Random Sampling of Sparse Trigonometric Polynomials

Holger Rauhut
NuHAG, Faculty of Mathematics, University of Vienna
Nordbergstrasse 15, A-1090 Wien, Austria
rauhut@ma.tum.de

December 29, 2005

Abstract

We study the problem of reconstructing a multivariate trigonometric polynomial having only few non-zero coefficients from few random samples. Inspired by recent work of Candes, Romberg and Tao we propose to recover the polynomial by Basis Pursuit, i.e., by ℓ^1-minimization. Numerical experiments show that in many cases the trigonometric polynomial can be recovered exactly provided the number N of samples is high enough compared to the "sparsity" – the number of non-vanishing coefficients. However, N can be chosen small compared to the assumed maximal degree of the trigonometric polynomial. Hence, the proposed scheme may overcome the Nyquist rate. We present two theorems that explain this observation. Unexpectedly, they establish a connection to an interesting combinatorial problem concerning set partitions, which seemingly has not yet been considered before.

Key Words: random sampling, trigonometric polynomials, Basis Pursuit, ℓ^1-minimization, sparse recovery, set partitions, random matrices
AMS Subject classification: 94A20, 42A05, 15A52, 05A18, 90C05, 90C25

1 Introduction

Recently, Candes, Romberg and Tao observed the surprising fact that it is possible to recover certain discrete functions exactly from vastly incomplete information on their discrete Fourier transform [6, 7, 8, 9]. The crucial property of these functions is their sparsity with respect to the canonical (Dirac) basis, i.e., their (unknown) support is very small. The recovery procedure consists in minimizing the ℓ^1-norm of the signal subject to the constraint that the Fourier coefficients are matched. This task is also known as Basis Pursuit [5]. Since minimizing the total variation norm can be reformulated as minimizing the ℓ^1-norm there are relevant applications in image processing, in particular, computer tomography [6, 9].

This paper is concerned with the related problem of reconstructing a sparse trigonometric polynomial from few randomly chosen samples drawn from the continuous uniform distribution on $[0, 2\pi]^d$. By "sparse" we mean that only very few coefficients of the polynomial are non-zero. However, a priori we do not know the support of the coefficients. From a practical viewpoint considering such polynomials can be motivated as follows. First, trigonometric polynomials with a certain maximal degree model band-limited signals. Secondly, in many cases it seems reasonable that only few coefficients (with unknown location) are large. Such a signal can at least be approximated by a sparse one.

We propose to reconstruct the sparse polynomial from its random samples by Basis Pursuit similarly as in [6, 7, 8, 9]. From numerical experiments it is evident that this scheme can indeed reconstruct the polynomial exactly provided the number of samples is large enough with respect to the sparsity. When comparing the number of samples to the assumed maximal degree of the polynomial it turns out that this method may overcome the Nyquist rate by far. Thus, the described recovery method is very likely to have high potential for practical applications in signal processing.

We will present two theorems that explain the observed phenomenon. Similar to [6] the first one estimates the probability of exact reconstruction given an arbitrary sparse trigonometric polynomial. Hence, this can be viewed as a worst case estimate. Our second theorem is more directed towards an avarage case analysis. It gives a probability estimate for generic polynomials in the sense that the support of the coefficients is modelled as random set. A result of this type seems to be new. As one may expect it gives better probability estimates than the first one. However unexpectly, it relates the problem to a seemingly new and difficult combinatorial problem about set partitions. Unfortunately, we were not able to solve this problem in general, and as a consequence we cannot yet exploit fully our probability estimate. We have to leave the combinatorial aspect as an interesting open problem.

We would like to mention some recent related work. In [7, 8] Candes et al. study stability aspects of the problem and investigate also recovery from few inner products with random vectors following Gaussian distributions and binary distributions. In [9] some practical examples are presented. The recovery from Gaussian measurements via Basis Pursuit is also investigated by Rudelson and Vershynin in [18] in the context of error correcting codes, while Tropp [14] studies the reconstruction by Orthogonal Matching Pursuit. In [10, 11] Donoho and Tsaig introduce the terminology "compressed sensing" for this range of problems and in [12, 13] probabilistic results concerning Basis Pursuit are discussed. A randomized sublinear algorithm for reconstructing sparse Fourier data is introduced and analyzed in [21]. If the reader is interested in reconstructing not necessarily sparse trigonometric polynomial from random samples we refer to recent work of Bass and Gröchenig [1], where probabilistic estimates of related condition numbers are developed.

The paper is structured as follows. In Section 2 we describe the problem and present our main results. To this end we also need to introduce some background on set partitions. Section 3 will be devoted to the proofs. Section 4 gives some more information on the combinatorial problem related to our second theorem. In Section 5 we present some plots of the probability bounds resulting from our theorems and finally Section 6 describes some numerical experiments.

Acknowledgements: The author was supported by the European Union's Human Potential Programme, under contract HPRN–CT–2002–00285 (HASSIP). He would like to thank Stefan Kunis for stimulating discussions on numerical aspects of the topic. Also he acknowledges interesting conversations with Justin Romberg and his mail correspondence with Emmanuel Candes on the subject.

2 Description of the Main Results

2.1 The Setting

Let $\Pi_q = \Pi_q^d$ denote the space of all trigonometric polynomials of maximal order $q \in \mathbb{N}_0$ in dimension d. An element f of Π_q is of the form

$$f(x) = \sum_{k \in [-q,q]^d \cap \mathbb{Z}^d} c_k e^{ik\cdot x}, \qquad x \in [0, 2\pi]^d,$$

with some Fourier coefficients $c_k \in \mathbb{C}$. The dimension of Π_q^d will be denoted by $D := (2q+1)^d$. In the sequel we will use the short notation $[-q, q]^d$ instead of $[-q, q]^d \cap \mathbb{Z}^d$.

Through the rest of this paper we will be dealing with "sparse" trigonometric polynomials, i.e., we assume that the sequence of coefficients c_k is supported only on a set T, which is much smaller than the dimension D of Π_q. However, a priori nothing is known about T apart from a maximum size. Thus, it is useful to introduce the set $\Pi_q(M) = \Pi_q^d(M) \subset \Pi_q$ of all trigonometric polynomials whose Fourier coefficients are supported on a set $T \subset [-q, q]^d \cap \mathbb{Z}^d$ satisfying $|T| \leq M$, i.e., $f \in \Pi_q(M)$ is of the form $f(x) = \sum_{k\in T} c_k e^{ik\cdot x}$. Note that $\Pi_q(M)$ is not a linear space.

Our aim is to sample a trigonometric polynomial f of $\Pi_q(M)$ at N randomly chosen points and try to reconstruct f from these samples. We model the sampling points $x_1, \ldots, x_N$ as independent random variables having the uniform distribution on $[0, 2\pi]^d$. We collect them into the sampling set

$$X := \{x_1, \ldots, x_N\}.$$

Obviously, the cardinality $|X|$ equals the number of samples N with probability 1.

Motivated by results of Candes, Romberg and Tao in [6] we propose the following non-linear method of reconstructing $f \in \Pi_d(M)$ from its sampled values $f(x_1), \ldots, f(x_N)$. We minimize the ℓ_1-norm of the Fourier coefficients c_k,

$$\|(c_k)\|_1 := \sum_{k\in[-q,q]^d} |c_k|,$$

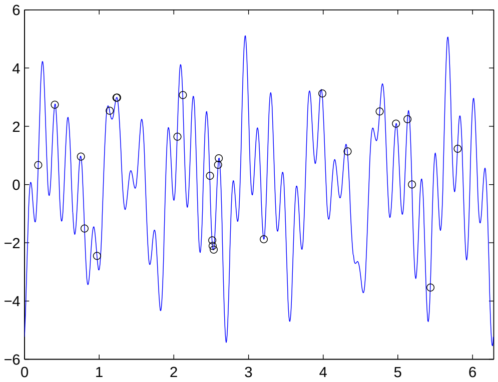
under the constraint that the corresponding trigonometric polynomial matches f on the sampling points. That is we solve the problem

$$\min \|(c_k)\|_1 \quad \text{subject to} \quad g(x_j) := \sum_{k\in[-q,q]^d} c_k e^{ik\cdot x_j} = f(x_j), \quad j = 1, \ldots, N. \tag{2.1}$$

This task – also referred to as Basis Pursuit [5] – can be performed with efficient convex optimization techniques [3], or even linear programming in case of real-valued coefficients c_k.

Once all the coefficients c_k, $k \in [-q, q]^d$, are known also f is known completely and can be evaluated efficiently at any point, e.g., with the Fourier transform for non-equispaced data developed by Daniel Potts et al. [16].

Surprisingly, there is numerical evidence that the above reconstruction scheme recovers f *exactly* provided the number of samples is large enough compared to the sparsity. Indeed, Figure 2.1 shows a sparse trigonometric polynomial with 8 non-zero coefficients and $N = 25$ sampling points while the maximal degree is $q = 40$, i.e., $D = 81$. The right hand side shows the



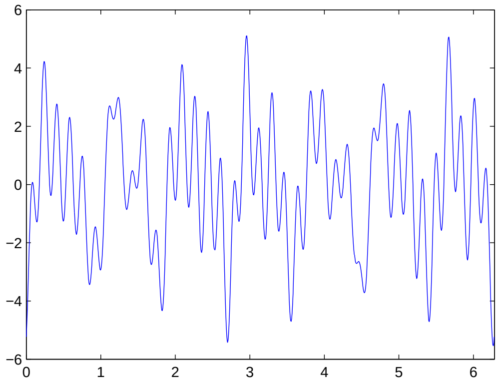


Figure 1: Original sparse trigonometric polynomial with samples (left) and reconstruction (right)

reconstruction from the samples by solving the minimization problem (2.1). The reconstruction is exact! We refer to Section 6 for more information on the numerical experiments.

Our main results are two theorems that give a theoretical explanation of this phenomenon. The first one treats any sparse polynomial in $\Pi_q^d(M)$ and the second one considers "generic" polynomials in the sense that the set T of non-vanishing coefficients is modelled as random set. Unexpectedly, both results involve combinatorial quantities connected to set partitions. We will spend the next section introducing the necessary notation.

2.2 Set Partitions

We denote $[n] := \{1, 2, \ldots, n\}$. A partition of $[n]$ is a set of subsets of $[n]$ – called blocks – such that each $j \in [n]$ is contained in precisely one of the subsets. By $P(n, k)$ we denote the set of all partitions of $[n]$ into exactly k blocks such that each block contains at least 2 elements. For example $P(4, 2)$ consists of

$$\{\{1,2\},\{3,4\}\}, \quad \{\{1,3\},\{2,4\}\} \quad \text{and} \quad \{\{1,4\},\{2,3\}\}.$$

Clearly, $P(n, k)$ is empty if $k > n/2$. The numbers $S_2(n, k) = |P(n, k)|$ are called associated Stirling numbers of the second kind. They have the following exponential generating function, see [17, formula (27), p.77],

$$\sum_{n=1}^{\infty} \sum_{k=1}^{\lfloor n/2 \rfloor} S_2(n,k) y^k \frac{x^n}{n!} \;=\; \exp(y(e^x - x - 1)). \tag{2.2}$$

Based on this one may deduce that the numbers $S_2(n, k)$ satisfy the recursion formula

$$S_2(n,k) \;=\; kS_2(n-1,k) + (n-1)S_2(n-2,k-1). \tag{2.3}$$

Also a combinatorial argument for this recursion exists, see Section 4 where also further information on the numbers $S_2(n, k)$ will be given.

We also need partitions of a different type. An adjacency is defined to be an occurence of two consecutive integers of $[n]$ in the same block of a partition. Hereby, consecutive is understood in the circular sense, i.e., also n and 1 are considered consecutive. We define $U(n,k)$ as the set of all partitions into k subsets having *no* adjacencies. For instance, $U(5,3)$ consists of the partitions

$$\begin{aligned}&\{\{1,4\},\{2,5\},\{3\}\}, \quad \{\{1,4\},\{2\},\{3,5\}\}, \quad \{\{1\},\{2,4\},\{3,5\}\},\\ &\{\{1,3\},\{2,5\},\{4\}\} \quad \text{and} \quad \{\{1,3\},\{2,4\},\{5\}\}.\end{aligned} \tag{2.4}$$

Clearly, $U(n,1)$ is empty. We remark that it was only very recently that D. Knuth [15] raised the problem of determining the number of partitions in $U(n,k)$.

We will also need a slight variation of this type of partitions. Let $[K]\times[m]=\{1,\ldots,K\}\times\{1,\ldots,m\}$ for some numbers $K,m\in\mathbb{N}$. We denote by $U^*(K,m,s)$ the set of all partitions of $[K]\times[m]$ such that (p,u) and $(p,u+1)$ are not contained in the same block for all $p=1,\ldots,K$ and $u=1,\ldots,m-1$. (So this sort of consecutiveness is *not* understood in the circular sense.) We remark that $U(K,1,k)$ is the set of all partitions of a K-element set into k subsets (without any restriction on the type of partition). In particular, the numbers $|U(K,1,k)|$ equal the (ordinary) Stirling numbers $S(K,k)$ of the second kind. The numbers $b_n:=\sum_{k=1}^n S(n,k)$ are called Bell numbers [17, 19].

Now let $\mathcal{A}=\{A_1,\ldots,A_t\}$ be a partition in $P(n,t)$ and $\mathcal{B}=\{B_1,\ldots,B_s\}\in U(n,s)$. By A_i+1 we understand the set whose elements are the ones of A_i incremented by 1 in the circular sense, i.e., $n+1\equiv 1$. We associate a $t\times s$ matrix $M=M(\mathcal{A},\mathcal{B})$ to the pair $\mathcal{A},\mathcal{B}$ by setting

$$M_{i,j} \;:=\; |A_i\cap B_j|-|(A_i+1)\cap B_j|, \qquad 1\le i\le t,\, 1\le j\le s. \tag{2.5}$$

Then we define $Q(n,t,s,R)$ to be the number of pairs of partitions $(\mathcal{A},\mathcal{B})$ with $\mathcal{A}\in P(n,t)$ and $\mathcal{B}\in U(n,s)$ such that the rank of $M(\mathcal{A},\mathcal{B})$ equals R, i.e.,

$$Q(n,t,s,R) \;:=\; \#\left\{(\mathcal{A},\mathcal{B}):\, \mathcal{A}\in P(n,t),\mathcal{B}\in U(n,s),\operatorname{rank}M(\mathcal{A},\mathcal{B})=R\right\}. \tag{2.6}$$

Observe that

$$\sum_{i=1}^t M_{i,j} \;=\; \sum_{i=1}^t \left(|A_i\cap B_j|-|(A_i+1)\cap B_j|\right) \;=\; |\{1,\ldots,n\}\cap B_j|-|\{1,\ldots,n\}\cap B_j| \;=\; 0$$

(since the A_i's are disjoint) and similarly $\sum_{j=1}^s M_{i,j}=0$. Thus, the rank of $M(\mathcal{A},\mathcal{B})$ is less or equal to $\min\{s,t\}-1$. In other words $Q(n,t,s,R)=0$ if $R\ge\min\{s,t\}$.

Similarly, let $(\mathcal{A},\mathcal{B})$ be a pair of partitions of $[K]\times[m]$ where $\mathcal{A}=\{A_1,\ldots,A_t\}\in P(Km,t)$ (identifying $[Km]$ with $[K]\times[m]$) and $\mathcal{B}=\{B_1,\ldots,B_s\}\in U^*(K,m,s)$. Let A_i-1 denote the sets whose elements are $\{(p,u-1),\,(p,u)\in A_i\}$. In contrast to above we do *not* calculate in the circular sense this time. So elements of the form $(p,0)$ may appear in A_i-1. Then to such a pair $(\mathcal{A},\mathcal{B})$ we associate a matrix $L=L(\mathcal{A},\mathcal{B})$ with entries

$$L_{i,j} \;=\; \sum_{(p,u)\in A_i\cap B_j}(-1)^p \;-\; \sum_{(p,u)\in(A_i-1)\cap B_j}(-1)^p. \tag{2.7}$$

Finally, we define

$$Q^*(K,m,t,s,R) \;:=\; \#\left\{(\mathcal{A},\mathcal{B}):\, \mathcal{A}\in P(Km,t),\mathcal{B}\in U^*(K,m,s),\operatorname{rank}L(\mathcal{A},\mathcal{B}) \;=\; R\right\}. \tag{2.8}$$

Later in Section 4 we will provide some more information on these combinatorial quantities.

2.3 The Main Theorems

In order to formulate our first theorem let $F_n(\theta), n \in \mathbb{N}$, denote the functions defined in terms of a generating function by

$$\sum_{n=1}^{\infty} F_n(\theta) \frac{x^n}{n!} \;=\; \exp(\theta(e^x - x - 1)). \tag{2.9}$$

Clearly, F_n is connected to the associated Stirling numbers of the second kind $S_2(n,k)$ by (2.2). We refer to Section 4 for a list of F_{2n} for $n = 1, \ldots, 6$. Further, we define

$$G_n(\theta) \;:=\; \theta^{-n} F_n(\theta).$$

Also recall that $D = (2q+1)^d$. Then our first theorem about exact reconstruction of sparse trigonometric polynomials reads as follows.

Theorem 2.1. *Assume $f \in \Pi_q^d(M)$ with some sparsity $M \in \mathbb{N}$. Let $x_1, \ldots, x_N \in [0, 2\pi]^d$ be independent random variables having the uniform distribution on $[0, 2\pi]^d$. Choose $n \in \mathbb{N}$, $\beta > 0, \kappa > 0$ and $K_1, \ldots, K_n \in \mathbb{N}$ such that*

$$a \;:=\; \sum_{m=1}^{n} \beta^{n/K_m} < 1 \qquad \text{and} \qquad \frac{\kappa}{1-\kappa} \;\leq\; \frac{1-a}{1+a} M^{-3/2}. \tag{2.10}$$

Set $\theta := N/M$. Then with probability at least

$$1 - \left(D\beta^{-2n} \sum_{m=1}^{n} G_{2mK_m}(\theta) + M\kappa^{-2} G_{2n}(\theta) \right) \tag{2.11}$$

f can be reconstructed exactly from its sample values $f(x_1), \ldots, f(x_N)$ by solving the minimization problem (2.1).

We will illustrate the probability bound (2.11) later in Section 5 with some plots. In particular, the probability of exact reconstruction is high if the "non-linear oversampling factor" $\theta = N/M$ is large enough. Of course, in order to obtain useful results one has to optimize with respect to the parameters occuring in (2.11). In particular, the choice of n is crucial. It may not be chosen too small but also not too large depending on θ. Indeed, pursuing this strategy leads to the following qualitative result.

Corollary 2.2. *There exists an absolute constant $C > 0$ such that the following is true. Assume $f \in \Pi_q^d(M)$ for some sparsity $M \in \mathbb{N}$. Let $x_1, \ldots, x_N \in [0, 2\pi]^d$ be independent random variables having the uniform distribution on $[0, 2\pi]^d$. If for some $\epsilon > 0$ it holds*

$$N \;\geq\; CM(\log D + \log(\epsilon^{-1}))$$

then with probability at least $1 - \epsilon$ the trigonometric polynomial f can be recovered from its sample values $f(x_j), j = 1, \ldots, N$, by solving the ℓ^1-minimization problem (2.1).

This formulation is similar to the main theorem in [6] concerned with exact reconstruction in the context of the discrete Fourier transform. Indeed, setting $\epsilon = D^{-\sigma}$ yields a probability of exact reconstruction of at least $1 - D^{-\sigma}$ provided $N \geq CM(\sigma + 1) \log D$.

We remark that (2.11) of Theorem 2.1 allows to actually compute precise bounds on the probability of exact reconstruction when the parameters M, N, D are given explicitly. But clearly, the previous corollary is easier to interpret. This is the reason why we have given both results.

For our next theorem we model also the set $T \subset [-q, q]^d$ of non-vanishing Fourier coefficients as random. So we will not treat arbitrary sparse polynomials, but only "generic" ones. The hope is, of course, that this provides even better estimates for the probability of exact reconstruction.

Let $0 < \tau < 1$. The probability that an index $k \in [-q, q]^d$ belongs to T is modelled as

$$\mathbb{P}(k \in T) \;=\; \tau \tag{2.12}$$

independently for each k. We also assume that the choice of T and the choice of the sampling set X are stochastically independent. Clearly, the expected size of T is $\mathbb{E}|T| = \tau D = \tau(2q+1)^d$ and $|T|$ follows the binomial distribution. For convenience we also introduce Π_T as the set of all trigonometric polynomials whose coefficients are supported on T.

We also need some auxiliary notation. For $n \in \mathbb{N}$ we define

$$W(n, N, \mathbb{E}|T|, D) \;:=\; N^{-2n} \sum_{t=1}^{\min\{n,N\}} \frac{N!}{(N-t)!} \sum_{s=2}^{2n} (\mathbb{E}|T|)^s \sum_{R=0}^{\min\{t,s\}-1} Q(2n, t, s, R) D^{-R} \tag{2.13}$$

and for $K, m \in \mathbb{N}$

$$Z(K, m, N, \mathbb{E}|T|, D) \;:=\; N^{-2Km} \sum_{t=1}^{\min\{Km,N\}} \frac{N!}{(N-t)!} \sum_{s=1}^{2Km} (\mathbb{E}|T|)^s \sum_{R=0}^{\min\{t,s\}} Q^*(2K, m, t, s, R) D^{-R}.$$

Our second theorem about reconstructing a sparse trigonometric polynomial from random samples by Basis Pursuit is given as follows.

Theorem 2.3. *Let $x_1, \dots, x_N \in [0, 2\pi]^d$ be independent random variables having the uniform distribution on $[0, 2\pi]^d$. Further assume that T is a random subset of $[-q, q]^d$ modelled by (2.12) (with T being independent of $x_1, \dots, x_N$) such that $\mathbb{E}|T| = \tau D \geq 1$. Choose $n \in \mathbb{N}$, $\alpha > 0, \beta > 0, \kappa > 0$ and $K_1, \dots, K_n \in \mathbb{N}$ such that*

$$a \;:=\; \sum_{m=1}^{n} \beta^{n/K_m} < 1 \quad \textit{and} \quad \frac{\kappa}{1-\kappa} \;\leq\; \frac{1-a}{1+a} \left((\alpha+1)\mathbb{E}|T|\right)^{-3/2}. \tag{2.14}$$

Then with probability at least

$$1 - \left(\kappa^{-2} W(n, N, \mathbb{E}|T|, D) + \beta^{-2n} D \sum_{m=1}^{n} Z(K_m, m, N, \mathbb{E}|T|, D) + \exp\left(-\frac{3\alpha^2}{6+2\alpha} \mathbb{E}|T| \right) \right) \tag{2.15}$$

any $f \in \Pi_T \subset \Pi_q^d(|T|)$ can be reconstructed exactly from its sample values $f(x_1), \dots, f(x_N)$ by solving the minimization problem (2.1).

Of course, the theorem has to be understood in the sense that the set T is not known a priori because with the knowledge of T it would be in fact much easier to reconstruct f.

(Although it seems that in higher dimensions $d \geq 2$ not many theoretical results are available, see e.g. [1].)

Like the previous theorem this result shows that the probability for reconstructing the original sparse polynomial is indeed high for appropriate choices for the number of sampling points N and the expected sparsity $\mathbb{E}|T|$. We will illustrate this later in Section 5 by computing numerical plots for the bound in (2.15). Since the theorem does not treat arbitrary f's but only "generic" ones in the sense that the set T is random one may expect that the bound for the probability for exact reconstruction is better than the one in Theorem 2.1. As we will see later, this is indeed the case if one takes the same n, see also Section 3.6. Unfortunately, we were not able to compute the bound explicitly for $n \geq 5$ so that practically up to now Theorem 2.1 gives the better bounds since here we are to evaluate the bound (2.11) for any n.

The reason for not being able to compute (2.15) for arbitrary n is due to the fact that we do not have an explicit expression (or a recursion formula, or a good estimation) of the numbers $Q(2n,t,s,R)$ and $Q^*(2K,m,t,s,R)$. We were only able to compute them on a computer up to $n = 4$ by checking the rank of $M(\mathcal{A},\mathcal{B})$ and $L(\mathcal{A},\mathcal{B})$ for all possible pairs $(\mathcal{A},\mathcal{B})$ of partitions. Already for $n = 5$ the computing times would exceed several days and with $n = 7$ at the latest the task nearly becomes an impossibility since the rank of $576535660478649 \approx 5.7 \times 10^{14}$ matrices would have to checked for computing the numbers $Q(14,t,s,R)$. So we have to leave it as an interesting open problem to provide more information on the numbers $Q(2n,t,s,R)$ and $Q^*(2K,m,t,s,R)$, see also Section 4. We hope that with a progress on this combinatorial problem we can improve significantly our probability bounds.

Remark 2.4. *(a) In both theorems it is reasonable to choose $K_m \approx \frac{m}{n}$, for instance rounding m/n to the nearest integer. In this way $mK_m \approx n$ for all m and further*

$$\sum_{m=1}^{n} \beta^{n/K_m} \approx \sum_{m=1}^{n} \beta^m \approx \frac{\beta}{1-\beta}.$$

Indeed, in the limit $n \to \infty$ all the above expressions become equal. As we require the left hand side to be less than 1, we should choose β approximately less than $1/2$. Actually, a choice near $1/2$ turned out to be good.

(b) There is nothing special about the underlying set $[-q,q]^d \cap \mathbb{Z}^d$. Indeed, both theorem still hold when taking any other finite subset of $\mathbb{Z}^d$ of size D instead.

(c) If one is interested in choosing the dimension $D = (2q+1)^d$ of the problem very large then one may observe that

$$\lim_{D\to\infty} W(n,N,\mathbb{E}|T|,D) = N^{-2n} \sum_{t=1}^{\min\{n,N\}} \frac{N!}{(N-t)!} \sum_{s=1}^{2n} Q(2n,t,s,0)\,(\mathbb{E}|T|)^s$$

and

$$\lim_{D\to\infty} Z(K,m,N,\mathbb{E}|T|,D) = N^{-2Km} \sum_{t=1}^{\min\{Km,N\}} \frac{N!}{(N-t)!} \sum_{s=1}^{2Km} Q(2K,m,t,s,0)\,(\mathbb{E}|T|)^s.$$

(Of course, we keep $\mathbb{E}|T|$ fixed in this limit so that $\tau = D/\mathbb{E}|T|$, see (2.12), has to be adjusted in the process of passing with D to infinity.) This shows that the numbers

$Q(2n,t,s,0)$ and $Q^*(2K,m,t,s,0)$ play the most important role in the probibility bound (2.15) of Theorem 2.3. In fact, the tables in the Appendix and Lemma 4.1 indicate that these numbers are quite small for $R=0$ compared to other values of R.

(d) *In practice, we usually do not have precisely sparse signals. However, signals that can be approximated by sparse ones may appear quite frequently (e.g. in the context of best n-term approximation). We leave the investigation of related questions to future contributions, see also [7] for the setting of the discrete Fourier transform.*

3 Proof of the Main Results

We will develop the proofs of both theorems in parallel. The basic idea is similar as in the paper [6] by Candes, Romberg and Tao. However, there are also significant differences and, in particular, it turns out that our approach leads to a simpler and slightly less technical proof (although still considerably elaborate). Also the idea of modelling the "sparsity set" as random is new and requires special treatment.

Let us first introduce some auxiliary notation. By $\ell^2([-q,q]^d), \ell^2(T), \ell^2(X)$ we denote the ℓ^2 space of sequences indexed by $[-q,q]^d$, $T \subset [-q,q]^d$ and X, respectively, endowed with the usual Euclidean norm. Moreover, we introduce the operator

$$\mathcal{F}_X : \ell^2([-q,q]^d) \to \ell^2(X), \qquad \mathcal{F}_X c(x_j) := \sum_{k \in [-q,q]^d} c_k e^{ik \cdot x_j}, \quad j = 1, \ldots, N.$$

By $\mathcal{F}_{TX} : \ell^2(T) \to \ell^2(X)$ we denote the restriction of $\mathcal{F}_X$ to sequences supported only on T. The adjoint operators are denoted by $\mathcal{F}_X^* : \ell^2(X) \to \ell^2([-q,q]^d)$ and $\mathcal{F}_{TX}^* : \ell^2(X) \to \ell^2(T)$.

Clearly, our problem is equivalent to reconstructing a sequence c from $\mathcal{F}_X c$ by solving the problem

$$\min \|c'\|_1 \quad \text{subject to} \quad \mathcal{F}_X c' = \mathcal{F}_X c. \tag{3.1}$$

For $c \in \ell^2([-q,q]^d)$ we introduce its sign by

$$\operatorname{sgn}(c)_k = \frac{c_k}{|c_k|}, \quad k \in \operatorname{supp} c, \quad \text{and} \quad \operatorname{sgn}(c)_k = 0, \quad k \notin \operatorname{supp} c.$$

Hereby, $\operatorname{supp} c$ denotes the support of c.

The key lemma for our proofs is the following duality principle.

Lemma 3.1. *Let $c \in \ell^2([-q,q]^d)$ and $T := \operatorname{supp} c$. Assume $\mathcal{F}_{TX} : \ell^2(T) \to \ell^2(X)$ to be injective. Suppose that there exists a vector $P \in \ell^2([-q,q]^d)$ with the following properties:*

(i) *$P_k = \operatorname{sgn} c_k$ for all $k \in T$,*

(ii) *$|P_k| < 1$ for all $k \notin T$,*

(iii) *there exists a vector $\lambda \in \ell^2(X)$ such that $P = \mathcal{F}_X^* \lambda$.*

Then c is the unique minimizer to the problem (3.1).

Proof: The proof mimiques the one by Candes, Romberg and Tao [6, Lemma 2.1]. For the sake of completeness we repeat the argument.

Let b be a vector with $\mathcal{F}_X b = \mathcal{F}_X c$ and set $h := b - c$. Clearly, $\mathcal{F}_X h$ vanishes. For any $k \in T$ we have

$$|b_k| \;=\; |c_k + h_k| \;=\; ||c_k| + h_k\overline{\mathrm{sgn}(c)_k}| \;\geq\; |c_k| + \mathrm{Re}(h_k\overline{\mathrm{sgn}(c)_k}) \;=\; |c_k| + \mathrm{Re}(h_k\overline{P_k}).$$

If $k \notin T$ then $|b_k| = |h_k| \geq \mathrm{Re}(h_k\overline{P_k})$ since $|P_k| < 1$. This gives

$$\|b\|_1 \;\geq\; \|c\|_1 + \sum_{k\in[-q,q]^d} \mathrm{Re}(h_k\overline{P_k}).$$

Further, observe that

$$\sum_{k\in[-q,q]^d} \mathrm{Re}(h_k\overline{P_k}) \;=\; \mathrm{Re}\left(\sum_{k\in[-q,q]^d} h_k\overline{(\mathcal{F}_X^*\lambda)_k}\right) \;=\; \mathrm{Re}\left(\sum_{j=1}^N (\mathcal{F}_X h)(x_j)\overline{\lambda(x_j)}\right) \;=\; 0$$

since $\mathcal{F}_X h$ vanishes. Altogether, we proved $\|b\|_1 \geq \|c\|_1$, and thus c is a minimizer of (3.1).

It remains to prove the uniqueness. The above argument shows that having the equality $\|b\|_1 = \|c\|_1$ forces $|h_k| = \mathrm{Re}(h_k\overline{P_k})$ for all $k \notin T$. Since $|P_k| < 1$ this means that h vanishes outside T. Since also $\mathcal{F}_X h$ vanishes, it follows from the injectivity of $\mathcal{F}_{TX}$ that h vanishes identically and hence, $b = c$. This shows that c is the unique minimizer of (3.1). ■

Concerning the assumption on the injectivity of $\mathcal{F}_{TX}$ we have the following simple result.

Lemma 3.2. *If $N \geq |T|$ then $\mathcal{F}_{TX}$ is injective almost surely.*

Proof: The proof is essentially contained in [1, Theorem 3.2]. There it is proved that any $|T| \times |T|$ submatrix of $\mathcal{F}_{TX}$ has non-vanishing determinant almost surely (even under slightly more general assumptions on the distribution of the random variables $x_1, \ldots, x_N$). This implies the result. ■

Now our strategy for proving Theorem 2.3 is obvious. We need to show that with high probability there exists a vector P with the properties assumed in Lemma 3.1. To this end we proceed similarly as in [6]. (Actually, the injectivity of $\mathcal{F}_{TX}$ will also follow from this finer analysis so that Lemma 3.2 will not be needed in the end.)

We introduce the restriction operator $R_T : \ell^2([-q,q]^d) \to \ell^2(T)$, $R_T c_k = c_k$ for $k \in T$. Its adjoint $R_T^* = E_T : \ell^2(T) \to \ell^2([-q,q]^d)$ is the operator that extends a vector outside T by zero, i.e., $(E_T d)_k = d_k$ for $k \in T$ and $(E_T d)_k = 0$ otherwise.

Now assume for the moment that $\mathcal{F}_{TX}^*\mathcal{F}_{TX} : \ell^2(T) \to \ell^2(T)$ is invertible. (By Lemma 3.2 this is true almost surely if $N \geq |T|$ since $\mathcal{F}_{TX}$ is then injective.) In this case we define P explicitly by

$$P \;:=\; \mathcal{F}_X^*\mathcal{F}_{TX}(\mathcal{F}_{TX}^*\mathcal{F}_{TX})^{-1}R_T\,\mathrm{sgn}(c),$$

where as before $T = \operatorname{supp} c$. Then clearly P has property (i) and property (iii) in Lemma 3.1 with

$$\lambda := \mathcal{F}_{TX}(\mathcal{F}_{TX}^*\mathcal{F}_{TX})^{-1}R_T\,\mathrm{sgn}(c) \in \ell^2(X).$$

We are left with proving that P has property (ii) of Lemma 3.1 with high probability.

To this end we introduce the auxiliary operators

$$H : \ell^2(T) \to \ell^2([-q,q]^d), \qquad H \;:=\; NE_T - \mathcal{F}_X^*\mathcal{F}_{TX}$$

and

$$H_0 : \ell^2(T) \to \ell^2(T), \qquad H_0 \;:=\; R_T H \;=\; N I_T - \mathcal{F}_{TX}^* \mathcal{F}_{TX},$$

where I_T denotes the identity on $\ell^2(T)$. Obviously, H_0 is self-adjoint, and H acts on a vector as

$$(Hc)_\ell \;=\; -\sum_{j=1}^{N} \sum_{\substack{k \in T \\ k \neq \ell}} c_k e^{i(k-\ell)\cdot x_j}.$$

Now we can write

$$P \;=\; (NE_T - H)\,(NI_T - H_0)^{-1}\, R_T \operatorname{sgn}(c).$$

As we are interested in property (ii) in Lemma 3.1 we consider only values of P on $T^c = [-q,q]^d \setminus T$. Since $R_{T^c} E_T = 0$ we have

$$P_k \;=\; -\frac{1}{N} R_{T^c} H (I_T - \frac{1}{N} H_0)^{-1} R_T \operatorname{sgn}(c) \quad \text{ for all } k \in T^c.$$

Let us look closer at the term $(I_T - \frac{1}{N} H_0)^{-1}$. To this end let $n \in \mathbb{N}$ be some arbitrary number. By the von Neumann series we can write

$$\left(I_T - \left(\frac{1}{N} H_0 \right)^n \right)^{-1} \;=\; I_T + A_n$$

with

$$A_n \;:=\; \sum_{r=1}^{\infty} (\frac{1}{N} H_0)^{rn}. \tag{3.2}$$

Using the identity

$$(1-M)^{-1} \;=\; (1-M^n)^{-1}(1 + M + \cdots + M^{n-1}) \tag{3.3}$$

we obtain

$$(I_T - \frac{1}{N} H_0)^{-1} \;=\; (I_T + A_n) \sum_{m=0}^{n-1} \left(\frac{1}{N} H_0 \right)^m.$$

Thus, on the complement of T, we may write

$$\begin{aligned}
R_{T^c} P \;&=\; -\frac{1}{N} H (I_T + A_n) \left(\sum_{m=0}^{n-1} (N^{-1} H_0)^m \right) R_T \operatorname{sgn}(c) \\
&=\; -\sum_{m=1}^{n} (N^{-1} H R_T)^m \operatorname{sgn}(c) - \frac{1}{N} H A_n R_T \sum_{m=0}^{n-1} (N^{-1} H R_T)^m \operatorname{sgn}(c) \\
&=\; -(P^{(1)} + P^{(2)}),
\end{aligned}$$

where

$$P^{(1)} \;=\; S_n \operatorname{sgn}(c), \quad \text{ and } \quad P^{(2)} \;=\; \frac{1}{N} H A_n R_T (I + S_{n-1}) \operatorname{sgn}(c),$$

with

$$S_n \;:=\; \sum_{m=1}^{n} (N^{-1} H R_T)^m.$$

Our aim is to estimate $\mathbb{P}(\sup_{k\in T^c}|P_k|\geq 1)$. To this end let $a_1,a_2>0$ be numbers satisfying $a_1+a_2=1$. Then

$$\mathbb{P}(\sup_{k\in T^c}|P_k|\geq 1)\;\leq\;\mathbb{P}\left(\{\sup_{k\in T^c}|P_k^{(1)}|\geq a_1\}\cup\{\sup_{k\in T^c}|P_k^{(2)}|\geq a_2\}\right). \tag{3.4}$$

Clearly,

$$\begin{aligned}\mathbb{P}(|P_k^{(1)}|\geq a_1) \;&=\; \mathbb{P}\left(\left|\left(\sum_{m=1}^{n}(N^{-1}HR_T)^m\operatorname{sgn}(c)\right)_k\right|\geq a_1\right)\\ &\leq\;\mathbb{P}\left(\sum_{m=1}^{n}|((N^{-1}HR_T)^m\operatorname{sgn}(c))_k|\geq a_1\right)\;=:\;\mathbb{P}(E_k).\end{aligned}\tag{3.5}$$

Consider $P^{(2)}$. Denoting $\ell^\infty=\ell^\infty([-q,q]^d)$ the space of sequences indexed by $[-q,q]^d$ with the supremum norm (and similarly defining $\ell^\infty(T)$) we have

$$\sup_{k\in T^c}|P_k^{(2)}|\leq\|P^{(2)}\|_\infty\leq\|\frac{1}{N}HA_n\|_{\ell^\infty(T)\to\ell^\infty}(1+\|R_TS_{n-1}\operatorname{sgn}(c)\|_{\ell^\infty(T)}) \tag{3.6}$$

In order to analyze the term $\|R_TS_{n-1}\operatorname{sgn}(c)\|_{\ell^\infty(T)}$ we observe that similarly as in (3.5)

$$\mathbb{P}(|(S_{n-1}\operatorname{sgn}(c))_k|\geq a_1)\;\leq\;\mathbb{P}\left(\sum_{m=1}^{n}|((N^{-1}HR_T)^m\operatorname{sgn}(c))_k|\geq a_1\right)\;=\;\mathbb{P}(E_k).$$

Let us now treat the operator norm appearing in (3.6). For simplicity we write $\|\cdot\|_\infty$ instead of $\|\cdot\|_{\ell^\infty\to\ell^\infty}$. It holds $\|A\|_\infty=\sup_r\sum_s|A_{rs}|$. Clearly,

$$\|\frac{1}{N}HA_n\|_{\ell^\infty}\leq\|\frac{1}{N}H\|_\infty\|A_n\|_{\ell^\infty(T)}.$$

Moreover, $\|\frac{1}{N}H\|_\infty\leq|T|$ as H has $|T|$ columns and each entry is bounded by N in absolute value.

In order to analyze A_n we will work with the Frobenius norm. For a matrix A it is defined as

$$\|A\|_F^2\;:=\;\operatorname{Tr}(AA^*)\;=\;\sum_{r,s}|A_{rs}|^2$$

where $\operatorname{Tr}(AA^*)$ denotes the trace of AA^*. Assume for the moment that

$$\|(\frac{1}{N}H_0)^n\|_F\;\leq\;\kappa<1. \tag{3.7}$$

Then it follows directly from the definition (3.2) of A_n that

$$\|A_n\|_F\;=\;\left\|\sum_{r=1}^{\infty}\left(\frac{1}{N}H_0\right)^{rn}\right\|_F\;\leq\;\sum_{r=1}^{\infty}\|(N^{-1}H_0)^n\|_F^r\;\leq\;\sum_{r=1}^{\infty}\kappa^r\;=\;\frac{\kappa}{1-\kappa}.$$

Moreover, since A_n has $|T|$ columns it follows from the Cauchy-Schwarz inequality that

$$\|A_n\|_\infty^2\;\leq\;\sup_i|T|\sum_j|A_n(i,j)|^2\;\leq\;|T|\|A_n\|_F^2.$$

So assuming (3.7) and $\|S_{n-1}\operatorname{sgn}(c)\|_\infty < a_1$ we have

$$\sup_{t\in T^c}|P_k^{(2)}| \le (1+a_1)|T|^{3/2}\frac{\kappa}{1-\kappa}.$$

In particular, if

$$\frac{\kappa}{1-\kappa} \le \frac{a_2}{1+a_1}|T|^{-3/2} \tag{3.8}$$

then $\sup_{t\in T^c}|P_k^{(2)}| \le a_2$ as desired. Also it follows from (3.8) that $\kappa < 1$ as $|T| \ge 1$ without loss of generality (if $T = \emptyset$ then $f = 0$ and ℓ^1-minimization will clearly recover f.)

Now we have to distinguish between the situation in Theorem 2.1 and the one in Theorem 2.3 since in the latter $|T|$ is a random variable while in the first it is deterministic.

1. Let us first treat the case of Theorem 2.3 where $|T|$ is random. If

$$|T| \le (\alpha+1)\mathbb{E}|T|$$

with $\alpha > 0$ and

$$\frac{\kappa}{1-\kappa} \le \frac{a_2}{1+a_1}((\alpha+1)\mathbb{E}|T|)^{-3/2} \tag{3.9}$$

then clearly (3.8) is satisfied and consequently

$$\sup_{t\in T^c}|P_k^{(2)}| \le a_2.$$

Using the union bound we altogether obtain from (3.4)

$$\begin{aligned}\mathbb{P}(\sup_{k\in T^c}|P_k| \ge 1) &\le \mathbb{P}\Bigg(\bigcup_{k\in T^c}\{|P_k^{(1)}| \ge a_1\} \cup \{\|R_T\operatorname{sgn}(c)\|_{\ell^\infty(T)} \ge a_1\} \cup \{\|(N^{-1}H_0)^n\|_F \ge \kappa\} \\ &\qquad \cup \{|T| \ge (\alpha+1)\mathbb{E}|T|\}\Bigg). \\ &\le \mathbb{P}\Bigg(\bigcup_{k\in[-q,q]^d} E_k \cup \{\|(N^{-1}H_0)^n\|_F \ge \kappa\} \cup \{|T| \ge (\alpha+1)\mathbb{E}|T|\}\Bigg) \\ &\le \sum_{k\in[-q,q]^d}\mathbb{P}(E_k) + \mathbb{P}(\|(N^{-1}H_0)^n\|_F \ge \kappa) + \mathbb{P}(|T| \ge (\alpha+1)\mathbb{E}|T|). \end{aligned} \tag{3.10}$$

As $|T|$ is the sum of independent random variables we obtain for the third term from the large deviation theorem (see for instance equation (6) in [2], where also slightly better estimates are available)

$$\begin{aligned}\mathbb{P}(|T| \ge \mathbb{E}|T| + \alpha\mathbb{E}|T|) &\le \exp\big(-(\alpha\mathbb{E}|T|)^2/(2\mathbb{E}|T| + 2(\alpha\mathbb{E}|T|)/3)\big) \\ &= \exp\left(-\frac{3\alpha^2}{6+2\alpha}\mathbb{E}|T|\right). \end{aligned} \tag{3.11}$$

So we are left with the two other expressions in (3.10).

2. In the situation of Theorem 2.1 we proceed in almost the same way with the only difference that we do not need to treat $|T|$ as random variable. Under the condition in (3.8) this yields

$$\mathbb{P}(\sup_{k\in T^c}|P_k|\geq 1)\;\leq\;\sum_{k\in[-q,q]^d}\mathbb{P}(E_k)+\mathbb{P}(\|(N^{-1}H_0)^n\|_F\geq\kappa). \tag{3.12}$$

Hence, also here we need to estime $\mathbb{P}(E_k)$ and $\mathbb{P}(\|(N^{-1}H_0)^n\|_F\geq\kappa)$.

3.1 Analysis of powers of H_0

In this section we treat the second term in (3.10) and (3.12), i.e., we estimate powers of the random matrix H_0 in the Frobenius norm. To this end Markov's inequality suggests to estimate the expectation of $\|H_0^n\|_F^2$. In the following lemma we only take the expectation with respect to the random sampling set $X=\{x_1,\dots,x_N\}$. For the situation of Theorem 2.3 we postpone the computation of the full expectation $\mathbb{E}=\mathbb{E}_T\mathbb{E}_X$ (the latter by Fubini's theorem).

Lemma 3.3. *It holds*

$$\mathbb{E}_X\left[\|H_0^n\|_F^2\right]\;=\;\sum_{t=1}^{\min\{n,N\}}\frac{N!}{(N-t)!}\sum_{\mathcal{A}\in P(2n,t)}\sum_{\substack{k_1,\dots,k_{2n}\in T\\ k_j\neq k_{j+1},j\in[2n]}}\prod_{A\in\mathcal{A}}\delta\left(\sum_{r\in A}(k_{r+1}-k_r)\right)$$

where $\delta(n)$ denotes the Kronecker δ_{0n} and $k_{2n+1}=k_1$.

<u>**Proof:**</u> As H_0 is self-adjoint we need to estimate $\|H_0^n\|_F^2=\mathrm{Tr}(H_0^{2n})$. Observe that

$$H_0(k,k')\;=\;h(k'-k),\qquad k,k'\in T$$

with

$$h(k)\;=\;-\delta(k)\sum_{\ell=1}^{N}e^{ik\cdot x_\ell}.$$

Thus, $H_0^2(k,k')\;=\;\sum_{t\in T,t\neq k,k'}h(t-k)h(k'-t)$ and

$$H_0^{2n}(k_1,k_1)\;=\;\sum_{k_2,\dots,k_{2n}\in T,k_j\neq k_{j+1}}h(k_2-k_1)\cdots h(k_1-k_{2n})$$

where we agree on the convention that $k_{2n+1}=k_1$. This yields

$$\mathrm{Tr}(H_0^{2n})\;=\;\sum_{k_1,\dots,k_{2n}\in T,k_j\neq k_{j+1}}h(k_2-k_1)h(k_3-k_2)\cdots h(k_1-k_{2n}).$$

Using linearity of expectation and the definition of h we get

$$\mathbb{E}_X\left[\mathrm{Tr}(H_0^{2n})\right]\;=\;\sum_{\ell_1,\dots,\ell_{2n}=1}^{N}\sum_{\substack{k_1,\dots,k_{2n}\in T\\ k_j\neq k_{j+1}}}\mathbb{E}_X\left[\exp\left(i\sum_{r=1}^{2n}(k_{r+1}-k_r)\cdot x_{\ell_r}\right)\right].$$

Let us consider the latter expected value. Here we have to take into accound that some of the indeces ℓ_r might be the same. This is where set partitions enter the game.

We associate a partition $\mathcal{A} = (A_1, \ldots, A_t)$ of $\{1, \ldots, 2n\}$ to a certain vector $(\ell_1, \ldots, \ell_{2n})$ such that $\ell_r = \ell_{r'}$ if and only if r and r' are contained in the same set $A_i \in \mathcal{A}$. This is allows us to unambiguously write ℓ_A instead of ℓ_r if $r \in A$. The independence of the x_{ℓ_A} yields

$$\begin{aligned}\mathbb{E}_X \left[\exp\left(i \sum_{r=1}^{2n} (k_{r+1} - k_r) \cdot x_{\ell_r} \right) \right] &= \mathbb{E}_X \left[\exp\left(i \sum_{A \in \mathcal{A}} \sum_{r \in A} (k_{r+1} - k_r) \cdot x_{\ell_A} \right) \right] \\ &= \prod_{A \in \mathcal{A}} \mathbb{E}_X \left[\exp\left(i \sum_{r \in A} (k_{r+1} - k_r) \cdot x_{\ell_A} \right) \right]. \end{aligned} \tag{3.13}$$

Since x_{ℓ_A} has the uniform distribution on $[0, 2\pi]^d$ we obtain

$$\begin{aligned}\mathbb{E}_X \left[\exp\left(i \sum_{r \in A} (k_{r+1} - k_r) \cdot x_{\ell_A} \right) \right] &= \int_{[0,2\pi]^d} \exp\left(i \sum_{r \in A} (k_{r+1} - k_r) \cdot x \right) dx \\ &= \delta\left(\sum_{r \in A} (k_{r+1} - k_r) \right). \end{aligned} \tag{3.14}$$

Observe that the last expression is independent of the precise values of the ℓ_r. Only the generated partition $\mathcal{A}$ plays a role. Moreover, if $A \in \mathcal{A}$ contains only one element then (3.14) vanishes due to the condition $k_{r+1} \neq k_r$. Thus, we only need to consider partitions $\mathcal{A}$ satisfying $|A| \geq 2$ for all $A \in \mathcal{A}$, i.e., partitions in $P(2n, t)$. Moreover, observe that the number of vectors $(\ell_{A_1}, \ldots, \ell_{A_t}) \in \{1, \ldots, N\}^t$ with different entries is precisely $N \cdots (N - t + 1) = N!/(N - t)!$ if $N \geq t$ and 0 if $N \leq t$. Finally, we obtain

$$\mathbb{E}_X[\|H_0^n\|_F^2] = \sum_{t=1}^{\min\{n,N\}} \sum_{\mathcal{A} \in P(2n,t)} \sum_{\substack{k_1,\ldots,k_{2n} \in T \\ k_j \neq k_{j+1}}} \prod_{A \in \mathcal{A}} \delta\left(\sum_{r \in A} (k_{r+1} - k_r) \right),$$

which is precisely the content of the lemma. ∎

In view of the previous lemma we define for simplicity and later reference

$$C(\mathcal{A}, T) := \sum_{\substack{k_1,\ldots,k_{2n} \in T \\ k_j \neq k_{j+1}}} \prod_{A \in \mathcal{A}} \delta\left(\sum_{r \in A} (k_{r+1} - k_r) \right) \tag{3.15}$$

$$= \#\left\{ (k_1, \ldots, k_{2n}) \in T^{2n} : k_j \neq k_{j+1}, j \in [2n], \text{ and } \sum_{r \in A} (k_{r+1} - k_r) = 0 \text{ for all } A \in \mathcal{A} \right\}.$$

3.2 Analysis of $\mathbb{P}(E_k)$

Let us now treat the first term $\mathbb{P}(E_k)$ in (3.10) resp. (3.12). To this end let $\beta_m, m = 1, \ldots, n$, be positive numbers satisfying

$$\sum_{m=1}^{n} \beta_m = a_1$$

and $K_m \in \mathbb{N}$, $m = 1, \ldots, n$, some natural numbers. Let $k \in [-q, q]^d$. Using Markov's inequality in the last step we obtain

$$\begin{aligned}
\mathbb{P}(E_k) &= \mathbb{P}\left(\sum_{m=1}^{n} |((N^{-1}HR_T)^m \operatorname{sgn}(c))_k| \geq a_1\right) \leq \sum_{m=1}^{n} \mathbb{P}(N^{-m}|((HR_T)^m \operatorname{sgn}(c))_k| \geq \beta_m) \\
&= \sum_{m=1}^{n} \mathbb{P}\left(N^{-2mK_m}|((HR_T)^m \operatorname{sgn}(c))_k|^{2K_m} \geq \beta_m^{2K_m}\right) \\
&\leq \sum_{m=1}^{n} \mathbb{E}\left[|((HR_T)^m \operatorname{sgn}(c))_k|^{2K_m}\right] N^{-2mK_m}\beta_m^{-2K_m}. \qquad (3.16)
\end{aligned}$$

Let us choose $\beta_m = \beta^{n/K_m}$, i.e., $\beta_m^{-2K_m} = \beta^{-2n}$. This yields

$$\mathbb{P}(E_k) \leq \beta^{-2n} \sum_{m=1}^{n} \mathbb{E}\left[|((HR_T)^m \operatorname{sgn}(c))_k|^{2K_m}\right] N^{-2mK_m} \qquad (3.17)$$

and the condition $a_1 = \sum_{m=1}^{n} \beta_m$ reads

$$a_1 = a = \sum_{m=1}^{n} \beta^{n/K_m} < 1.$$

The following lemma is concerned with the expectation appearing in (3.17). We first investigate the expectation with respect to X. The following proof is similar to the one of Lemma 3.3.

Lemma 3.4. *For $k \in [-q, q]^d$ and $c \in \ell^2([-q, q]^d)$ with $\operatorname{supp} c = T$ we have*

$$\begin{aligned}
&\mathbb{E}_X\left[|((HR_T)^m \operatorname{sgn}(c))_k|^{2K}\right] \\
&\leq \sum_{t=1}^{\min\{Km,N\}} \frac{N!}{(N-t)!} \sum_{\mathcal{A}\in P(2Km,t)} \sum_{\substack{k_1^{(1)},\ldots,k_m^{(1)}\in T \\ \vdots \\ k_1^{(2K)},\ldots,k_m^{(2K)}\in T \\ k_{j-1}^{(p)}\neq k_j^{(p)},\, j\in[m]}} \prod_{A\in\mathcal{A}} \delta\left(\sum_{(r,p)\in A} (-1)^p (k_r^{(p)} - k_{r-1}^{(p)})\right)
\end{aligned}$$

with $k_0^{(p)} := k$ for $p = 1, \ldots, 2K$. Hereby, we identify partitions of $[2Km]$ in $P(2Km, t)$ with partitions of $[2K] \times [m]$ in an obvious way.

<u>**Proof:**</u> Set $\sigma := \operatorname{sgn}(c)$. An elementary calculation yields

$$((HR_T)^m\sigma)_k = (-1)^m \sum_{\ell_1,\ldots,\ell_m=1}^{N} \sum_{\substack{k_1,\ldots,k_m\in T \\ k_{j-1}\neq k_j,\, j=1,\ldots,m}} \sigma(k_m)e^{i(k_m-k_{m-1})\cdot x_{\ell_m}} \cdots e^{i(k_1-k_0)\cdot x_{\ell_1}}$$

with $k_0 := k$. Thus,

$$\begin{aligned}
|((HR_T)^m\sigma)_{k_0}|^2 &= \sum_{\ell_1^{(1)},\ldots,\ell_m^{(1)}=1}^{N} \sum_{\ell_1^{(2)},\ldots,\ell_m^{(2)}=1}^{N} \sum_{\substack{k_1^{(1)},\ldots,k_m^{(1)}\in T \\ k_1^{(2)},\ldots,k_m^{(2)}\in T \\ k_{j-1}^{(p)}\neq k_j^{(p)},\, j\in[m],\, p=1,2}} \sigma(k_m^{(1)})\overline{\sigma(k_m^{(2)})} \times \\
&\times e^{i\sum_{r=1}^{m}(k_r^{(1)}-k_{r-1}^{(1)})\cdot x_{\ell_r^{(1)}}} e^{-i\sum_{r=1}^{m}(k_r^{(2)}-k_{r-1}^{(2)})\cdot x_{\ell_r^{(2)}}}
\end{aligned}$$

where $k_0^{(1)} = k_0^{(2)} = k_0 = k$. Taking a $2K$-th power yields

$$
\begin{aligned}
|((HR_T)^m \operatorname{sgn}(c))_k|^{2K} \ = & \sum_{\substack{\ell_1^{(1)},\dots,\ell_m^{(1)}=1 \\ \vdots \\ \ell_1^{(2K)},\dots,\ell_m^{(2K)}=1}}^{N} \ \sum_{\substack{k_1^{(1)},\dots,k_m^{(1)}\in T \\ \vdots \\ k_1^{(2K)},\dots,k_m^{(2K)}\in T \\ k_{j-1}^{(p)}\neq k_j^{(p)}}} \sigma(k_m^{(1)})\overline{\sigma(k_m^{(2)})}\cdots\sigma(k_m^{(2K-1)})\overline{\sigma(k_m^{(2K)})}\times \\
& \times \exp\left(i \sum_{p=1}^{2K} (-1)^p \sum_{r=1}^{m} (k_r^{(p)} - k_{r-1}^{(p)}) \cdot x_{\ell_r^{(p)}} \right)
\end{aligned}
$$

with $k_0^{(p)} = k$, $p = 1, \dots, 2K$. Further, recall that $|\sigma(k)| = 1$ on T. Taking the expected value $\mathbb{E}_X$ yields

$$
\begin{aligned}
& \mathbb{E}_X \left[|((HR_T)^m \operatorname{sgn}(c))_k|^{2K} \right] \\
& \quad \leq \sum_{\substack{\ell_1^{(1)},\dots,\ell_m^{(1)}=1 \\ \vdots \\ \ell_1^{(2K)},\dots,\ell_m^{(2K)}=1}}^{N} \ \sum_{\substack{k_1^{(1)},\dots,k_m^{(1)}\in T \\ \vdots \\ k_1^{(2K)},\dots,k_m^{(2K)}\in T \\ k_{j-1}^{(p)}\neq k_j^{(p)}}} \mathbb{E}_X \left[\exp\left(i \sum_{p=1}^{2K} (-1)^p \sum_{r=1}^{m} (k_r^{(p)} - k_{r-1}^{(p)}) \cdot x_{\ell_r^{(p)}} \right) \right] \qquad (3.18)
\end{aligned}
$$

(with equality if all the entries of σ are equal on T).

Let us consider the expected value appearing in the sum. As in the proof of Lemma 3.3 we have to take into account that some of the indeces $\ell_r^{(p)}$ might coincide. This affords to introduce some additional notation. Let $(\ell_r^{(p)})_{r=1,\dots,m}^{p=1,\dots,2K} \subset \{1, \dots, N\}^{2Km}$ be some vector of indeces and let $\mathcal{A} = (A_1, \dots, A_t)$, $A_i \subset \{1, \dots, m\} \times \{1, \dots 2K\}$ be a corresponding partition such that (r, p) and (r', p') are contained in the same block if and only if $\ell_r^{(p)} = \ell_{r'}^{(p')}$. For some $A \in \mathcal{A}$ we may unambigously write ℓ_A instead of $\ell_r^{(p)}$ if $(r, p) \in A$.

Like in (3.13), using that all ℓ_A for $A \in \mathcal{A}$ are different and that the x_{ℓ_A} are independent we may write the expectation in the sum in (3.18) as

$$
\begin{aligned}
& \mathbb{E} \left[\exp\left(i \sum_{p=1}^{2K} (-1)^p \sum_{r=1}^{m} (k_r^{(p)} - k_{r-1}^{(p)}) \cdot x_{\ell_r^{(p)}} \right) \right] \\
& \quad = \prod_{A\in\mathcal{A}} \mathbb{E} \left[\exp\left(i \sum_{(r,p)\in A} (-1)^p (k_r^{(p)} - k_{r-1}^{(p)}) \cdot x_{\ell_A} \right) \right] \ = \ \prod_{A\in\mathcal{A}} \delta \left(\sum_{(r,p)\in A} (-1)^p (k_r^{(p)} - k_{r-1}^{(p)}) \right).
\end{aligned}
$$

Once again, if $A \in \mathcal{A}$ contains only one element then the last expression vanishes due to the condition $k_r^{(p)} \neq k_{r-1}^{(p)}$. Thus, we only need to consider partitions $\mathcal{A}$ in $P(2Km, t)$. Now we are able to rewrite the inequality in (3.18) as

$$
\begin{aligned}
&\mathbb{E}_X\left[|((HR_T)^m\sigma)_k|^{2K}\right] \\
&\le \sum_{t=1}^{Km}\sum_{\mathcal{A}\in P(2Km,t)}\sum_{\substack{\ell_{(1)},\dots,\ell_{(t)}=1\\ \ell_{(1)},\dots,\ell_{(t)}\text{p.w. different}}}^{N}\sum_{\substack{k_1^{(1)},\dots,k_m^{(1)}\in T\\ \vdots\\ k_1^{(2K)},\dots,k_m^{(2K)}\in T\\ k_{j-1}^{(p)}\neq k_j^{(p)}}}\prod_{A\in\mathcal{A}}\delta\left(\sum_{(r,p)\in A}(-1)^p(k_r^{(p)}-k_{r-1}^{(p)})\right) \\
&= \sum_{t=1}^{\min\{Km,N\}}\frac{N!}{(N-t)!}\sum_{\mathcal{A}\in P(2Km,t)}\sum_{\substack{k_1^{(1)},\dots,k_m^{(1)}\in T\\ \vdots\\ k_1^{(2K)},\dots,k_m^{(2K)}\in T\\ k_{j-1}^{(p)}\neq k_j^{(p)}}}\prod_{A\in\mathcal{A}}\delta\left(\sum_{(r,p)\in A}(-1)^p(k_r^{(p)}-k_{r-1}^{(p)})\right).
\end{aligned}
$$

This proves the lemma. ■

In view of the previous lemma and for the sake of simple notation we denote

$$
B(\mathcal{A},T) := \sum_{\substack{k_1^{(1)},\dots,k_m^{(1)}\in T\\ \vdots\\ k_1^{(2K)},\dots,k_m^{(2K)}\in T\\ k_{j-1}^{(p)}\neq k_j^{(p)},j\in[m],p\in[2K]}}\prod_{A\in\mathcal{A}}\delta\left(\sum_{(r,p)\in A}(-1)^p(k_r^{(p)}-k_{r-1}^{(p)})\right). \tag{3.19}
$$

3.3 Proof of Theorem 2.1

Let us assemble all the pieces to complete the proof of Theorem 2.1. By Lemma 3.3 we need to investigate the quantity $C(\mathcal{A},T)$ defined in (3.15) for $\mathcal{A}\in P(2n,t)$. Here the indeces $(k_1,\dots,k_{2n})\in T^{2n}$ are subjected to the $|\mathcal{A}|=t$ linear constraints $\sum_{r\in A}(k_{r+1}-k_r)=0$ for all $A\in\mathcal{A}$. These constraints are independent except for $\sum_{r=1}^{2n}(k_{r+1}-k_r)=0$. Thus, we can estimate

$$
C(\mathcal{A},T) \le |T|^{2n-t+1} \le M^{2n-t+1}. \tag{3.20}
$$

By Lemma 3.3 we obtain (note that in the situation of Theorem 2.1 T is not random, so $\mathbb{E}=\mathbb{E}_X$)

$$
\mathbb{E}[\|H_0^n\|_F^2] \le \sum_{t=1}^{\min\{n,N\}}\frac{N!}{(N-t)!}\sum_{\mathcal{A}\in P(2n,t)}|T|^{2n+1-t} \le M^{2n+1}\sum_{t=1}^{n}(N/M)^t S_2(2n,t).
$$

where $S_2(n,t)=|P(2n,t)|$ are the associated Stirling numbers of the second kind. Set $\theta=N/M$. From the generating function (2.2) of the numbers $S_2(n,k)$ we know that

$$
\sum_{t=1}^{n} S_2(2n,t)\theta^t = F_{2n}(\theta)
$$

with F_{2n} defined by (2.9). Markov's inequality yields

$$\begin{aligned}\mathbb{P}(\|(N^{-1}H_0)^n\|_F \geq \kappa) &= \mathbb{P}(\|H_0^n\|_F^2 \geq N^{2n}\kappa^2) \\ &\leq N^{-2n}\kappa^{-2}\mathbb{E}[\|H_0^n\|_F^2] \leq \kappa^{-2}M\theta^{-2n}F_{2n}(\theta) = \kappa^{-2}MG_{2n}(\theta).\end{aligned}$$

We remark that by (3.7) we have $\kappa < 1$. In the event that $\|(N^{-1}H_0)^n\|_F \leq \kappa$ this implies that $(I_T - (N^{-1}H_0)^n)$ is invertible by the von Neumann series and by (3.3) also

$$\mathcal{F}_{TX}^*\mathcal{F}_{TX} = N(I_T - N^{-1}H_0)$$

is invertible. In particular, $\mathcal{F}_{TX}$ is injective. So this basic condition in Lemma 3.1 is satisfied automatically with a probability that can be derived from the estimation above, and we do not even need to invoke Lemma 3.2.

Let us now consider $\mathbb{P}(E_k)$. By Lemma 3.4 we need to bound $B(\mathcal{A}, T)$ defined in (3.19), i.e., the number of vectors $(k_j^{(p)}) \in T^{2Km}$ satisfying $\sum_{(r,p)\in A}(-1)^p(k_r^{(p)} - k_{r-1}^{(p)}) = 0$ for all $A \in \mathcal{A}$ with $\mathcal{A} \in P(2Km, t)$. These are t independent linear constraints. So the number of these indices is bounded from above by $|T|^{2Km-t} \leq M^{2Km-t}$. Thus, similarly as above we obtain

$$\mathbb{E}\left[|((HR_T)^m \operatorname{sgn}(c))_k|^{2K}\right] \leq \sum_{t=1}^{Km} N^t S_2(2Km, t) M^{2Km-t} = M^{2Km}F_{2Km}(\theta).$$

By (3.17) this yields

$$\mathbb{P}(E_k) \leq \beta^{-2n}\sum_{m=1}^{n}\theta^{-2mK_m}F_{2mK_m}(\theta) = \beta^{-2n}\sum_{m=1}^{n}G_{2mK_m}(\theta).$$

Let $\mathbb{P}(\text{failure})$ denote the probability that exact reconstruction of f by ℓ^1-minimization fails. By Lemma 3.1, (3.12) and by the union bound we finally obtain

$$\begin{aligned}\mathbb{P}(\text{failure}) &\leq \mathbb{P}\left(\{\mathcal{F}_{TX} \text{ is injective}\} \cup \{\sup_{k\in T^c}|P_k| \geq 1\}\right) \\ &\leq \sum_{k\in[-q,q]^d}\mathbb{P}(E_k) + \mathbb{P}(\|(N^{-1}N_0)^n\|_F \geq \kappa) \leq D\beta^{-2n}\sum_{m=1}^{n}G_{2mK_m}(\theta) + \kappa^{-2}MG_{2n}(\theta)\end{aligned}$$

under the conditions

$$a_1 = a = \sum_{m=1}^{n}\beta^{n/K_m} < 1, \qquad a_2 + a_1 = 1 \qquad \text{i.e.} \quad a_2 = 1 - a,$$

$$\frac{\kappa}{1-\kappa} \leq \frac{a_2}{1+a_1}M^{-3/2} = \frac{1-a}{1+a}M^{-3/2},$$

see (3.8). This proves Theorem 2.1.

3.4 Proof of Theorem 2.3

Recall that here T is a random set modelled by (2.12). The completion of the proof of Theorem 2.3 will be slightly more complicated as above because we still need to take the expectation with respect to the set T in Lemmas 3.3 and 3.4. Let us start with the expectation of $C(\mathcal{A}, T)$ defined in (3.15).

Lemma 3.5. *For $\mathcal{A} \in P(2n,t)$ it holds*

$$\mathbb{E}[C(\mathcal{A},T)] \leq \sum_{s=2}^{n} (\mathbb{E}|T|)^s \sum_{R=0}^{\min\{s,t\}-1} D^{-R} \#\{\mathcal{B} \in U(2n,s), \operatorname{rank} M(\mathcal{A},\mathcal{B}) = R\}.$$

Proof: Using linearity of expectation we obtain

$$\begin{aligned}
\mathbb{E}[C(\mathcal{A},T)] &= \mathbb{E}\left[\sum_{k_1,\ldots,k_{2n} \in [-q,q]^d, k_j \neq k_{j+1}} \prod_{j=1}^{2n} I_{\{k_j \in T\}} \prod_{A \in \mathcal{A}} \delta\left(\sum_{r \in A} (k_{r+1} - k_r)\right)\right] \\
&= \sum_{k_1,\ldots,k_{2n} \in [-q,q]^d, k_j \neq k_{j+1}} \mathbb{E}\left[\prod_{j=1}^{2n} I_{\{k_j \in T\}}\right] \prod_{A \in \mathcal{A}} \delta\left(\sum_{r \in A} (k_{r+1} - k_r)\right).
\end{aligned}$$

Hereby, $I_{\{k \in T\}}$ denotes an indicator variable which is 1 if and only if $k \in T$. The expression $\mathbb{E}\left[\prod_{j=1}^{2n} I_{\{k_j \in T\}}\right]$ depends on how many different k_j's there are. So once again partitions enter the game. If $(k_1,\ldots,k_{2n}) \in ([-q,q]^d)^{2n}$ is a vector satisfying $k_j \neq k_{j+1}$ then we associate a partition $\mathcal{B} = (B_1,\ldots,B_s)$ of $\{1,\ldots,2n\}$ such that j and j' are in the same set B_i if and only if $k_j = k_{j'}$. Obviously, j and $j+1$ must be contained in different blocks for all j due to the condition $k_j \neq k_{j+1}$ (once again we agree on the convention that $2n+1 \equiv 1$). In other words $\mathcal{B}$ has no adjacencies, i.e., $\mathcal{B} \in U(2n,s)$. Now if $\mathcal{B}$ has $|\mathcal{B}| = s$ blocks then by the probability model (2.12) for T and stochastic independence

$$\mathbb{E}\left[\prod_{i=1}^{2n} I_{\{k_i \in T\}}\right] = \mathbb{E}\left[\prod_{j=1}^{s} I_{\{k_{B_j} \in T\}}\right] = \prod_{j=1}^{s} \mathbb{E}[I_{\{k_{B_j} \in T\}}] = \tau^s, \tag{3.21}$$

where (unambiguously) $k_{B_j} = k_i$ if $i \in B_j$. We further introduce the notation $\sigma_{\mathcal{B}}(r) = j$ if and only if $r \in B_j \in \mathcal{B}$. This leads to

$$\mathbb{E}[C(\mathcal{A},T)] = \sum_{s=2}^{2n} \tau^s \sum_{\mathcal{B} \in U(2n,s)} \sum_{\substack{k_1,\ldots,k_s \in [-q,q]^d \\ k_i \text{p.w. different}}} \prod_{A \in \mathcal{A}} \delta\left(\sum_{r \in A} (k_{\sigma_{\mathcal{B}}(r+1)} - k_{\sigma_{\mathcal{B}}(r)})\right).$$

Clearly, the expression $\prod_{A \in \mathcal{A}} \delta\left(\sum_{r \in A} (k_{\sigma_{\mathcal{B}}(r+1)} - k_{\sigma_{\mathcal{B}}(r)}\right)$ is 1 if and only if

$$\sum_{r \in A} (k_{\sigma_{\mathcal{B}}(r)} - k_{\sigma_{\mathcal{B}}(r+1)}) = 0 \quad \text{for all } A \in \mathcal{A} \tag{3.22}$$

and 0 otherwise. For $j \in \{1,\ldots,s\}$ the term k_j appears $|A_i \cap B_j|$ times as $k_{\sigma_{\mathcal{B}}(r)}$ when r runs through $A_i \in \mathcal{A}$. Let $M = M(\mathcal{A},\mathcal{B})$ denote the $t \times s$ matrix whose entries are defined by (2.5). Then (3.22) is satisfied if and only if $(k_1,\ldots,k_s) \in ([-q,q]^d)^s$ is contained in the kernel of $M(\mathcal{A},\mathcal{B})$. Thus, if the rank of $M(\mathcal{A},\mathcal{B})$ equals R then the number of vectors $(k_1,\ldots,k_s) \in ([-q,q]^d)^s$ for which (3.22) is satisfied can be bounded by D^{s-R} where $D = (2q+1)^d$. (Here we even neglected the condition that the $k_1,\ldots,k_s$ should be pairwise different). So finally we obtain

$$
\begin{aligned}
\mathbb{E}[C(\mathcal{A},T)] \;&\le\; \sum_{s=2}^{n}\tau^s \sum_{R=0}^{\min\{s,t\}-1} D^{s-R}\#\{B\in U(n,s), \operatorname{rank} M(\mathcal{A},\mathcal{B})=R\}\\
&=\; \sum_{s=2}^{n}(\mathbb{E}|T|)^s \sum_{R=0}^{\min\{s,t\}-1} D^{-R}\#\{B\in U(n,s), \operatorname{rank} M(\mathcal{A},\mathcal{B})=R\},
\end{aligned}
$$

where we substituted $\mathbb{E}|T| = \tau D$. ■

Since $\mathbb{E}=\mathbb{E}_X\mathbb{E}_T$ by Fubini's theorem and stochastic independence of T and X the previous result yields together with Lemma 3.3

$$
\begin{aligned}
&\mathbb{E}[\|H_0^n\|_F^2]\\
&\le \sum_{t=1}^{\min\{n,N\}} \frac{N!}{(N-t)!} \sum_{\mathcal{A}\in P(2n,t)} \sum_{s=2}^{n}(\mathbb{E}|T|)^s \sum_{R=0}^{\min\{s,t\}-1} D^{-R}\#\{B\in U(n,s), \operatorname{rank} M(\mathcal{A},\mathcal{B})=R\}\\
&= \sum_{t=1}^{\min\{n,N\}} \frac{N!}{(N-t)!} \sum_{s=2}^{n}(\mathbb{E}|T|)^s \sum_{R=0}^{\min\{s,t\}-1} D^{-R} Q(2n,t,s,R) \;=\; N^{2n} W(n,N,\mathbb{E}|T|,D)
\end{aligned}
$$

by definition (2.6) of the numbers $Q(2n,t,s,R)$ and by definition (2.13) of the function W. Markov's inequality yields

$$
\mathbb{P}(\|(N^{-1}H_0)^n\|_F \ge \kappa) \;\le\; N^{-2n}\kappa^{-2}\mathbb{E}[\|H_0^n\|_F^2] \;\le\; \kappa^{-2} W(n,N,\mathbb{E}|T|,D).
$$

We remark that by the same argument as in the proof of Theorem 2.3 $\mathcal{F}_{TX}$ is injective in the event $\|(N^{-1}H_0)^n\|_F < 1$.

Let us turn now to the estimation of $\mathbb{P}(E_k)$. From Lemma 3.4 one realizes that we need to estimate the expected value of $B(\mathcal{A},T)$ defined in (3.19).

Lemma 3.6. *For $\mathcal{A}\in P(2Km,t)$ it holds*

$$
\mathbb{E}[B(\mathcal{A},T)] \;\le\; \sum_{s=1}^{2Km}(\mathbb{E}|T|)^s \sum_{R=0}^{\min\{s,t\}} D^{-R}\#\{\mathcal{B}\in U^*(2K,m,s), \operatorname{rank} L(\mathcal{A},\mathcal{B})=R\}.
$$

Proof: As in the proof of the previous lemma we may write

$$
\mathbb{E}[B(\mathcal{A},T)] \;=\; \sum_{\substack{(k_j^{(p)})\in([-q,q]^d)^{2Km}\\ k_{j-1}^{(p)}\neq k_j^{(p)}}} \mathbb{E}\left[\prod_{(p,j)\in[2K]\times[m]} I_{\{k_j^{(p)}\in T\}}\right] \prod_{A\in\mathcal{A}} \delta\left(\sum_{(r,p)\in A}(-1)^p(k_r^{(p)}-k_{r-1}^{(p)})\right).
$$

Once again $\mathbb{E}\left[\prod_{(p,j)\in[2K]\times[m]} I_{\{k_j^{(p)}\in T\}}\right]$ depends on how many different $k^{(p)}$'s there are. So if $(k_1^{(1)},\ldots,k_m^{(2K)})\in([-q,q]^d)^{(2Km)}$ is a vector satisfying

$$
k_j^{(p)} \;\neq\; k_{j-1}^{(p)} \qquad \text{for all } j\in[m],\; p\in[2K] \tag{3.23}
$$

then we associate a partition $\mathcal{B} = (B_1, \dots, B_s)$ of $[2K] \times [m]$ such that (p, j) and (p', j') are contained in the same block if and only if $k_j^{(p)} = k_{j'}^{(p')}$. Obviously, (p, j) and $(p, j-1)$ cannot be contained in the same block due to the condition (3.23). In other words, $\mathcal{B}$ belongs to $U^*(2K, m, s)$. Now, if $\mathcal{B}$ has s blocks, i.e., there are s different values of $k_j^{(p)}$, then

$$\mathbb{E}\left[\prod_{(p,j)\in[2K]\times[m]} I_{\{k_j^{(p)}\in T\}}\right] = \tau^s$$

as in (3.21). Once more, we use the notation $\sigma_{\mathcal{B}}(p, j) = i$ if $(p, j) \in B_i \in \mathcal{B}$ and $\sigma(p, 0) = 0$. (Recall that by definition $k_0^{(p)} = k_0 = k$.) Thus,

$$\mathbb{E}[B(\mathcal{A}, T)] = \sum_{s=1}^{2n} \tau^s \sum_{\mathcal{B}\in U^*(2K,m,s)} \sum_{\substack{k_1,\dots,k_s\in[-q,q]^d \\ k_i \text{ p.w. different}}} \prod_{A\in\mathcal{A}} \delta\left(\sum_{(p,j)\in A} (-1)^p (k_{\sigma_{\mathcal{B}}(p,j)} - k_{\sigma_{\mathcal{B}}(p,j-1)})\right).$$

The term $\prod_{A\in\mathcal{A}} \delta\left(\sum_{(p,j)\in A} (-1)^p (k_{\sigma_{\mathcal{B}}(p,j)} - k_{\sigma_{\mathcal{B}}(p,j-1)})\right)$ contributes to the sum if and only if

$$\sum_{(p,j)\in A} (-1)^p (k_{\sigma_{\mathcal{B}}(p,j)} - k_{\sigma_{\mathcal{B}}(p,j-1)}) = 0 \quad \text{for all } A \in \mathcal{A}.$$

By definition (2.7) of the matrix $L(\mathcal{A}, \mathcal{B})$ and since $k_0 = k$ this is equivalent to

$$L(\mathcal{A}, \mathcal{B})(k_1, \dots, k_s)^T = kv(\mathcal{A}, \mathcal{B}), \tag{3.24}$$

where $v = v(\mathcal{A}, \mathcal{B})$ is the t-dimensional vector with entries

$$v_i = \sum_{(p,1)\in A_i} (-1)^p, \quad i = 1, \dots, t.$$

(If $d > 1$ then (3.24) has to interpreted vector-valued, i.e., for each component of $k \in [-q, q]^d$ and of $k_1, \dots, k_s \in [-q, q]^d$ we have one equation with the same $L(\mathcal{A}, \mathcal{B})$ and the same $v(\mathcal{A}, \mathcal{B})$.) If the rank of $L(\mathcal{A}, \mathcal{B})$ equals R then we can bound the number of solutions to (3.24) by D^{s-R}. Hence, we obtain the bound

$$\mathbb{E}[B(\mathcal{A}, T)] \le \sum_{s=1}^{2Km} \tau^s \sum_{R=0}^{\min\{s,t\}} D^{s-R} \#\{\mathcal{B} \in U^*(2K, m, s), \operatorname{rank} L(\mathcal{A}, \mathcal{B}) = R\}.$$

Since $\mathbb{E}|T| = \tau D$ this proves the lemma. ■

Together with Lemma 3.4 the previous result yields

$$\begin{aligned}\mathbb{E}\left[|((HR_T)^m\sigma)_k|^{2K}\right] &\le \sum_{t=1}^{\min\{Km,N\}} \frac{N!}{(N-t)!} \sum_{s=1}^{2Km} (\mathbb{E}|T|)^s \sum_{R=0}^{\min\{s,t\}} Q^*(2K, m, t, s, R)\, D^{-R} \\ &= N^{2Km} Z(K, m, N, \mathbb{E}|T|, D)\end{aligned}$$

where $Q^*(2K, m, t, s, R)$ are the numbers defined in (2.8). By (3.17) we obtain

$$\mathbb{P}(E_k) \le \beta^{-2n} \sum_{m=1}^{n} Z(K_m, m, N, \mathbb{E}|T|, D)$$

Finally, let $\mathbb{P}(\text{failure})$ denote the probability that exact reconstruction of f fails. By Lemma 3.1, (3.10), (3.11) and using that $\{\mathcal{F}_{TX} \text{ is not injective}\} \subset \{\|(N^{-1}H_0)^n\|_F \ge \kappa\}$ we obtain

$$\begin{aligned}
\mathbb{P}(\text{failure}) &\le \mathbb{P}\left(\{\mathcal{F}_{TX} \text{ is injective}\} \cup \{\sup_{k\in T^c} |P_k| \ge 1\} \right) \\
&\le \sum_{k\in[-q,q]^d} \mathbb{P}(E_k) + \mathbb{P}(\|(N^{-1}N_0)^n\|_F \ge \kappa) + \mathbb{P}(|T| \ge (\alpha+1)\mathbb{E}|T|) \\
&\le D\beta^{-2n} \sum_{m=1}^{n} Z(K_m, m, N, \mathbb{E}|T|, D) + \kappa^{-2} W(n, N, \mathbb{E}|T|, D) + \exp(-\frac{3\alpha^2}{6+2\alpha}\mathbb{E}|T|)
\end{aligned}$$

under the conditions

$$\begin{aligned}
a_1 &= a = \sum_{m=1}^{n} \beta^{n/K_m} < 1, \qquad a_2 + a_1 = 1 \qquad \text{i.e.} \quad a_2 = 1 - a, \\
\frac{\kappa}{1-\kappa} &\le \frac{a_2}{1+a_1}((\alpha+1)\mathbb{E}|T|)^{-3/2} = \frac{1-a}{1+a}((\alpha+1)\mathbb{E}|T|)^{-3/2},
\end{aligned}$$

see (3.9). This proves Theorem 2.3.

3.5 Proof of Corollary 2.2

We have to show that a finer analysis of the probability bound (2.11) of Theorem 2.1 gives Corollary 2.2. We first claim that the associated Stirling numbers satisfy the estimate

$$S_2(n, k) \le (3n/2)^{n-k} \quad \text{for all } k = 1, \ldots, \lfloor n/2 \rfloor. \tag{3.25}$$

Indeed, the claim is true for $S_2(1, k) = 0$ and $S_2(2, 1) = 1$. Now suppose, the claim is true for all $S_2(m, k)$ with $m < n$. Then from the recursion formula (2.3) it follows

$$\begin{aligned}
S_2(n, k) &= kS_2(n-1, k) + (n-1)S_2(n-2, k-1) \\
&\le k(3(n-1)/2)^{n-k-1} + (n-1)(3n/2-3)^{n-k-1} \le (n-1+k)(3n/2)^{n-k-1} \\
&\le (3n/2)^{n-k}
\end{aligned}$$

since $n - 1 + k \le 3n/2$. This proves (3.25). Pluggin this into the definition of G_{2n} yields

$$\begin{aligned}
G_{2n}(\theta) &= \theta^{-2n} \sum_{k=1}^{n} S_2(2n, k)\theta^k \le \theta^{-2n} \sum_{k=1}^{n} (3n)^{2n-k}\theta^k = (3n/\theta)^{2n} \sum_{k=1}^{n} (\theta/3n)^k \\
&= (3n/\theta)^{2n} \frac{(\theta/3n)^{n+1} - (\theta/3n)}{(\theta/3n) - 1} = (3n/\theta)^{2n-1} \frac{(\theta/3n)^n - 1}{(\theta/3n) - 1}.
\end{aligned}$$

Now assume we have chosen n such that $n \le \theta/6$. Then we further obtain

$$G_{2n}(\theta) \le (3n/\theta)^{n-1}.$$

Now consider the term $D\beta^{-2n}\sum_{m=1}^{n} G_{2mK_m}(\theta)$ from the probability bound (2.11). We choose $K_m = r(n/m)$ where r denotes the function that rounds to the nearest integer. Then it is easy to see that

$$mK_m \in \{\lceil 2n/3 \rceil, \dots, \lfloor 4n/3 \rfloor\}, \quad m \in \{1, \dots, n\}.$$

Thus,

$$\sum_{m=1}^{n} G_2 m K_m(\theta) \le n \max_{k\in\{\lceil 2n/3\rceil,\dots,\lfloor 4n/3\rfloor\}} G_{2k}(\theta) \le n \left(\frac{4n}{\theta}\right)^{2n/3-1}$$

provided $k \le \theta/6$ for all $k \in \{\lceil 2n/3 \rceil, \dots, \lfloor 4n/3 \rfloor\}$, i.e., $6\lfloor 4n/3 \rfloor \le \theta$. This yields

$$D\beta^{-2n}\sum_{m=1}^{n} G_{2mK_m}(\theta) \le Dn\left(\frac{4n}{\theta}\right)^{-1}\left(\beta^{-3}\frac{4n}{\theta}\right)^{2n/3}.$$

In order to make this expression small it is certainly a good strategy to make the last term smaller than 1. Indeed, choose

$$n = n(\theta) := \left\lfloor \frac{\beta^3\theta}{8} \right\rfloor \tag{3.26}$$

implying $\beta^{-3}4n/\theta \le 1/2$. (This choice for n is certainly valid since $\beta < 1$ as it must satisfy condition (2.10).) We obtain

$$D\beta^{-2n}\sum_{m=1}^{n} G_{2mK_m}(\theta) \le \frac{1}{4}D\theta\, 2^{-2n(\theta)/3}.$$

A simple calculation yields that the latter term is less than $\epsilon/2$ if

$$\frac{2\ln(2)}{3}n(\theta) - \ln(\theta) \ge \ln(D) + \ln(\epsilon^{-1}) - \ln(2).$$

Furthermore, a simple numerical test shows that a valid choice for β is $\beta = 0.47$. The corresponding $a = \sum_{m=1}^{n}\beta^{n/K_m}$ is always less than 0.957 and $n(\theta) \approx \lfloor 0.013\,\theta \rfloor$. Recalling that $\theta = M/N$ it follows that there exists a constant C_1 such that $D\beta^{-2n}\sum_{m=1}^{n} G_{2mK_m}(\theta) \le \epsilon/2$ provided

$$N \ge C_1 M(\ln(D) + \ln(\epsilon^{-1})).$$

Now consider the other term $M\kappa^{-2}G_{2n}(\theta)$ in the probability bound (2.11). We choose κ such that there is equality in (2.10), i.e.,

$$\kappa = \frac{(1-a)/(1-a)M^{-3/2}}{1 + (1-a)/(1+a)M^{-3/2}} \ge \frac{1-a}{2(1+a)}M^{-3/2}.$$

Hence,

$$M\kappa^{-2}G_{2n}(\theta) \le \left(\frac{1-a}{2(1+a)}\right)^2 M^4 G_{2n}(\theta)$$

Now we do not have the freedom anymore to choose n. We have to make the same choice (3.26) as above. This yields

$$M\kappa^{-2}G_{2n(\theta)}(\theta) \le \left(\frac{2(1+a)}{(1-a)}\right)^2 M^4 \left(\frac{3\beta^3}{8}\right)^{n(\theta)-1}.$$

Requiring that the latter expression is less than $\epsilon/2$ is equivalent to

$$(n(\theta)-1)\ln\left(\frac{8}{3\beta^3}\right) \geq \ln\left(8\left(\frac{1+a}{1-a}\right)^2\right) + 4\ln(M) + \ln(\epsilon^{-1}).$$

As already remarked the choice $\beta = 0.47$ results in $a \leq 0.957$ and $n(\theta) \approx \lfloor 0.013\,\theta \rfloor$. Hence, $\ln(8/(3\beta^3)) \approx 3.2459$ and $\ln(8((1+a)/(1-a))^2) \approx 9.7153$. Since $M \leq D$ there exists a constant C_2 (whose precise value may be calculated from the numbers above) such that $M\kappa^{-2}G_{2n(\theta)}(\theta) \leq \epsilon/2$ provided

$$N \geq C_2 M(\ln(D) + \ln(\epsilon^{-1})).$$

Choosing $C := \max\{C_1, C_2\}$ completes the proof of Corollary 2.2.

We remark that analyzing numerical plots for $\beta^{-2n'(\theta)}\sum_{m=1}^{n'(\theta)} G_{2mK_m(\theta)}(\theta)$ and $G_{2n'(\theta)}(\theta)$ for $n'(\theta) = \lceil\theta/12\rceil$ indicates that one may choose the constant C much smaller as the ones resulting from the theoretical analysis above. It seems that $C \lessapprox 20$ is a valid choice.

3.6 Remarks

We conclude this section with some remarks.

(a) Let us give a more detailed reason why we believe that the probilistic model for the "sparsity set" T is likely to give better probability bounds for exact reconstruction than the deterministic approach holding for all T of a given size. Indeed the main difference in the two previous proofs lies in the estimation of $C(\mathcal{A}, T)$ and $B(\mathcal{A}, T)$ defined in (3.15) and (3.19). If $|\mathcal{A}| = t$ then for deterministic T we used the estimation (3.20), i.e., $C(\mathcal{A}, T) \leq |T|^{2n-t+1}$. Indeed, if T is an arithmetic progression then $C(\mathcal{A}, T)$ may come very close to this upper bound. However, for generic sets T the bound is quite pessimistic. In fact, in the probabilistic model the expected size of $C(\mathcal{A}, T)$ can be bounded by

$$\mathbb{E}[C(\mathcal{A}, T)] \leq \sum_{s=2}^{n}(\mathbb{E}|T|)^s \sum_{R=0}^{\min\{s,t\}-1} D^{-R}\#\{\mathcal{B} \in U(2n, s), \operatorname{rank} M(\mathcal{A}, \mathcal{B}) = R\},$$

see Lemma 3.5. In particular, if D is large (and $\mathbb{E}|T|$ not too small) then the latter estimate should be much better. Let us illustrate this with two examples.

1. Let $\mathcal{A} = \{\{1, 2, 3, 5\}, \{4, 6\}\}$, i.e., $2n = 6$ and $t = 2$. Then (3.20) yields $C(\mathcal{A}, T) \leq |T|^5$ while computing 3.27) explicitly gives

   $$\mathbb{E}[C(\mathcal{A}, T)] = D^{-1}\left[(\mathbb{E}|T|)^2 + 10(\mathbb{E}|T|)^3 + 20(\mathbb{E}|T|)^4 + 9(\mathbb{E}|T|)^5 + (\mathbb{E}|T|)^6\right].$$

   Clearly, if D is sufficiently large then the probabilistic estimate is much better than the deterministic one.

2. Let $\mathcal{A} = \{\{1, 2, 3\}, \{4, 5, 6\}\}$, so again $2n = 6$ and $t = 2$. Then the deterministic estimate gives again $C(\mathcal{A}, T) \leq |T|^5$ while (3.27) results in

   $$\mathbb{E}[C(\mathcal{A}, T)] \leq (\mathbb{E}|T|)^2 + 3(\mathbb{E}|T|)^3 + (\mathbb{E}|T|)^4 + D^{-1}\left[7(\mathbb{E}|T|)^3 + 19(\mathbb{E}|T|)^4 + 9(\mathbb{E}|T|)^5 + (\mathbb{E}|T|)^6\right].$$

   So here one has to choose both $\mathbb{E}|T|$ and $D >> \mathbb{E}|T|$ large to see that potentially the probabilistic estimate is much better.

(b) **Discrete Fourier transforms:** The whole proofs work without essential change if one replaces our setting by the following one similar to the situation investigated by Candes, Romberg and Tao in [6]. Consider functions on the cyclic group $\mathbb{Z}_p^d = \{0, \ldots, p-1\}^d$, $p \in \mathbb{N}$, rather than on $[0, 2\pi]^d$. The discrete Fourier transform is defined by

$$\hat{f}(\omega) := \sum_{x \in \mathbb{Z}_p^d} f(x) e^{-2\pi i x \cdot \omega / p}, \quad \omega \in \mathbb{Z}_p^d.$$

We draw $x_1, \ldots, x_N$ from the uniform distribution on $\mathbb{Z}_p^d$. Note that in contrast to sampling from $[0, 2\pi]^d$ it may occur with non-zero probability that some elements of $\mathbb{Z}_p^d$ are drawn more than once. But this will not do much harm.

Let f be such that $\hat{f}$ is a sparse vector on $\mathbb{Z}_p^d$. Once again we try to reconstruct f from its sample values $f(x_j)$ by minimizing the ℓ^1-norm of $\hat{f}$ under the constraint that the observed values $f(x_j)$ are matched.

Theorems 2.1 and 2.3 will also apply to this situation. Indeed, the only thing that differs in the proofs is that we have to calculate modulo p in the definition of $C(\mathcal{A}, T)$ and $B(\mathcal{A}, T)$, see (3.15) and (3.19). This is apparent from (3.14) where the integral is replaced by a sum of exponentials. Nevertheless, the deterministic and probabilistic estimates for the quantities $C(\mathcal{A}, T)$ and $B(\mathcal{A}, T)$ still hold and so everything goes through in completely the same manner.

Of course, one can also exchange the role of f and $\hat{f}$, aiming at reconstructing a sparse signal on $\mathbb{Z}_p^d$ from random samples of its Fourier transform. Indeed, this situation is investigated in [6] with a different probability model for the sampling points. In other words, we presented a slightly different approach for the main result in [6].

4 Some more on set partitions

From Theorem 2.3 we realize that we have to investigate the functions $F_n(\theta)$ connected to set partitions in $P(n,t)$ and also the numbers $Q(n,t,s,R)$ and $Q^*(K,m,t,s,R)$, respectively. We already gave some information on the number $S_2(n,t)$ of partitions in $P(n,t)$ earlier. Let us be a bit more detailed here. Clearly, by definition (2.9) of F_n and the generating function (2.2) we see that

$$F_n(\theta) = \sum_{k=1}^{\lfloor n/2 \rfloor} S_2(n,k)\theta^k.$$

(This follows also directly from the proof of Theorem 2.1.) In particular, F_{2n} is a polynomial of degree n. There are different ways of computing F_n explicitly. One possibility is to use the generating function (2.2) leading to

$$F_n(\theta) = \frac{\partial^n}{\partial x^n} \exp(\theta(e^x - x - 1))_{|x=0}.$$

One may also compute the numbers $S_2(n,k)$ explicitly. Indeed, differentiating (2.2) k times with respect to y and setting $y = 0$ yields

$$\sum_{n=1}^{\infty} S_2(n,k) \frac{x^n}{n!} = \frac{1}{k!} \left(e^x - x - 1\right)^k.$$

Expanding the right hand side into a power series and comparing coefficients yields (after some computations)

$$S_2(n,k) \;=\; \frac{1}{k!}\left(k^n + \sum_{j=1}^{k-1}(-1)^j\binom{k}{j}\sum_{\ell=0}^{j}\binom{j}{\ell}\frac{n!}{(n-\ell)!}(k-j)^{n-\ell}\right) \tag{4.1}$$

valid for $n \geq 2k$ (otherwise $S_2(n,k) = 0$). In the special case $k = 2$ we obtain $S_2(n,2) = 2^{n-1} - n - 1$. Further, a combinatorial argument shows that $S_2(2n,n) = \frac{2n!}{2^n n!}$. (One uses that $P(2n,n)$ consists only of partitions where each block has precisely 2 elements.)

Let us give the first of the functions F_{2n} explicitly in the following list,

$$F_2(y) \;=\; y, \qquad F_4(y) \;=\; y + 3y^2, \qquad F_6(y) \;=\; y + 25y^2 + 15y^3,$$
$$F_8(y) \;=\; y + 119y^2 + 490y^3 + 105y^4, \qquad F_{10}(y) \;=\; y + 501y^2 + 6825y^3 + 9450y^4 + 945y^5,$$
$$F_{12}(y) \;=\; y + 2035y^2 + 74316y^3 + 302995y^4 + 190575y^5 + 10395y^6.$$

Of course, explicit values of $S_2(n,k)$ can be read off this list.

Now consider the number $p_n = \sum_k S_2(n,k)$ of all partitions of $[n]$ into subsets having at least two elements. Setting $y = 1$ in the exponential generating function (2.2) yields

$$\sum_{n=1}^{\infty} p_n \frac{x^n}{n!} \;=\; \exp(e^x - x - 1). \tag{4.2}$$

Unfortunately, much less is known about the number of partitions in $U(n,s)$. As already mentioned, it was only very recently that D. Knuth [15] posed the problem of determining $|U(n,s)|$. Let us denote by $u_n = \sum_{k=2}^{n}|U(n,k)|$ the number of all partitions of $\{1,\dots,n\}$ having no adjacencies (recall that $U(n,1) = \emptyset$). Recently, it was proved in [4] that $u_n = p_n$. So (4.2) is also the exponential generating function of the numbers u_n. Concerning the size of $U^*(K,m,s)$, up to now, we cannot say more than that it is bounded by the number of all partitions into s blocks of a set with Km elements, i.e., by the (ordinary) Stirling number of the second kind $S(Km,s)$. If $m = 1$ then $|U^*(K,1,s)| = S(K,s)$ as already remarked. The Stirling numbers $S(n,k)$ have the generating function [17, 19]

$$\sum_{n=1}^{\infty}\sum_{k=1}^{n} S(n,k)y^k\frac{x^n}{n!} \;=\; \exp(y(e^x - 1)). \tag{4.3}$$

Let us denote $u^*_{K,m} = \sum_{k=1}^{Km}|U^*(K,n,k)|$. Then clearly $u^*_{K,m} \leq b_{Km} = \sum_{k=1}^{n} S(Km,k)$ with equality if $m = 1$. A lower bound for $u^*_{K,m}$ is given by the numbers p_{Km}.

Now some elementary observations concerning the numbers $Q(n,t,s,R)$ and $Q^*(K,n,t,s,R)$ can be made. Disregarding the rank of $M(\mathcal{A},\mathcal{B})$, the number of all pairs $(\mathcal{A},\mathcal{B})$ with $\mathcal{A} \in P(n,t)$ and $\mathcal{B} \in U(n,s)$ is $|P(n,t)| \times |U(n,s)|$, hence, $\sum_{R=0}^{\min\{s,t\}} Q(n,t,s,R) = |P(n,t)| \times |U(n,s)|$ and similarly for $Q^*(K,m,t,s,R)$. Summing also over t and s gives

$$\sum_t\sum_s\sum_R Q(n,t,s,R) \;=\; u_n p_n \;=\; p_n^2$$

and $\sum_{t,s,R} Q^*(K,m,t,s,R) = p_{Km}u^*_{K,m}$. In the following table we give some values of p_n, and b_n for even $n = 2,4,6,\dots$ (we omit the odd numbers since we do not need them for Theorem 2.3).

n	2	4	6	8	10	12	14	16
$p_n = u_n$	1	4	41	715	17 722	580 317	24 011 157	1 216 070 380
$b_n = u^*_{n,1}$	2	15	203	4140	115 975	4 213 597	190 899 322	10 480 142 147

We determined $Q(n, t, s, R)$ and $Q^*(K, m, t, s, R)$ for certain small n, K, m on a computer in the following way. First all partitions in $P(n, t)$ and $U(n, s)$ (resp. $U^*(K, m, s)$) are computed recursively. For $P(n, t)$ we have the following procedure:

1. if $n < 2$ or $t > n/2$ then RETURN $P(n, t) = \emptyset$.
2. if $t = 1$ then RETURN $P(n, 1) = \{\{1, \ldots, n\}\}$.
3. $P(n, t) = \emptyset$
4. compute (recursively) $P(n - 1, t)$ and $P(n - 2, t - 1)$.
5. for each $\mathcal{A} \in P(n - 1, t)$:
   for j from 1 to t:
   create new partition $\mathcal{A}'$ by adding the element n to the j-th subset of $\mathcal{A}$
   add $\mathcal{A}'$ to $P(n, t)$
6. for each $\mathcal{A} \in P(n - 2, t - 1)$:
   for ℓ from 1 to $n - 1$:
   create new partition $\mathcal{A}'$ from $\mathcal{A}$ by incrementing each element $i \in [n - 2]$ by 1 if $i \geq \ell$
   and adding the subset $\{\ell, n\}$
   add $\mathcal{A}'$ to $P(n, t)$
7. RETURN $P(n, t)$

We remark that from this procedure also the recursion formula (2.3) follows.

The partitions in $U(n, s)$ are determined by first computing the set $V(n, s)$ of *all* partitions of $[n]$ into s blocks and then omitting those that have adjacencies. Similarly $U^*(K, n, s)$ is computed. Hereby, we have the following recursive procedure to compute $V(n, s)$:

1. if $s = 1$ RETURN $V(n, s) = \{\{1, \ldots, n\}\}$
2. if $s = n$ RETURN $V(n, s) = \{\{1\}, \ldots, \{n\}\}$
3. $V(n, s) = \emptyset$
4. compute (recursively) $V(n - 1, s)$ and $V(n - 1, s - 1)$
5. for each $\mathcal{A} \in V(n - 1, s)$:
   for j from 1 to s:
   create new partition $\mathcal{A}'$ by adding the element n to the j-th subset of $\mathcal{A}$
   add $\mathcal{A}'$ to $V(n, s)$
6. for each $\mathcal{A} \in V(n - 1, s - 1)$:
   create new partition $\mathcal{A}' = \mathcal{A} \cup \{n\}$
   add $\mathcal{A}'$ to $V(n, s)$
7. RETURN $V(n, s)$

One may easily deduce the recursion formula $S(n, k) = kS(n - 1, k) + S(n - 1, k - 1)$ for the Stirling numbers of the second kind $S(n, k) = |V(n, k)|$ from this procedure.

After determining $P(n,t)$ and $U(n,s)$ for each pair $(\mathcal{A},\mathcal{B})$ with $\mathcal{A}\in P(n,t)$ and $\mathcal{B}\in U(n,s)$ (or $\mathcal{B}\in U^*(K,n,s)$ resp.) we set up the matrix $M(\mathcal{A},\mathcal{B})$ (or $L(\mathcal{A},\mathcal{B})$), see (2.5) and (2.7), and compute its rank. By counting the number of matrices $M(\mathcal{A},\mathcal{B})$ that have rank R we determine $Q(n,t,s,R)$ or $Q^*(K,m,t,s,R)$, respectively. The results of these computations for certain n,K,m are given in the appendix. Considering the table of the numbers p_n (recall that p_n^2 equals the overall number of matrices whose rank has to be determined) we realize that this procedure is practicable only for small values of n. Even for $n=10$ the computing time reaches several days and for $n=14$ it seems impossible to do the task in a reasonable time as $p_{14}^2 = 576535660478649$.

The following lemma is concerned with $Q(n,t,s,0)$ for some special cases.

Lemma 4.1. *(a)* $Q(n,1,s,0) = |U(n,s)|$.

(b) It holds $Q(2n,2,2,0) = \frac{(2n)!}{2(n!)^2} - 1$ *and* $Q(2n,2,2,1) = 2^{2n-1} - 2n - \frac{(2n)!}{2(n!)^2}$.

(c) If $n>2$ *and* $2s \geq 3n$ *then* $Q(2n,n,s,0)=0$.

(d) If $t\neq 1$ *then* $Q(2n,t,2n,0)=0$.

(e) If $n>3$ *and* $3t\geq 2n$ *then* $Q(2n,t,2n-1,0)=0$.

Proof: (a) There is only one partition in $P(n,1)$ and the maximal rank of $M(\mathcal{A},\mathcal{B})$ is $\min\{t,s\}-1=0$. Thus, $Q(n,1,s,0)=|U(n,s)|$.

(b) Clearly, $U(2n,2)$ consists of only 1 partition $\mathcal{B}=(B_1,B_2)$, i.e.,

$$B_1 = \{1,3,5,\ldots,2n-1\}, \qquad B_2 = \{2,4,6,\ldots,2n\}.$$

The associated matrix $M=M(\mathcal{A},\mathcal{B})$, $\mathcal{A}=\{A_1,A_2\}\in P(2n,2)$ has entries

$$M_{i,j} = |A_i\cap B_j| - |(A_i+1)\cap B_j| = |A_i\cap B_j| - |A_i\cap(B_j-1)| = |A_i\cap B_j| - |A_i\cap B_{3-j}|$$

since $B_1-1=B_2$ and $B_2-1=B_1$. Thus, $M(\mathcal{A},\mathcal{B})$ has rank 0, i.e., $M(\mathcal{A},\mathcal{B})=0$ if and only if

$$|A_1\cap B_1| = |A_1\cap B_2| \quad\text{and}\quad |A_2\cap B_1| = |A_2\cap B_2|. \tag{4.4}$$

So A_1 and A_2 must have the same number of elements from B_1 and from B_2. So we can construct all possible partitions $\mathcal{A}$ satisfying (4.4) in the following way. Choose $m\in\{1,\ldots,n-1\}$ and then form A_1 by taking m elements from B_1 and m elements from B_2. The set A_2 is formed of all the remaining elements. Then (4.4) is clearly satisfied. We can do this in $\binom{n}{m}^2$ different ways. However, if we run with m through $\{1,\ldots,n-1\}$ every possible partition appears once as $\{A_1,A_2\}$ and once as $\{A_2,A_1\}$, so that altogether we have the formula

$$Q(2n,2,2,0) = 2^{-1}\sum_{m=1}^{n-1}\binom{n}{m}^2 = \frac{(2n)!}{2(n!)^2} - 1.$$

The second equality follows from the fact that $\sum_{m=0}^{n}\binom{n}{m}^2 = \binom{2n}{n}$, see e.g. [20]. Now the second assertion follows easily since

$$Q(2n,2,2,1) = |P(2n,2)| - Q(2n,2,2,0) = 2^{2n-1} - 1 - 2n - Q(2n,2,2,0).$$

(c)-(e) For all the remaining cases we have to prove that for all relevant partitions $\mathcal{A} \in P(2n,t), \mathcal{B} \in U(2n,s)$ we never have $M(\mathcal{A},\mathcal{B}) = 0$ (the zero-matrix). Observe that $M(\mathcal{A},\mathcal{B}) = 0$ means that

$$|A \cap B| \; = \; |(A+1) \cap B| \quad \text{for all } A \in \mathcal{A}, B \in \mathcal{B} \tag{4.5}$$

(where $A+1$ is computed modulo n as usual). So for all three cases we assume that $\mathcal{A} \in P(2n,t)$ and $\mathcal{B} \in U(2n,s)$ are given (with t,s satisfying the respective conditions) and show that the condition (4.5) leads to a contradiction.

(c) Clearly, a partition $\mathcal{A}$ in $P(2n,n)$ has only subsets consisting of precisely 2 elements. The condition $2s \geq 3n$ implies that a partition $\mathcal{B} \in U(2n,s)$ has at least n singletons (i.e. subsets consisting of only one element). Indeed, if there would be less than n singletons than the overall number of elements would be larger than $n-1+2(s-(n-1))$ (i.e. $n-1$ sets with 1 element and $(s-(n-1))$ sets with at least 2 elements). Since $n-1+2(s-(n-1)) = 2s-n+1 \geq 3n-n+1 = 2n+1$ this produces a contradiction as there are only $2n$ elements.

Now, if $\{k\}$ is a singleton of $\mathcal{B}$ and $k \in A$ for $A \in \mathcal{A}$ then condition (4.5) implies that also $(k-1) \in A$. As all subsets A in $\mathcal{A}$ have precisely two elements this means that $A = \{k-1,k\}$. Using once more (4.5) we further see that this implies that neither $\{k-1\}$ nor $\{k+1\}$ can be singletons in $\mathcal{B}$. So $\mathcal{A}$ has the form

$$\{\{1,2\},\{3,4\},\ldots,\{2n-1,2n\}\}$$

and the singletons of $\mathcal{B}$ are $\{2\},\{4\},\ldots,\{2n\}$ up to shifting all elements by 1 (modulo n). We still have to distribute the remaining numbers $1,3,5,\ldots,2n-1$ onto subsets in $\mathcal{B}$. If $1 \in B \in \mathcal{B}$ then condition (4.5) with $A = \{1,2\}$ tells us that also $3 \in B$. The same argument for 3 and $A = \{3,4\}$ implies that also $5 \in B$ and so on. So $B = \{1,3,5,\ldots,2n-1\}$ and thus, $s = n+1$. Since $n > 2$ this is a contradiction to $s \geq 3n/2$. Thus, there is no pair of partitions $\mathcal{A} \in P(2n,n), \mathcal{B} \in U(2n,s)$ with $M(\mathcal{A},\mathcal{B}) = 0$.

(d) The only partition in $U(2n,2n)$ is $\mathcal{B} = \{\{1\},\{2\},\ldots,\{2n\}\}$. Thus the condition (4.5) implies that whenever $j \in A \in \mathcal{A}$ then also $j-1 \in A$. As j is arbitrary this means that the only possibility for A is $\{1,2,\ldots,2n\}$, i.e., $t=1$.

(e) The condition on t implies that there is at least one subset $A_1 \in \mathcal{A} \in P(2n,t)$ that has precisely 2 elements. Moreover, any partition in $U(2n,2n-1)$ has precisely $2n-2$ singletons and one subset B_1 consisting of precisely 2 elements. We write $A_1 = \{j_1,j_2\}$, $j_2 > j_1$ and $B_1 = \{k_1,k_2\}$.

We distinguish two cases. Let us first assume $j_2 \neq j_1+1$ and $\{j_1,j_2\} \neq \{1,2n\}$. All singletons in $\mathcal{B}$ are given by $\{k\}$ with $k \neq k_1,k_2$. Checking the condition (4.5) with A_1 and $\{k\}$ shows that necessarily $j_2 \neq k$ for all $k \neq k_1,k_2$. Without loss of generality this means $j_2 = k_2$. Condition (4.5) with A_1 and B_1 thus yields

$$|\{j_1,j_2\} \cap \{k_1,j_2\}| = |\{j_1+1,j_2+1\} \cap \{k_1,j_2\}|$$

It is not possible that the sets on both sides have both cardinality 2. Thus, the relation implies $j_1 \neq k_1$. Moreover, since by assumption $j_1+1 \neq j_2$ either $k_1 = j_1+1$ or $k_1 = j_2+1$. In both cases the singleton $\{j_1\}$ belongs to $\mathcal{B}$. Condition (4.5) yields $|\{j_1,j_2\} \cap \{j_1\}| = |\{j_1+1,j_2+1\} \cap \{j_1\}|$ which is not possible since $j_1 \neq j_2+1$ by the assumptions $j_2 \geq j_1$ and $\{j_1,j_2\} \neq \{1,2n\}$.

Next we treat the case $A_1 = \{j_1,j_2\} = \{j,j+1\}$. Without loss of generality we may assume $j=1$, so $A_1 = \{1,2\}$. Checking condition (4.5) with A_1 and $\{k\}$, $k \neq 1,2$ shows that $k \neq 1,3$. Thus $B_1 = \{1,3\}$ and the singletons of $\mathcal{B}$ are the sets $\{2\},\{4\},\{5\},\{6\},\ldots,\{2n\}$.

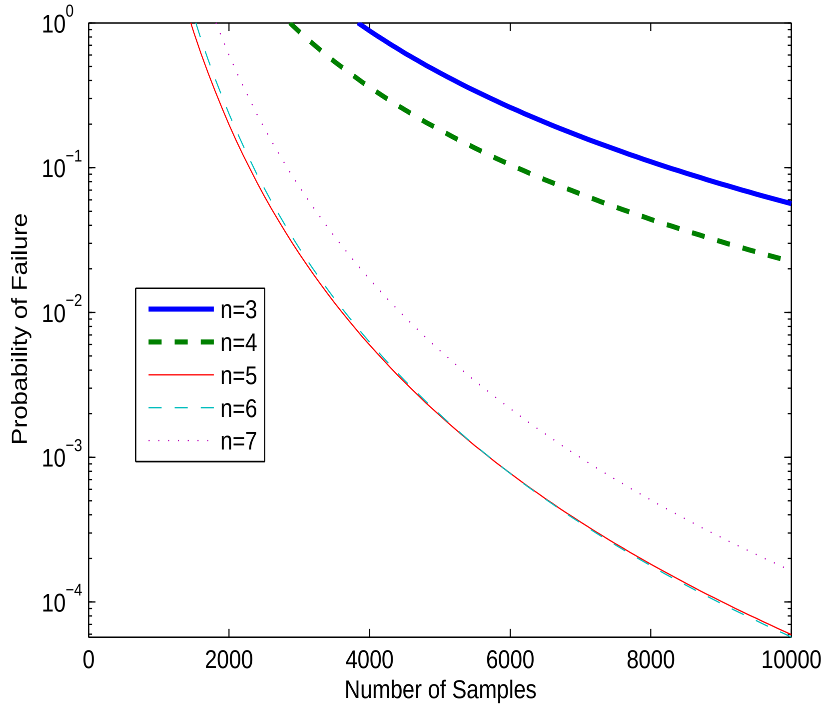


Figure 2: Bounds for the probability of failure of exact reconstruction due to Theorem 2.1 with $M = 10,\ D = 10000$.

Then condition (4.5) with A_1 and B_1 is satisfied. Now, let A be the subset of $\mathcal{A}$ containing the element 3 and write $A = \{3\} \cup A'$. Then condition (4.5) with $B = \{4\}$ reads $|(A' \cup \{3\}) \cap \{4\}| = |((A'+1) \cup \{4\}) \cap \{4\}| = 1$. Thus, A' and hence also A must contain the element 4. We may continue in this way to show that $A = \{3, 4, 5, \dots, 2n\}$. In particular, $t = 2$. Since $n > 3$ this is a contradiction to $3t > 2n$. ■

One may compare the assertions of this lemma with the tables in the appendix. For $Q^*(K, m, t, s, 0)$ certainly a similar analysis can be done but we have not further pursued this issue here.

5 Bounds for the probability of exact reconstruction

In this section we illustrate the bounds in Theorems 2.1 and 2.3 for the probability of exact reconstruction by drawing some plots. Hereby, we always plotted the bound of the probability of *failure* of exact reconstruction, i.e., 1 minus the expressions in (2.11) and (2.15).

In figure 5 we have chosen $M = 10$, $D = 10000$ and $n = 3, 4, 5, 6, 7$ to show a logarithmic plot of the probability bound (2.11) of Theorem 2.1 versus the number of samples. The parameter β was chosen always near to $1/2$ and then κ was determined such that there is equality in (2.10). One can see clearly, that here $n = 5$ or $n = 6$ is the optimal choice depending on the precise value of the number of samples N. Unfortunately, it seems that

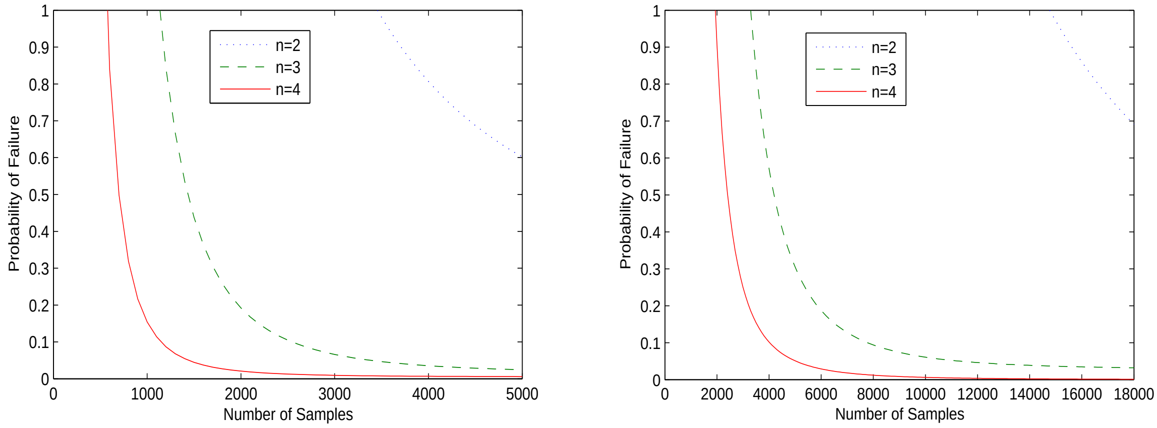


Figure 3: Probability of failure of exact reconstruction for $\mathbb{E}|T| = 4,\ D = 5000$ (left) and $\mathbb{E}|T| = 8, D = 20000$ (right) due to Theorem 2.3

these bounds are quite pessimistic when compared to the numerical experiments (see next section). In the given example one needs at least about $N = 2000$ samples (corresponding to a "non-linear oversampling factor" of 200) in order that the bound becomes non-trivial.

Based on the computation of the explicit values of the numbers Q and Q^* we can also illustrate the probability bound (2.15) in Theorem 2.3. Unfortunately, we may only take $n \leq 4$ since for higher values of n the corresponding numbers Q and Q^* are not at our disposal. Figure (5) shows a plot of the bound (2.15). We have chosen $n = 2, 3, 4$ and $(\mathbb{E}|T|, D) = (4, 5000)$, $(8, 20000)$ and varied the number N of sampling points. For $n = 2$ we have chosen $K_1 = 2, K_2 = 1$, for $n = 3$: $K_1 = 3, K_2 = 2, K_3 = 1$ and for $n = 4$ we took $K_1 = 4, K_2 = 2, K_3 = 1, K_4 = 1$ as suggested in Remark 2.4(a). It turned out that good choices for β are around $1/2$ and for $\kappa \approx 10^{-3}$ (with slight variations for the different choices of the other parameters). The remaining parameter α was chosen such that there is equality in (2.14).

Looking at the plot one realizes clearly that the bound becomes better for larger n. However, as above the bounds are still quite pessimistic. Nevertheless, as already remarked one expects them to be at least better than the ones of Theorem 2.1. Figure 5 supports this intuition. Indeed, we plotted the different bounds for $M = \mathbb{E}|T| = 4$, $D = 5000$, $n = 4$ and $K_1 = 4, K_2 = 2, K_3 = 1$ and $K_4 = 1$. Apparently the curve for the bound of Theorem 2.3 is far below the one of Theorem 2.1. Unfortunately, we cannot yet use the full strength of Theorem 2.3 as we are still lacking an efficient way to actually compute the bound explicitly for higher values of n. Actually up to now Theorem 2.1 still gives the better bound in most situations because we are able to evaluate (2.11) for arbitrary n.

Let us finally discuss possible reasons why the theoretical bounds are quite pessimistic. Both theorems give bounds for the probability that exact reconstruction holds for *all* choices of the coefficients f on T, while the numerical experiments in the next section choose also the coefficients on T at random. (Of course, it is impossible to check all possible coefficients by some algorithm.) Intuitively, it is very plausible that in such an experiment the probability of failure of exact reconstruction is much lower than for the situation in our main Theorem 2.3. We remark that it seems to be an interesting project to investigate theoretically also the

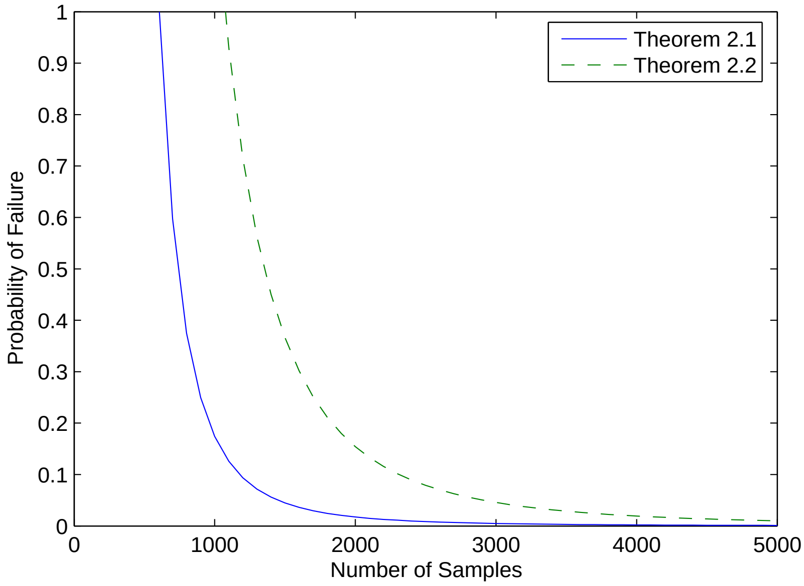


Figure 4: Comparison of the bounds of Theorem 2.1 and Theorem 2.3 for $M = \mathbb{E}|T| = 4$, $D = 5000$ and $n = 4$.

case that the coefficients of f on T are chosen at random, see also Section 5 in [7]. We plan to pursue this issue in a follow-up paper.

Of course, the theoretical bounds may also be pessimistic compared to reality since some of the estimates in the proof are perhaps not sharp. However, it seems to be hard to improve on the method of our proof.

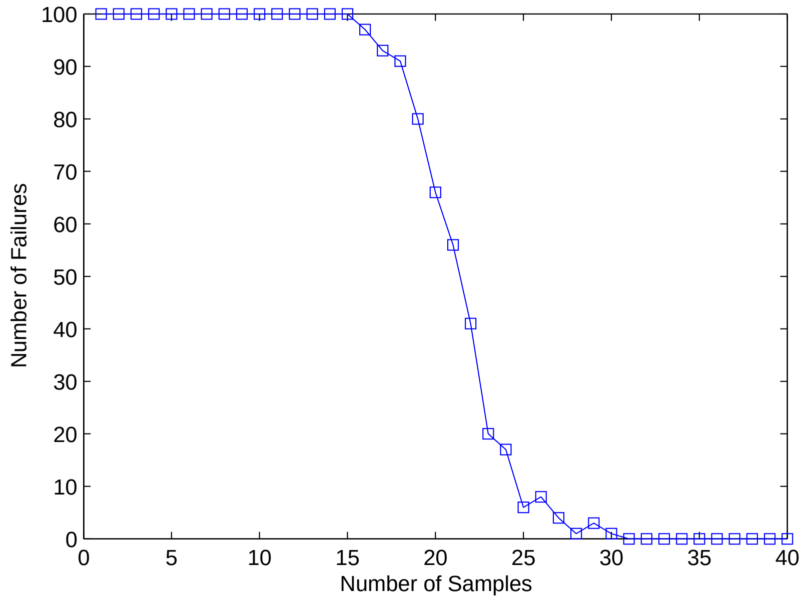
6 Numerical experiments

Let us describe some numerical tests of the proposed sampling resp. reconstruction method. In order to use convex optimization techniques we reformulate the optimization problem 2.1 as the following equivalent problem,

$$\min \sum_k u_k \quad \text{subject to } \sqrt{(c_k^{(1)})^2 + (c_k^{(2)})^2} \leq u_k, \qquad (6.1)$$
$$\sum_k (c_k^{(1)} + ic_k^{(2)}) e^{ik\cdot x_j} = f(x_j)$$

with u_k and $c_k^{(1)}$ and $c_k^{(2)}$, $k \in [-q,q]^d$, as real optimization variables. The solution to the original problem 2.1 is then given as $c_k = c_k^{(1)} + ic_k^{(2)}$.

A problem of the above type (6.1) is known as second order cone program [3]. Efficient algorithms to solve such problems exist. We have used the toolbox MOSEK (in connection with MATLAB), which provides an interior point solver for cone problems. We remark that if the coefficients c_k are real-valued then the minimization problem (2.1) can be recast as a linear program.



Numerical results: number of failures out of 100 trials for $M = |T| = 8$ and $q = 40$ versus number of samples

Our numerical experiment has the following form. We first choose the sparsity M, the maximal degree q (we only tested for $d = 1$) and the number of samples N. Then the following steps are done:

1. Choose a random subset $T \subset [-q, q]$ of size M from the uniform distribution. (Generate a random permutation of $[-q, q]$ and take the first M elements.)
2. Randomly generate the coefficients c_k for $k \in T$ by choosing their real part and imaginary part from a standard normal distribution.
3. Randomly select $x_1, \ldots, x_N$ independently from the uniform distribution on $[0, 2\pi]$.
4. Generate $f(x_j) = \sum_{k \in T} c_k e^{ikx_j}$, $j = 1, \ldots, N$.
5. Solve the minimization problem (6.1).
6. Compare the result to the original vector of coefficients.

For figure 6 we have chosen $q = 40$, i.e., $D = (2q + 1) = 81$ and $M = |T| = 8$. Then for each N between 1 and 40 we ran the above procedure 100 times and counted how often exact reconstruction failed. The result is illustrated in the plot. As one can see for N larger than 30 (corresponding to a non-linear oversampling factor of about 4) our reconstruction method always succeeded in giving back the original function exactly!

Comparing these results with the bounds of Theorem 2.3 as illustrated in the previous section one realizes that in practice the method works even much better than we are able to predict theoretically. So this method seems to have quite a lot of potential for practical applications of signal reconstruction.

A Appendix

A.1 Tables for $Q(n,t,s,R)$ and $Q^*(K,m,t,s,R)$

In the following tables we list some values for the numbers $Q(n,t,s,R)$ and $Q^*(K,m,t,s,R)$ that were computed by the procedures described in Section 4. For the probability estimation (2.15) the numbers $Q^*(2K_m,m,t,s,R)$, $m=1,\dots,n$ are needed and we have chosen $n=4$ and $K_1=4, K_2=2, K_3=1, K_4=1$ (since for higher numbers of n and K_m computing times are absurdly long). So the numbers $Q^*(K,m,t,s,R)$ have to be computed for $(K,m) = (8,1),(4,2),(2,3),(2,4)$.

Note that $Q(n,t,s,R)=0$ if $R \geq \min\{t,s\}$ or $s=1$ or $t>n/2$ and $Q^*(n,t,s,R)=0$ if $R>\min\{t,s\}$, which is the reason why we do not reproduce these cases. Also recall that $U(n,1)=\emptyset$ and, unless $m=1$, $U(K,m,1)=\emptyset$, hence, $Q(n,t,1,R)=0$ and $Q^*(K,m,t,1,R) = 0$.

$Q(4,1,s,R)$	$R=0$
$s=2$	1
$s=3$	2
$s=4$	1

$Q(4,2,s,R)$	$R=0$	$R=1$
$s=2$	2	1
$s=3$	2	4
$s=4$	0	3

$Q(6,1,s,R)$	$R=0$
$s=2$	1
$s=3$	10
$s=4$	20
$s=5$	9
$s=6$	1

$Q(6,2,s,R)$	$R=0$	$R=1$
$s=2$	9	16
$s=3$	46	204
$s=4$	45	455
$s=5$	9	216
$s=6$	0	25

$Q(6,3,s,R)$	$R=0$	$R=1$	$R=2$
$s=2$	6	9	0
$s=3$	15	78	57
$s=4$	5	87	208
$s=5$	0	18	117
$s=6$	0	0	15

$Q(8,1,s,R)$	$R=0$
$s=2$	1
$s=3$	42
$s=4$	231
$s=5$	294
$s=6$	126
$s=7$	20
$s=8$	1

$Q(8,2,s,R)$	$R=0$	$R=1$
$s=2$	34	85
$s=3$	674	4324
$s=4$	1970	25519
$s=5$	1386	33600
$s=6$	308	14686
$s=7$	20	2360
$s=8$	0	119

$Q(8,3,s,R)$	$R=0$	$R=1$	$R=2$
$s=2$	72	418	0
$s=3$	732	892	10896
$s=4$	1218	27446	84526
$s=5$	504	19944	123612
$s=6$	56	4556	57128
$s=7$	0	304	9496
$s=8$	0	0	490

$Q(8,4,s,R)$	$R=0$	$R=1$	$R=2$	$R=3$
$s=2$	24	81	0	0
$s=3$	112	1208	3090	0
$s=4$	84	2018	11944	10209
$s=5$	14	800	9368	20688
$s=6$	0	86	2236	10908
$s=7$	0	0	156	1944
$s=8$	0	0	0	105

$Q^*(8,1,t,s,R)$:

$t=1$	$R=0$	$R=1$
$s=1$	1	0
$s=2$	34	93
$s=3$	72	894
$s=4$	24	1677
$s=5$	0	1050
$s=6$	0	266
$s=7$	0	28
$s=8$	0	1

$t=2$	$R=0$	$R=1$	$R=2$
$s=1$	34	85	0
$s=2$	610	6598	7905
$s=3$	792	20420	93742
$s=4$	168	13736	188515
$s=5$	0	3380	121570
$s=6$	0	408	31246
$s=7$	0	16	3316
$s=8$	0	0	119

$t=3$	$R=0$	$R=1$	$R=2$	$R=3$
$s=1$	72	418	0	0
$s=2$	792	14572	46866	0
$s=3$	792	25704	210638	236206
$s=4$	144	10104	192512	630730
$s=5$	0	1368	63134	449998
$s=6$	0	72	8700	121568
$s=7$	0	0	400	13320
$s=8$	0	0	0	490

$t=4$	$R=0$	$R=1$	$R=2$	$R=3$	$R=4$
$s=1$	24	81	0	0	0
$s=2$	168	3744	9423	0	0
$s=3$	144	4440	38472	58374	0
$s=4$	24	1296	21060	96384	59841
$s=5$	0	96	3816	37302	69036
$s=6$	0	0	216	5292	22422
$s=7$	0	0	0	240	2700
$s=8$	0	0	0	0	105

$Q^*(4,2,t,s,R)$:

$t=1$	$R=0$	$R=1$
$s=2$	3	5
$s=3$	37	171
$s=4$	56	596
$s=5$	21	555
$s=6$	2	186
$s=7$	0	24
$s=8$	0	1

$t=2$	$R=0$	$R=1$	$R=2$
$s=2$	47	415	490
$s=3$	274	5866	18612
$s=4$	226	9537	67825
$s=5$	46	3946	64552
$s=6$	2	480	21890
$s=7$	0	12	2844
$s=8$	0	0	119

$t=3$	$R=0$	$R=1$	$R=2$	$R=3$
$s=2$	50	744	3126	0
$s=3$	134	4930	42410	54446
$s=4$	54	4070	70998	244358
$s=5$	4	776	31452	250008
$s=6$	0	26	4422	87672
$s=7$	0	0	166	11594
$s=8$	0	0	0	490

$t=4$	$R=0$	$R=1$	$R=2$	$R=3$	$R=4$
$s=2$	8	108	724	0	0
$s=3$	10	412	4910	16508	0
$s=4$	2	186	4377	30778	33117
$s=5$	0	17	962	14607	44894
$s=6$	0	0	53	2266	17421
$s=7$	0	0	0	102	2418
$s=8$	0	0	0	0	105

$Q^*(2,3,t,s,R)$:

$t=1$	$R=0$	$R=1$
$s=2$	1	1
$s=3$	7	15
$s=4$	6	25
$s=5$	1	10
$s=6$	0	1

$t=2$	$R=0$	$R=1$	$R=2$
$s=2$	3	27	20
$s=3$	10	200	340
$s=4$	4	172	599
$s=5$	0	29	246
$s=6$	0	0	25

$t=3$	$R=0$	$R=1$	$R=2$	$R=3$
$s=2$	1	9	20	0
$s=3$	1	23	166	140
$s=4$	0	7	132	326
$s=5$	0	0	21	144
$s=6$	0	0	0	15

$Q^*(2,4,t,s,R)$:

$t=1$	$R=0$	$R=1$
$s=2$	1	1
$s=3$	31	63
$s=4$	90	301
$s=5$	65	350
$s=6$	15	140
$s=7$	1	21
$s=8$	0	1

$t=2$	$R=0$	$R=1$	$R=2$
$s=2$	11	135	92
$s=3$	157	4225	6804
$s=4$	222	11981	34326
$s=5$	69	8438	40878
$s=6$	5	1902	16538
$s=7$	0	124	2494
$s=8$	0	0	119

$t=3$	$R=0$	$R=1$	$R=2$	$R=3$
$s=2$	11	223	746	0
$s=3$	71	2974	23519	19496
$s=4$	48	3907	63319	124316
$s=5$	5	1171	42404	159770
$s=6$	0	85	9135	66730
$s=7$	0	0	572	10208
$s=8$	0	0	0	490

$t=4$	$R=0$	$R=1$	$R=2$	$R=3$	$R=4$
$s=2$	2	36	172	0	0
$s=3$	4	164	2393	7309	0
$s=4$	1	101	2865	20438	17650
$s=5$	0	11	820	12988	29756
$s=6$	0	0	58	2643	13574
$s=7$	0	0	0	156	2154
$s=8$	0	0	0	0	105

References

[1] R.F. Bass, K. Gröchenig, Random Sampling of Multivariate Trigonometric Polynomials, SIAM J. Math. Anal. 36(3), 773–795, 2004.

[2] S. Boucheron, G. Lugosi, P. Massart, A sharp concentration inequality with applications, Random Structures Algorithms 16, 277–292, 2000.

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

[4] D. Callan, On Conjugates for Set Partitions and Integer Compositions, Preprint arXiv:math.Co/0508052v3, 2005.

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

[6] E. Candes, J. Romberg, T. Tao, Robust Uncertainty Principles: Exact Signal Reconstruction from Highly Incomplete Frequency Information, Preprint arXiv:math.NA/0409186, 2004.

[7] E. Candes, T. Tao, Near Optimal Signal Recovery From Random Projections: Universal Encoding Strategies?, Preprint arXiv:math.CA/0410542, 2004.

[8] E. Candes, J. Romberg, T. Tao, Stable Signal Recovery from Incomplete and Inaccurate Measurements, Preprint 2005.

[9] E. Candes and J. Romberg, Practical Signal Recovery from Random Projections, Preprint 2004.

[10] D.L. Donoho, Compressed Sensing, Preprint, 2004.

[11] D.L. Donoho, Y. Tsaig, Extensions of Compressed Sensing, Preprint, 2004.

[12] D.L. Donoho, For most large underdetermined systems of linear equations the minimal ℓ^1-norm solution is also the sparsest solution, Preprint 2004.

[13] D.L. Donoho, J. Tanner, Sparse Nonnegative Solutions of Underdetermined Linear Equations, Preprint, 2005.

[14] A. Gilbert, J. Tropp, Signal Recovery from partial information via Orthogonal Matching Pursuit, Preprint, 2005.

[15] D. Knuth, Problem 11151, Amer. Math. Monthly 112, p. 367, 2005.

[16] D. Potts, G. Steidl, M. Tasche, Fast Fourier transforms for nonequispaced data: A tutorial, In: Modern Sampling Theory: Mathematics and Applications, J.J. Benedetto and P. Ferreira (Eds.), Chapter 12, pages 249–274. 1998.

[17] J. Riordan, *Combinatorial Analysis*, Wiley, 1958.

[18] M. Rudelson, R. Vershynin, Geometric approach to error correcting codes and reconstruction of signals, Preprint, arXiv:math.FA/0502299, 2005.

[19] R.P. Stanley, *Enumerative Combinatorics*, Combridge University Press, 1997.

[20] E.W. Weisstein, Binomial Sums, From MathWorld – A Wolfram Web Resource. `http://mathworld.wolfram.com/BinomialSums.html`.

[21] J. Zou, A. Gilbert, M. Strauss, I. Daubechies, Theoretical and experimental analysis of a randomized algorithm for Sparse Fourier transform analysis, J. Comp. Phys. 211, 572–595, 2005.