case the 2-element block is $\{5,6\}$ then the single element blocks are $\{7\},\{8\}$ and (5.1) requires $k_3 = k_4 = k_0$. Now, if we form the combination with $\{\{1,4\},\{2,3\}\}$ then (5.1) requires also $k_2 = k_3 = k_0$ and $k_1 = k_4 = k_0$ and hence, $C_5(\mathcal{A}) = 1$ for the corresponding partition. If we form the combination with $\{\{1,2\},\{3,4\}\}$ then (5.1) requires $k_1 = k_2$, and we conclude $C_5(\mathcal{A}) = M$ for $\mathcal{A} = \{\{1,2\},\{3,4\},\{5,6\},\{7\},\{8\}\}$. A similar situation occurs for the 2-element blocks $\{5,8\},\{6,7\},\{7,8\}$. Now, if the 2-element block is $\{5,7\}$ then (5.1) for the remaining 1-element blocks requires $k_2 = k_4 = k_0$. Now for both $\{\{1,2\},\{3,4\}\}$ and $\{\{1,4\},\{2,3\}\}$ this implies that also $k_1 = k_4 = k_0$, and hence $C_5(\mathcal{A}) = 1$ for the corresponding partition $\mathcal{A}$. The same occurs for the 2-element block $\{6,8\}$. Counting all cases we have $C_5(\mathcal{A}) = 1$ for 8 partitions $\mathcal{A}$ and $C_5(\mathcal{A}) = M$ for 4 partitions.

Altogether we have

$$D(5) = 1 + 16 + 8 + 4M = 4M + 25.$$

Now let $r = 4$. Then we have the following possibilities.

- We take $\{1,2,3,4\}$ and add one of the numbers $5, \dots, 8$ to it. Further, we choose the remaining three numbers as single element blocks. Then (5.1) requires $k_1 = k_2 = k_3 = k_4 = k_0$ and, hence, $C_4(\mathcal{A}) = 1$ for each of the 4 of such partitions.
- We take $\{\{1,2\},\{3,4\}\}$, add two of the numbers $5,6,7,8$ to one of these blocks, and take the remaining numbers as single element blocks. Consider first the resulting partition $\mathcal{A} = \{\{1,2,5,6\},\{3,4\},\{7\},\{8\}\}$. Then (5.1) is equivalent to $k_3 = k_4 = k_0$, hence, $C_4(\mathcal{A}) = M^2$, and similarly for $\mathcal{A} = \{\{1,2\},\{3,4,7,8\},\{5\},\{6\}\}$. The remaining 6 partitions constructed in this way satisfy $C_4(\mathcal{A}) = M$, as can easily be seen. The same considerations hold, of course, if we start with $\{\{1,4\},\{2,3\}\}$. Counting cases we get 4 times $C_4(\mathcal{A}) = M^2$ and 12 times $C_4(\mathcal{A}) = M$.
- We take $\{\{1,2\},\{3,4\}\}$, add one of the numbers $5,6,7,8$ to $\{1,2\}$ and another one to $\{3,4\}$, and let the remaining numbers form a single element block each. Assume first that the resulting partition is $\mathcal{A} = \{\{1,2,5\},\{3,4,7\},\{6\},\{8\}\}$. Then (5.1) means $k_2 = k_4 = k_0$, hence, $C_4(\mathcal{A}) = M^2$, and similarly for the other 3 partitions where 5 or 6 is added to $\{1,2\}$ and 7 or 8 is added to $\{3,4\}$. If $\mathcal{A} = \{\{1,2,5\},\{3,4,6\},\{7\},\{8\}\}$ then (5.1) is satisfied iff $k_2 = k_3 = k_4 = k_0$,

hence, $C_4(\mathcal{A}) = M$, and as well for the other 3 partitions where either 5 and 6 or 7 and 8 are added to the first blocks. The same considerations apply if we start with $\{\{1,4\},\{2,3\}\}$. Counting cases yields 8 times $C_4(\mathcal{A}) = M^2$ and 8 times $C_4(\mathcal{A}) = M$.

- As one block of the partition we take $\{1,2,3,4\}$ and as the remaining three blocks we take a partition of $\{5,6,7,8\}$ into two 1-element blocks and one 2-element block. If the 1-element blocks are $\{5\},\{6\}$ then $k_1 = k_2 = k_0$ and (5.1) applied to $\{1,2,3,4\}$ yields also $k_3 = k_4$ and then (5.1) is also satisfied for the remaining block $\{7,8\}$. Hence, in this case $C_4(\mathcal{A}) = M$. The same appears for the 2-element blocks $\{5,6\}$,$\{5,8\}$,$\{6,7\}$. Now if the 1-element blocks are $\{5\},\{7\}$ then $k_1 = k_3 = k_0$. Further, (5.1) is satisfied for both $A = \{1,2,3,4\}$ and $\{6,8\}$ if and only if $k_2 + k_4 = 2k_0$. This is satisfied in particular if $k_2 = k_4 = k_0$, but at most for M pairs $\{k_2,k_4\}$, hence $1 \le C_4(\mathcal{A}) \le M$ for this partition. The same holds for the 2-element block $\{5,7\}$. Counting cases we get 4 times $C_4(\mathcal{A}) = M$ and twice $1 \le C_4(\mathcal{A}) = M$.
- We take $\{\{1,2\},\{3,4\}\}$ and as the remaining two blocks we take a partition of $\{5,6,7,8\}$. If the partition of $\{5,\dots,8\}$ is $\{\{5,6\},\{7,8\}\}$ then (5.1) is satisfied if and only if $k_1 = k_2$ and $k_3 = k_4$, hence $C_4(\mathcal{A}) = M^2$ in this case. If the added two blocks are $\{5,8\}$ and $\{6,7\}$ then (5.1) is satisfied if and only if $k_1 = k_2 = k_3 = k_4$, hence $C_4(\mathcal{A}) = M$. If we add $\{5,7\}$ and $\{6,8\}$ then (5.1) is satisfied if and only if $k_1 = k_2 = k_3 = k_4 = k_0$, hence $C_4(\mathcal{A}) = 1$. If the remaining blocks are $\{5\}$ and $\{6,7,8\}$ then (5.1) is satisfied iff $k_1 = k_2 = k_0$ and $k_3 = k_4$, hence, $C_4(\mathcal{A}) = M$, and this holds as well if the single element block is $\{6\}$, $\{7\}$ or $\{8\}$. Clearly, similar observations hold if we start with $\{\{1,4\},\{2,3\}\}$. Counting cases yields twice $C_4(\mathcal{A}) = M^2$, 10 times $C_4(\mathcal{A}) = M$ and twice $C_4(\mathcal{A}) = 1$.

We conclude that

$$D(4) \le 4 + 4M^2 + 12M + 8M^2 + 8M + 4M + 2M + 2M^2 + 10M + 2 \;=\; 14M^2 + 36M + 6,$$

(but also $D(4) \ge 14M^2 + 34M + 8$). Now let $r = 3$. We distinguish the following cases.

- We take $\{1,2,3,4\}$ and add two of the numbers $5,\dots,8$ to it and take the remaining two as single element blocks. Suppose first that these are $\{5\},\{6\}$. Then (5.1) is satisfied if and only if $k_1 = k_2 = k_0$ and $k_3 = k_4$, hence $C_3(\mathcal{A}) = M$. The same holds if the single element blocks are $\{5\},\{8\}$ or $\{6\},\{7\}$ or $\{7\},\{8\}$ giving a total of 4 possibilities. Further, if the single element blocks are $\{5\},\{7\}$ or $\{6\},\{8\}$ then (5.1) is satisfied if and only if $k_1 = k_3 = k_0$ and $k_2 + k_4 = 2k_0$, hence $1 \le C_3(\mathcal{A}) \le M$. Althogether, we have 4 times $C_3(\mathcal{A}) = M$ and twice $1 \le C_3(\mathcal{A}) \le M$.
- Take $\{1,2,3,4\}$ and add one of the numbers $5,\dots,8$ to it and form two blocks from the remaining three numbers, i.e., one single element block and one two element block. Suppose first that these blocks are $\{5\},\{7,8\}$. Then (5.1) is satisfied iff $k_1 = k_2 = 0$ and $k_3 = k_4$, hence $C_3(\mathcal{A}) = M$. The same holds for $\{6\},\{7,8\}$ and $\{7\},\{5,6\}$ and $\{8\},\{5,6\}$. Now suppose we have the blocks $\{5\},\{6,8\}$. Then (5.1) is satisfied iff $k_1 = k_3 = k_0$ and $k_2 + k_4 = 2k_0$, hence $1 \le C_3(\mathcal{A}) \le M$. This holds as well for $\{6\},\{5,7\}$. Counting cases yields 4 times $C_3(\mathcal{A}) = M$ and twice $1 \le C_3(\mathcal{A}) \le M$.
- We start with $\{1,2\},\{3,4\}$ and take $\{5,6,7,8\}$ as third block. Then (5.1) is satisfied iff $k_1 = k_2$ and $k_3 = k_4$, hence $C_3(\mathcal{A}) = M^2$. The same holds for the partition $\mathcal{A} = \{\{1,4\},\{2,3\},\{5,6,7,8\}\}$.

- Take $\{1,2\},\{3,4\}$, form a partition of $\{5,6,7,8\}$ into two blocks and add one these two blocks to one of the sets $\{1,2\},\{3,4\}$. Consider first the resulting partition $\mathcal{A}=\{\{1,2,5,6\},\{3,4\},\{7,8\}\}$. Then (5.1) is satisfied iff $k_3=k_4$, and hence, $C_3(\mathcal{A})=M^3$. This holds as well for the partition $\{\{1,2\},\{3,4,7,8\},\{5,6\}\}$. The partition $\mathcal{A}=\{\{1,2,7,8\},\{3,4\},\{5,6\}\}$ requires $k_1=k_2$ and $k_3=k_4$, hence $C_3(\mathcal{A})=M^2$, and similarly for $\mathcal{A}=\{\{3,4,5,6\},\{1,2\},\{7,8\}\}$. If we have $\mathcal{A}=\{\{1,2,5,7\},\{3,4\},\{6,8\}\}\}$ then (5.1) is satisfied iff $k_2=k_4=2k_0$ and $k_3=k_4$ while k_1 is free. Thus, $M\leq C_3(\mathcal{A})\leq M^2$, and similarly for the three partitions $\{\{1,2,6,8\},\{3,4\},\{5,7\}\}$, $\{\{3,4,5,7\},\{1,2\},\{6,8\}\}$, $\{\{3,4,6,8\},\{1,2\},\{5,7\}\}$. If $\mathcal{A}=\{\{1,2,5,8\},\ \{3,4\},\{6,7\}\}$ then (5.1) is satisfied iff $k_2=k_3=k_4$, hence $C_3(\mathcal{A})=M^2$, and similarly for the remaining three partitions consisting of one 4-element block and 2 two-element blocks.

  If the resulting partition is $\mathcal{A}=\{\{1,2,5\},\{3,4\},\{6,7,8\}\}$ then (5.1) is satisfied iff $k_2=k_0$ and $k_3=k_4$, hence, $C_3(\mathcal{A})=M^2$, and similarly if 6 is added to $\{1,2\}$, or 7 or 8 are added to $\{3,4\}$, giving a total of 4 partitions with $C_3(\mathcal{A})=M^2$. Now consider $\mathcal{A}=\{\{1,2,7\},\{3,4\},\{5,6,8\}\}$. Then (5.1) holds iff $k_3=k_4=k_0$ and $k_1=k_2$, hence, $C_3(\mathcal{A})=M$. Similarly, if 8 is added to $\{1,2\}$, or 5 or 6 are added to $\{3,4\}$. Next, consider $\mathcal{A}=\{1,2,5,6,7\},\{3,4\},\{8\}\}$. Then (5.1) is satisfied iff $k_3=k_4=k_0$, hence, $C_3(\mathcal{A})=M^2$. The same holds as well for the partitions where 7 is a single element block and $\{5,6,8\}$ is added to $\{1,2\}$; and 5 or 6 are single element blocks and the remaining numbers are added to $\{3,4\}$. Finally, consider $\mathcal{A}=\{\{1,2,6,7,8\},\{3,4\},\{5\}\}$. Then (5.1) is equivalent to $k_1=k_0$ and $k_3=k_4$, thus, $C_3(\mathcal{A})=M^2$, and similarly if 6 is a single element block and the remaining numbers are added to $\{1,2\}$, or 7 or 8 are single element blocks and the remaining numbers are added to $\{3,4\}$. Counting cases yields twice $C_3(\mathcal{A})=M^3$, 18 times $C_3(\mathcal{A})=M^2$, 4 times $M\leq C_3(\mathcal{A})\leq M^2$ and 4 times $C_3(\mathcal{A})=M$.

  Similar considerations as above apply if we start with $\{1,4\},\{2,3\}$, hence, we have to take all the above quantities into account twice.

- Finally, take $\{1,2\},\{3,4\}$, partition the remaining numbers $\{5,6,7,8\}$ into three blocks and add one of them to $\{1,2\}$ and another one to $\{3,4\}$. Suppose first that we result in $\{\{1,2,5,6\},\{3,4,7\},\{8\}\}$. Then (5.1) is satisfied iff $k_4=k_0$, thus, $C_3(\mathcal{A})=M^3$. The same holds if the role of 7 and 8 is interchanged, or the 4-element block is $\{3,4,7,8\}$, resulting in 4 possible partitions giving $C_3(\mathcal{A})=M^3$. Now, suppose we have $\mathcal{A}=\{\{1,2,7,8\},\{3,4,5\},\{6\}\}$. Then (5.1) requires $k_2=k_0$ and $k_3-k_4=k_1-k_0$, hence $M\leq C_3(\mathcal{A})\leq M^2$, and there are three further partitions for which similar considerations hold. Next, consider $\mathcal{A}=\{\{1,2,5,7\},\{3,4,6\},\{8\}\}$. This implies $k_4=k_0$ and $k_2+k_3=2k_0$, hence, $M\leq C_3(\mathcal{A})\leq M^2$, and the same holds for 7 further partitions. The partition $\mathcal{A}=\{\{1,2,5\},\{3,4,8\},\{6,7\}\}$ implies $k_2=k_3=k_0$, hence, $C_3(\mathcal{A})=M^2$, and the same holds if 5 is replaced by 6 and / or 8 by 7 (giving a total of 4 partitions). Consider $\mathcal{A}=\{\{1,2,7\},\{3,4,8\},\{5,6\}\}$. Then $k_3=k_0$ and $k_1=k_2$, hence, $C_4(\mathcal{A})=M^2$, and similarly if 7 and 8 are interchaned, and as well if the roles of $5,6$ and $7,8$ are exchanged (giving again a total of 4 partitions). Next, take $\mathcal{A}=\{\{1,2,8\},\{3,4,5\},\{6,7\}\}$. Then (5.1) is equivalent to $k_2=k_3$ and $k_2-k_4=k_1-k_0$, thus, $M\leq C_3(\mathcal{A})\leq M^2$, and similarly if both 8 is interchanged with 7 and 5 with 6. Finally, consider $\mathcal{A}=\{\{1,2,8\},\{3,4,6\},\{5,7\}\}$. Then (5.1) is satisfied iff $k_1+k_3=2k_0$ and $k_3+k_2=k_4+k_0$, thus $1\leq C_4(\mathcal{A})\leq M^2$, and similarly for $\mathcal{A}=\{\{1,2,7\},\{3,4,6\},\{6,8\}\}$. Counting cases gives 4 times $C_3(\mathcal{A})=M^3$, 14 times $M\leq C_3(\mathcal{A})\leq M^2$, 8 times $C_3(\mathcal{A})=M^2$ and twice $1\leq C_3(\mathcal{A})\leq M^2$.

Again, the same consideration as above are valid if we start from $\{1,4\},\{2,3\}$. Collecting all the above cases we conclude that

$$
\begin{aligned}
D(3) \leq\ & 4M + 2M + 4M + 2M + 2M^2 + 2(2M^3 + 18M^2 + 4M^2 + 4M) \\
& + 2(4M^3 + 14M^2 + 8M^2 + 2M^2) \\
=\ & 12M^3 + 94M^2 + 20M,
\end{aligned}
$$

(but $D(3) \geq 12M^3 + 54M^2 + 52M + 8$). Finally, let $r = 2$. Then we distinguish the following cases:

- Consider $\mathcal{A} = \{\{1,2,3,4\},\{5,6,7,8\}\}$. Then (5.1) is equivalent to $k_1 - k_2 + k_3 - k_4 = 0$, and hence, $M^2 \leq C_2(\mathcal{A}) \leq M^3$.
- Take $A = \{1,2,3,4\}$, form a partition of $\{5,6,7,8\}$ into two blocks and add the elements of one these blocks to A. If a single element is added to A, say 5 then $k_0 - k_2 = k_3 - k_4$, $M^2 \leq C_2(\mathcal{A}) \leq M^3$ and there are 4 possibilities of doing this. If an element, say 5, is kept as a single element block then $k_1 = k_0$, hence, $C_2(\mathcal{A}) = M^3$ and again there are 4 possibilities for this. If $\{5,6\}$ is added to A and $\{7,8\}$ remains as 2-element block then $k_3 = k_4$, hence, $C_2(\mathcal{A}) = M^3$, and the same holds with the roles of $\{5,6\}$ and $\{7,8\}$ interchanged, and furthermore if we replace both blocks by $\{5,8\}$, $\{6,7\}$. Now, for $\mathcal{A} = \{\{1,2,3,4,5,7\},\{6,8\}\}$ (5.1) is equivalent to $k_2 + k_4 = 2k_0$, and hence, $M^2 \leq C_2(\mathcal{A}) \leq M^3$, and similarly for $\mathcal{A} = \{\{1,2,3,4,6,8\},\{5,7\}\}$. Altogether we have 6 times $M^2 \leq C_2(\mathcal{A}) \leq M^3$ and 8 times $C_2(\mathcal{A}) = M^3$.
- Take $\{1,2\},\{3,4\}$, form a partition of $\{5,6,7,8\}$ into two blocks and add one block to $\{1,2\}$ and the other one to $\{3,4\}$. If the resulting partition is $\mathcal{A} = \{\{1,2,5,6\},\{3,4,7,8\}\}$ then (5.1) is always satisfied, hence $C_2(\mathcal{A}) = M^4$. If $\mathcal{A} = \{\{1,2,7,8\},\{3,4,5,6\}\}$ then (5.1) is equivalent to $k_1 - k_2 = k_3 - k_4$, hence $M^2 \leq C_2(\mathcal{A}) \leq M^3$. For $\mathcal{A} = \{\{1,2,5,8\},\{3,4,6,7\}\}$ (5.1) is equivalent to $k_2 = k_4$, thus, $C_2(\mathcal{A}) = M^3$, and similarly for $\mathcal{A} = \{\{1,2,6,7\},\{3,4,5,8\}\}$. If $\mathcal{A} = \{\{1,2,5,7\},\{3,4,6,8\}\}$ then $k_2 + k_3 = 2k_0$ and $M^2 \leq C_2(\mathcal{A}) \leq M^3$; and similarly for $\mathcal{A} = \{\{1,2,6,8\},\{3,4,5,7\}\}$.

  If $\mathcal{A} = \{\{1,2,5\},\{3,4,6,7,8\}\}$ then (5.1) means $k_2 = k_0$ and $C_2(\mathcal{A}) = M^3$, and similarly, if 6 is added to $\{1,2\}$, or 7 or 8 is added to $\{3,4\}$. If $\mathcal{A} = \{\{1,2,7\},\{3,4,5,6,8\}\}$ then (5.1) requires $k_1 - k_2 = k_3 - k_0$, hence $M^2 \leq C_2(\mathcal{A}) \leq M^3$, and the same holds if 8 is added to $\{1,2\}$, or 5 or 6 is added to $\{3,4\}$. Counting cases yields once $C_2(\mathcal{A}) = M^4$, 7 times $M^2 \leq C_2(\mathcal{A}) \leq M^3$ and 6 times $C_2(\mathcal{A}) = M^3$.

  Clearly, the same considerations apply if we start with $\{1,4\},\{2,3\}$.

Summing up all possibilities we obtain

$$
D(2) \leq M^3 + 6M^3 + 8M^3 + 2(M^4 + 7M^3 + 6M^3) = 2M^4 + 41M^3
$$

(and $D(2) \geq 2M^4 + 21M^3 + 21M^2$). Finally, we obtain

$$
\begin{aligned}
\mathbb{E}[|\langle y, \phi_{k_0}\rangle|^4] &= \sum_{r=1}^{6} \frac{N!}{(N-r)!} D(r) \leq \sum_{r=1}^{6} N^r D(r) \\
&\leq NM^4 + N^2(2M^4 + 41M^3) + N^3(12M^3 + 94M^2 + 20M) \\
&+ N^4(14M^2 + 36M + 6) + N^5(4M + 25) + 2N^6,
\end{aligned}
$$

which is precisely the statement of Lemma 4.4.

References

[1] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Constr. Approx.*, to appear.

[2] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inform. Theory*, 52(2):489–509, 2006.

[3] E. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Comm. Pure Appl. Math.*, 59(8):1207–1223, 2006.

[4] E. Candès and T. Tao. Near optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Inform. Theory*, 52(12):5406–5425, 2006.

[5] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by Basis Pursuit. *SIAM J. Sci. Comput.*, 20(1):33–61, 1999.

[6] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k-term approximation. *preprint*, 2006.

[7] D. L. Donoho. Compressed sensing. *IEEE Trans. Inform. Theory*, 52(4):1289–1306, 2006.

[8] D. L. Donoho. For most large underdetermined systems of linear equations the minimal l^1 solution is also the sparsest solution. *Commun. Pure Appl. Anal.*, 59(6):797–829, 2006.

[9] A. C. Gilbert and J. A. Tropp. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Inform. Theory*, to appear.

[10] S. Kunis and H. Rauhut. Random sampling of sparse trigonometric polynomials II - orthogonal matching pursuit versus basis pursuit. *Found. Comput. Math.*, to appear.

[11] D. Needell and R. Vershynin. Uniform uncertainty principle and signal recovery via regularized orthogonal matching pursuit, 2007. Preprint.

[12] H. Rauhut. Stability results for random sampling of sparse trigonometric polynomials. *preprint*, 2006.

[13] H. Rauhut. Random sampling of sparse trigonometric polynomials. *Appl. Comput. Harmon. Anal.*, 22(1):16–42, 2007.

[14] M. Rudelson and R. Vershynin. Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements. In *Proc. CISS 2006 (40th Annual Conference on Information Sciences and Systems)*, 2006.

[15] M. Rudelson and R. Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Comm. Pure Appl. Math.*, to appear.

[16] J. A. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Trans. Inform. Theory*, 50(10):2231–2242, 2004.