Random Sampling of Sparse Trigonometric Polynomials II - Orthogonal Matching Pursuit versus Basis Pursuit

Stefan Kunis, Holger Rauhut

April 24, 2006

Abstract

We continue investigating the problem of reconstructing a multivariate trigonometric polynomial having only few non-zero coefficients from few random samples. Both for a continuous and a discrete probability model for the sampling points we prove theoretical results on the success probability of reconstruction when using Orthogonal Matching Pursuit (OMP) or Basis Pursuit (BP). Although our theoretical estimates are the same for both methods, our numerical experiments indicate that OMP outperforms BP slightly. Moreover, OMP is significantly faster than BP in practice.

Key Words: random sampling, trigonometric polynomials, Orthogonal Matching Pursuit, Basis Pursuit, sparse recovery, set partitions, random matrices, fast Fourier transform, nonequispaced fast Fourier transform
AMS Subject classification: 94A20, 42A05, 15A52, 05A18, 90C05, 90C25

1 Introduction

Recently, the surprising fact that it is possible to recover functions having only few non-zero coefficients with respect to some basis from vastly incomplete information has gained much attention. Such functions are commonly called sparse or compressible and they naturally appear in a wide range of applications such as signal and image processing, e.g. biomedical imaging.

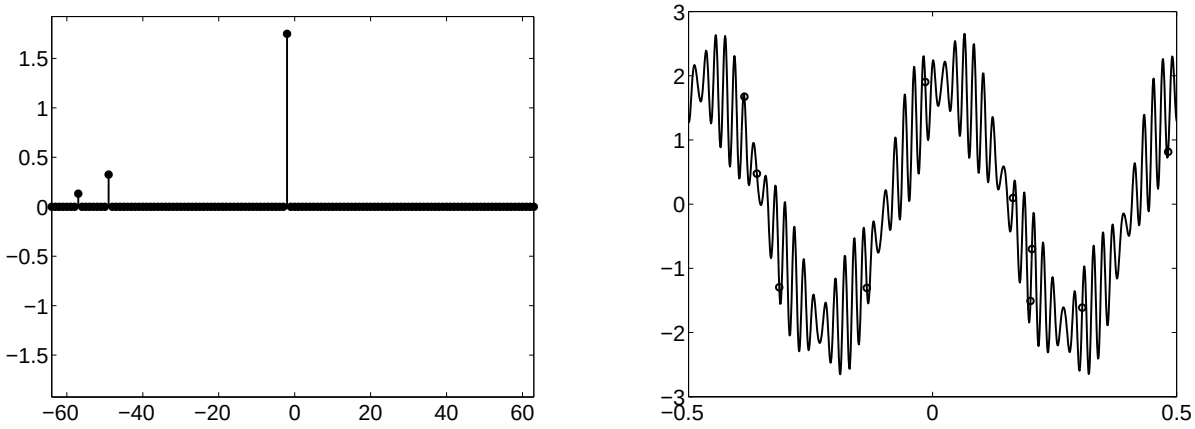
Previous work on this topic includes the reconstruction of Fourier coefficients from samples taken randomly on a lattice and the use of *Randomized Algorithms for Sparse Fourier Analysis* [33] or the *Basis Pursuit* (BP) principle [6, 7]. The latter approach consists in minimizing the ℓ^1-norm of the Fourier coefficients subject to the condition that the corresponding trigonometric polynomial matches the sampling points. Indeed, it was proven by Candes, Romberg and Tao in [6] in the setting of the discrete Fourier transform that this scheme recovers the coefficients exactly with high probability provided the number of samples is high enough compared to the sparsity, i.e., the number of non-vanishing coefficients. This result has been generalized by the second author of the present paper in [23] for the case of samples taken uniformly at random from the cube $[0, 2\pi]^d$.

Although there has been made much progress on interior point methods for solving convex problems (in particular ℓ^1-minimization) [4, 25, 20], reliable algorithms that take advantage of structured, but fully populated, matrices are somewhat rare. A second body of literature,

cf. [19, 16, 31, 32], suggests (Orthogonal) Matching Pursuit for sparse reconstruction tasks. Orthogonal Matching Pursuit (OMP) is a greedy approach that is conceptually simple to implement and potentially faster than BP. In particular, it may easily take into account fast algorithms for the involved matrices.

This paper is devoted to the theoretical and numerical investigation and comparison of the use of OMP and BP for the recovery of sparse trigonometric polynomials from randomly taken samples. Our results indicate that indeed both methods are suitable for this task. However, our numerical experiments suggest that OMP outperforms BP both with respect to computation time and – maybe more surprisingly – also slightly with respect to the success probability of reconstruction.

For related work on this topic, also known as compressed sensing, we refer to [6, 7, 8, 9, 10, 11, 12, 13, 14, 16, 26, 27] and the references therein. For more information on sampling of (not necessarily sparse) trigonometric polynomials in a random setting the reader may consult [1, 2].



(a) Sparse coefficient vector.

(b) Trigonometric polynomial and a few samples.

Figure 1: Sparse vector of Fourier coefficients and the corresponding trigonometric polynomial (real part). After sampling at a few randomly chosen points, the orthogonal matching pursuit algorithm (OMP), i.e., Algorithm 1, as well as the Basis pursuit principle recover the coefficient vector perfectly with high probability.

The paper is organized as follows: After introducing the necessary notation, including the orthogonal matching pursuit algorithm, we present the main result in Theorem 2.1. Subsequently, Corollary 2.2 reveals the asymptotic performance of OMP and BP in the present setting, whereas Corollary 2.3 comments on a slight modification within the discrete probability model. Commonly applied, the coherence parameter proves to be less useful here, cf. Subsection 2.3. In Section 3 all proofs of the obtained results are given, Section 4 presents numerical experiments, showing the superior performance of OMP over BP. Finally, Section 5 makes conclusions and discusses possible future work.

2 Main Results

2.1 The Setting

For some finite subset $\Gamma \subset \mathbb{Z}^d$, $d \in \mathbb{N}$, we let Π_Γ denote the space of all trigonometric polynomials in dimension d whose coefficients are supported on Γ. Clearly, an element f of Π_Γ is of the form

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

with some Fourier coefficients $c_k \in \mathbb{C}$. The dimension of Π_Γ will be denoted by $D := |\Gamma|$. Taking $\Gamma = \{-q, -q+1, \ldots, q\}^d$ yields the space $\Pi_\Gamma = \Pi_q^d$ of all trigonometric polynomials of maximal order q.

We will mainly deal 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 Γ. However, a priori nothing is known about T apart from a maximum size. Thus, it is useful to introduce the set $\Pi_\Gamma(M) \subset \Pi_\Gamma$ of all trigonometric polynomials whose Fourier coefficients are supported on a set $T \subset \Gamma$ satisfying $|T| \le M$. Note that

$$\Pi_\Gamma(M) = \bigcup_{T\subset\Gamma, |T|\le M} \Pi_T$$

is not a linear space.

Our aim is to sample a trigonometric polynomial f of $\Pi_\Gamma(M)$ at N points $x_1, \ldots, x_N \in [0,2\pi]^d$ and try to reconstruct f from these samples. If for some $m \in \mathbb{N}$ the sampling points are located on the grid

$$\frac{2\pi}{m}\mathbb{Z}_m^d = \left\{0, \frac{2\pi}{m}, \ldots, \frac{2\pi(m-1)}{m}\right\}$$

then this problem can also be interpreted as reconstructing a sparse vector from partial information on its discrete Fourier transform.

Basis Pursuit consists in solving the following ℓ^1-minimization problem

$$\min \|d\|_1 := \sum_{k\in\Gamma} |d_k| \quad \text{subject to} \quad \sum_{k\in\Gamma} d_k e^{2\pi ik\cdot x_j} = f(x_j), \quad j = 1, \ldots, N. \tag{2.1}$$

This task can be performed with convex optimization techniques [4]. For real-valued coefficients (2.1) can be reformulated as a linear program while for complex-valued coefficients we obtain a second order cone program. For both kind of problems standard software exists, such as MOSEK [20] or CVX [15] (internally using SeDuMi [28]) and since recently also L1MAGIC [25] (only for real-valued coefficients).

On the other hand, we use Orthogonal Matching Pursuit (OMP), cf. Algorithm 1, to recover the Fourier coefficients of f from a few samples. We need to introduce some notation. Let $X = \{x_1, \ldots, x_N\}$ be the set of (random) sampling points. We denote by $\mathcal{F}_X$ the $N \times D$ matrix (recall that $D = |\Gamma|$) with entries

$$(\mathcal{F}_X)_{j,k} = e^{ik\cdot x_j}, \quad 1 \le j \le N,\ k \in \Gamma. \tag{2.2}$$

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

$$\phi_k = \begin{pmatrix} e^{ik\cdot x_1} \\ \vdots \\ e^{ik\cdot x_N} \end{pmatrix},$$

Algorithm 1 OMP

Input: sampling set $X \subset [0, 2\pi]^d$, sampling vector $\mathrm{f} := (f(x_j))_{j=1}^N$, set $\Gamma \subset \mathbb{Z}^d$.
Optional: maximum allowed sparsity M or residual tolerance ε.

1: Set $s = 0$, the residual vector $r_0 = \mathrm{f}$, and the index set $T_0 = \emptyset$.
2: **repeat**
3: Set $s = s + 1$.
4: Find $k_s = \arg\max_{k\in\Gamma} |\langle r_{s-1}, \phi_k\rangle|$ and augment $T_s = T_{s-1} \cup \{k_s\}$.
5: Project onto $\mathrm{span}\{\phi_k, k \in T_s\}$ by solving the least squares problem

$$\|\mathcal{F}_{T_s X} d_s - \mathrm{f}\|_2 \xrightarrow{d_s} \min .$$

6: Compute the new residual $r_s = \mathrm{f} - \mathcal{F}_{T_s X} d_s$.
7: **until** $s = M$ or $\|r_s\| \le \varepsilon$
8: Set $T = T_s$, the non-zeros of the vector c are given by $(c_k)_{k\in T} = d_s$.

Output: vector of coefficients $(c_k)_{k\in\Gamma}$ and its support T.

so $\mathcal{F}_X = (\phi_{k_1}|\phi_{k_2}|\dots|\phi_{k_D})$. By

$$(\mathcal{F}_{TX})_{j,k} = e^{ik\cdot x_j}, \quad 1 \le j \le N,\, k \in T. \tag{2.3}$$

we denote the restriction of $\mathcal{F}_X$ to sequences supported only on T. Furthermore, let $\langle\cdot,\cdot\rangle$ denote the usual Euclidean scalar product and $\|\cdot\|_2$ the associated norm. We have $\|\phi_k\|_2 = \sqrt{N}$ for all $k \in \Gamma$, i.e., all the columns of $\mathcal{F}_X$ have the same ℓ^2-norm. We postpone a detailed discussion on the implementation of Algorithm 1 to Section 4.

Of course, the hope is that running OMP or BP on samples of some $f \in \Pi_\Gamma(M)$ will recover its Fourier coefficients. To analyze the performance of the algorithms we will use two probabilistic models for the sampling points (one for the continuous case and the other one for the discrete Fourier transform case). This random modeling should be understood in the sense that the sampling set X is 'generic': reconstruction is allowed to fail for certain choices of X, as long as the probability of encountering such a pathological case is very small.

We will work with the following two probability models for the sampling points:

- In our first model we assume that the sampling points $x_1, \dots, x_N$ are independent random variables having the uniform distribution on $[0, 2\pi]^d$. Obviously, the cardinality of the sampling set $X = \{x_1, \dots, x_N\}$ equals the number of samples N with probability 1.

- Our second model aims at analyzing the discrete Fourier transform, i.e., at studying the problem of reconstructing sparse vectors from (partial) information on its discrete Fourier transform. Indeed, we suppose that the sampling points $x_1, \dots, x_N$ have the uniform distribution on the finite set $\frac{2\pi}{m}\mathbb{Z}_m^d$. Moreover, it will always be assumed implicitly that $\Gamma \subset \mathbb{Z}_m^d$ when we discuss this second model.

  Observe that it happens with non-zero probability that some of the sampling points coincide, so the cardinality of the sampling set $X = \{x_1, \dots, x_N\}$ might be smaller than

N. However, this fact will not do much harm to our analysis. Moreover, we will later be able to replace this model by the one where the random sampling set X is chosen uniformly at random from all the subsets of $\frac{2\pi}{m}\mathbb{Z}_m^d$ of size N.

We will often refer to the first model as the "continuous model" while the second will be called the "discrete model". It turns out that one can treat both probability models in parallel. We will have to make a slight distinction only at one point of the proof of our results.

2.2 The Main Theorem

Although originally, our main aim was on investigating OMP, it surprisingly turned out that our method for approaching the recovery problem is applicable to both OMP and BP at the same time. Indeed, it is based on results by Tropp in [31, 32] that apply to both schemes. Also, our result is very similar to Theorem 2.1 in [23] concerning BP, although the proof is somewhat different there. In order to formulate it we first need to introduce functions $F_n(\theta)$, $n \in \mathbb{N}$, by

$$F_n(\theta) \;=\; \sum_{k=1}^{\lfloor n/2 \rfloor} S_2(n,k)\theta^k, \qquad \theta \in \mathbb{R}, \tag{2.4}$$

where $S_2(n,k)$ are the associated Stirling numbers of the second kind. They can be computed by means of their exponential generating function, see [24, formula (27), p.77] or Sloane's A008299 in [30],

$$\sum_{n=1}^{\infty} F_n(\theta)\frac{x^n}{n!} \;=\; \exp(\theta(e^x - x - 1)). \tag{2.5}$$

Further, we define

$$G_n(\theta) \;:=\; \theta^{-n}F_n(\theta). \tag{2.6}$$

Then our main theorem reads as follows.

Theorem 2.1. *Assume $f \in \Pi_\Gamma(M)$, with some sparsity $M \in \mathbb{N}$ and $\Gamma \subset \mathbb{Z}^d$ with $|\Gamma| = D$. Let $x_1, \ldots, x_N \in [0, 2\pi]^d$ be independent random variables either*

(a) having the uniform distribution on $[0, 2\pi]^d$, or

(b) having the uniform distribution on $\frac{2\pi}{m}\mathbb{Z}_m^d = \left\{0, \frac{2\pi}{m}, \ldots, \frac{2\pi(m-1)}{m}\right\}^d$.

Choose $n \in \mathbb{N}$, $\beta > 0, 0 < \kappa < 1$ and $K_1, \ldots, K_n \in \mathbb{N}$ such that

$$\sum_{\nu=1}^{n} \beta^{n/K_\nu} \;\le\; \left(1 + \sqrt{M}\frac{\kappa}{1-\kappa}\right)^{-1} \tag{2.7}$$

Set the 'generalized oversampling factor' $\theta := N/M$. Then with probability at least

$$1 - \left(M\kappa^{-2}G_{2n}(\theta) + (D-M)\beta^{-2n}\sum_{\nu=1}^{n} G_{2\nu K_\nu}(\theta)\right) \tag{2.8}$$

the sparse trigonometric polynomial f can be reconstructed exactly from its sample values $f(x_1), \ldots, f(x_N)$ via Orthogonal Matching Pursuit and also via Basis Pursuit.

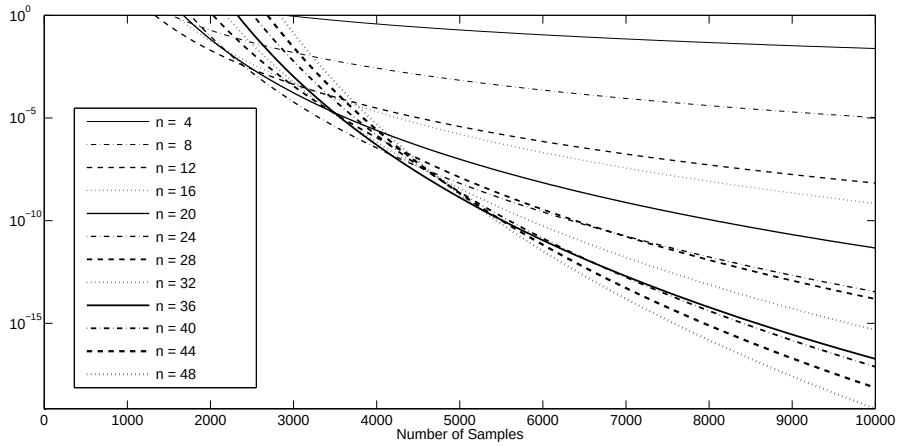


Figure 2: Bound for the probability of *failure* of exact reconstruction by OMP and BP due to Theorem 2.1 for different choices of n. (The other parameters were chosen (near) optimal). The sparsity is $M = 10$, the dimension $D = 10000$ and the number of samples N is varied.

The theorem allows to compute explicit numerical bounds for the probability of exact reconstruction via OMP and via BP given the sparsity M, the number of samples N and the dimension D of Π_Γ, see Figure 2. However, since it is not very obvious to interpret we also provide the following corollary, which gives more information on the asymptotic behavior of the probability bound.

Corollary 2.2. *There exists an absolute constant $C > 0$ such that the following is true. Assume $f \in \Pi_\Gamma(M)$ for some sparsity $M \in \mathbb{N}$ and $|\Gamma| = D$. Let $x_1, \dots, x_N \in [0, 2\pi]^d$ be independent random variables having the uniform distribution either on $[0, 2\pi]^d$ or on $\frac{2\pi}{m}\mathbb{Z}_m^d$. If for some $\epsilon > 0$ it holds*

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

then with probability at least $1 - \epsilon$ the trigonometric polynomial f can be recovered from its sample values $f(x_j), j = 1, \dots, N$, via Orthogonal Matching Pursuit and also via Basis Pursuit.

In particular, if the sparsity M is small and the dimension D large then we may choose the number N of samples much smaller than D (but larger than M), and we are still able to recover a polynomial $f \in \Pi_\Gamma(M)$ exactly – at least with high probability.

As already remarked taking random samples from the uniform distribution on $\frac{2\pi}{m}\mathbb{Z}_m^d$ has the disadvantage that some of the sampling points $x_1, \dots, x_N$ might coincide with non-zero probability. The following corollary removes this drawback.

Corollary 2.3. *Choose $X = \{x_1, \dots, x_N\}$ uniformly at random from all subsets of $\frac{2\pi}{m}\mathbb{Z}_m^d$ of size N. Then Theorem 2.1 and Corollary 2.2 still hold.*

This result corresponds precisely to the random model used by Candes, Romberg and Tao in Theorem 1.3 of [6]. Since our result is also applicable to Basis Pursuit we are even able to provide an alternative proof of their theorem, which is considerably shorter. Nevertheless we were still influenced by nice ideas of Candes et al. in [6].

Let us shortly comment on the proof of Theorem 2.1. First, it is not possible to apply directly ideas of Gilbert and Tropp in [16] concerning OMP because they assume that the columns of the measurement matrix (in our case the vectors ϕ_k) are stochastically independent (see property (M2) in [16, Section 3.1]). Clearly, this property does not hold in our case. Only the rows of $\mathcal{F}_X$ are stochastically independent.

Instead, the proof is based on investigating the exact reconstruction condition (ERC) as introduced by Tropp, see [31, Theorem A], [32]. This is the reason why our analysis is also applicable to Basis Pursuit. It even improves (very slightly) our previous result in [23]. However, although we start differently here, in the end, surprisingly, we have to do very similar computations as in [23]. So parts of our proof rely on auxiliary results in [23].

Furthermore, we note that we may also model the support set T of the Fourier coefficients as a random set as in [23]. This would rather correspond to an average case analysis than to a worst case analysis, as provided by Theorem 2.1. The methods in [23] can easily be applied also to our situation here, so one may formulate the analogue of Theorem 2.3 in [23] without difficulties. In particular, this would yield an improved bound for the probability of reconstruction compared to Theorem 2.1, see also Remark (a) in Section 3.6 in [23]. However, since a related combinatorial problem (stated in [23]) has not yet been solved, the corresponding estimate is not easy to interpret, and hence, we abstain from giving it here.

2.3 A Result based on Coherence

Many results for OMP rely on the so called coherence parameter μ [31, 32]. It measures the correlation of different columns of the measurement matrix $\mathcal{F}_X$. It requires that they have unit norm, so we define $\widetilde{\phi}_k := N^{-1/2}\phi_k$ resulting in $\|\widetilde{\phi}_k\|_2 = 1$. Then μ is defined as

$$\mu := \max_{j \neq k} |\langle \widetilde{\phi}_j, \widetilde{\phi}_k \rangle| = N^{-1} \max_{j \neq k} |\langle \phi_j, \phi_k \rangle|. \tag{2.9}$$

Reformulating Corollary 3.6 in [31] for our context yields the following result.

Theorem 2.4. *Suppose that $\mathcal{F}_{TX}$ is injective for all subsets $T \subset \Gamma$ with $|T| \leq M$. Assume $(2M-1)\mu < 1$. Then both OMP and BP recover* every *$f \in \Pi_\Gamma(M)$.*

Proof: The injectivity of $\mathcal{F}_{TX}$ for all T with $|T| \leq M$ is needed to ensure that if $f \in \Pi_\Gamma(M)$ has support T then it is the unique polynomial in Π_T having the vector $(f(x_1), \ldots, f(x_N))$ of sample values. The rest of the proof is the same as the one of Corollary 3.6 in [31]. ■

In order to state a result on the coherence parameter μ we first introduce a variant $S^1(n,t)$ of the Stirling numbers by means of the following generating function, see also [29, formula (18)],

$$\sum_{n=1}^{\infty} \sum_{t=1}^{n} S^1(n,t) x^t \frac{z^n}{(n!)^2} = \exp\left(x \sum_{n=1}^{\infty} \frac{z^n}{(n!)^2} \right). \tag{2.10}$$

A table of these numbers is given in [29, Table II]. (The proof of the following result will even reveal a combinatorial interpretation of them, which does not seem to be known yet.)

Theorem 2.5. *Assume $M, N \in \mathbb{N}$, $\epsilon > 0$ and $\Gamma \subset \mathbb{Z}^d$. Choose $x_1, \ldots, x_N$ independent random variable having the uniform distribution either on $[0, 2\pi]^d$ or on $\frac{2\pi}{m}\mathbb{Z}_m^d$. Let $\Gamma' = \{j - k : j, k \in$*

$\Gamma, j \neq k\}$ *and* $D' = |\Gamma'|$. *Suppose*

$$D' \left(\frac{2M-1}{N} \right)^{2n} \sum_{t=1}^{n} S^1(n,t) N^t \leq \epsilon \tag{2.11}$$

for some $n \in \mathbb{N}$, *which is assumed to be less than* m *in the discrete probability model. Then with probability at least* $1-\epsilon$ *the coherence satisfies* $(2M-1)\mu < 1$.

Observe that the influence of D' and ϵ decreases in (2.11) when n gets larger. For given parameters D', M, N one may optimize (2.11) with respect to n to obtain a good probability bound. We abstain from giving a detailed analysis here, but rather suggest to do this task numerically. Also observe that if $\Gamma = \{-q, \dots, q\}^d$ for instance then $\Gamma' = \{-2q, \dots, 2q\}^d \setminus \{0\}$ and $D' = (4q+1)^d - 1$.

In the continuous probability model the matrices $\mathcal{F}_{TX}$ are injective almost surely if $|T| \leq N$ by Lemma 3.2 in [23]. So combining Theorem 2.4 and 2.5 gives

Theorem 2.6. *Make the same assumptions on* M, N, ϵ *and* Γ *as in Theorem 2.5. Let* $x_1, \dots, x_N$ *be independent random variables having the uniform distribution on* $[0, 2\pi]^d$. *Then with probability at least* $1-\epsilon$ every $f \in \Pi_\Gamma(M)$ *can be recovered both by OMP and BP.*

The main difference of the previous result to Theorem 2.1 is the uniformity in f. Indeed here, once the $x_1, \dots, x_N$ are chosen then reconstruction is successful *for all* f with a certain probability, whereas in Theorem 2.1 the probability of reconstruction is estimated for a (deterministically) given single f. (Although inspecting the proof of Theorem 2.1 shows that with the stated probability the reconstruction is successful for all f with a given support T of their Fourier coefficients.) In other words, we cannot deduce from Theorem 2.1 that a single sampling set X can be equally good for all sparse f (resp. all support sets T).

Having a closer look at condition (2.11) reveals that for fixed D and ϵ we have to choose $N \geq C_{D,\epsilon} M^2$. Thus, the number N of samples that Theorem 2.5 suggests to take are far too many for practical purposes, especially when comparing this observation with Corollary 2.2. It is interesting to note that Gilbert and Tropp obtain a similar result for Gaussian and Bernoulli measurements [16, Proposition 10]. So it seems that coherence is not the optimal concept for studying the recovery problem. To further support this statement we note that one can hardly improve Theorem 2.5 in this regard by using other methods than ours. Indeed, its proof shows that $\mathbb{E}[|\langle \phi_j, \phi_k \rangle|^2] = S^1(1,1)N = N$ for $k \neq j$, see (3.20). Thus, it does not seem very unlikely that $|\langle \phi_j, \phi_k \rangle| \geq \sqrt{N}$ for at least one pair (k, j), in particular, $\mu = N^{-1} \max_{j \neq k} |\langle \phi_j, \phi_k \rangle| \geq N^{-1/2}$. This means that it is rather likely that $(2M-1)\mu > 1$ if $(2M-1)^2 \geq N$. (We are, of course, aware that this is not a precise mathematical argument, but nevertheless gives some intuition.)

A refinement of the coherence parameter is given by the cumulative coherence function, defined as

$$\mu_1(M) = N^{-1} \max_{T \subset \Gamma, |T| = M} \max_{j \in \Gamma \setminus T} \sum_{k \in T} |\langle \phi_j, \phi_k \rangle|.$$

It is easy to see that $\mu_1(M) \leq M\mu$. Theorem 2.4 still holds if the condition $(2M-1)\mu$ is replaced by $\mu_1(M) + \mu_1(M-1) < 1$, see [31, Theorem B]. (Clearly, the latter implies the first condition.) One might ask whether in our situation an investigation of $\mu_1(M)$ would give better results than analyzing μ. However, a similar informal argument as above leaves little hope. Indeed, as $\mathbb{E}[|\langle \phi_j, \phi_k \rangle|^2] = N$ for all $j \neq k$, it seems rather likely that $\sum_{k \in T} |\langle \phi_j, \phi_k \rangle| \geq M\sqrt{N}$

for some T with $|T| = M$ and as a consequence, $\mu_1(M) + \mu_1(M-1) \geq (2M-1)N^{-1/2}$. In particular, using the cumulative coherence function does not seem to improve our results based on the coherence μ.

Of course, we may also formulate a result analogous to Theorem 2.6 for the discrete probability model. However, the almost sure injectivity of $\mathcal{F}_{TX}$ is not guaranteed anymore. The proof of Theorem 2.1 contains an analysis of the probability of injectivity of $\mathcal{F}_{TX}$ (for a single T). So in principle we could state the analogue of Theorem 2.6 but the probability estimate would be even worse than (2.11). We remark that based on the so called Uniform Uncertainty Principle, Candes and Tao were able to formulate much better results for the discrete probability model that hold uniformly for all sparse f [7, 8], see also further improved results by Rudelson and Vershynin [27].

3 Proof of the Main Results

3.1 Proof of Theorem 2.1

Let us first introduce some notation. By $\ell^p(T), \ell^p(X)$, $p = 1, 2$ we denote the usual ℓ^p-spaces of sequences indexed by $T \subset \Gamma$ and X, respectively, endowed with the usual norm $\|c\|_p = (\sum_k |c_k|^p)^{1/p}$. The matrices $\mathcal{F}_X$, $\mathcal{F}_{TX}$ defined in (2.2) and (2.3) can be identified with operators from $\ell^2(\Gamma)$ and $\ell^2(T)$, resp., into $\ell^2(X)$. The adjoint operators are denoted by $\mathcal{F}_X^* : \ell^2(X) \to \ell^2(\Gamma)$ and $\mathcal{F}_{TX}^* : \ell^2(X) \to \ell^2(T)$. Further, let $\mathcal{F}_{TX}^\dagger : \ell^2(X) \to \ell^2(T)$ be the Moore-Penrose pseudo-inverse of $\mathcal{F}_{TX}$. If $\mathcal{F}_{TX}^*\mathcal{F}_{TX}$ is invertible then it is given by $\mathcal{F}_{TX}^\dagger = (\mathcal{F}_{TX}^*\mathcal{F}_{TX})^{-1}\mathcal{F}_{TX}^*$.

The next lemma is the key to our proof. It is essentially due to Tropp [31, Theorem A], [32, Theorem 5.2].

Lemma 3.1. *Let T denote the support of the Fourier coefficients of $f \in \Pi_\Gamma(M)$. Assume $\mathcal{F}_{TX}$ is injective and*

$$\max_{j \in \Gamma \setminus T} \|\mathcal{F}_{TX}^\dagger \phi_j\|_1 < 1. \tag{3.1}$$

Then f can be recovered from its sample values on the set X via OMP and also via Basis Pursuit.

Proof: The injectivity of $\mathcal{F}_{TX}$ ensures that there does not exist another $\tilde{f}$ whose Fourier coefficients are also supported on T such that $\tilde{f}(x_\ell) = f(x_\ell)$ for all $\ell = 1, \ldots, N$. (In the context of sparse approximation this property is satisfied automatically for the sparsest representation, and thus the corresponding hypothesis is not needed in [31].) The rest of the proof is completely analogous to the one of Theorem A in [31], and hence, we omit it here. We just note that the set Λ_{opt} in [31] corresponds to T in our situation. ∎

We remark that (3.1) is called the exact reconstruction condition. In the following we will prove in several steps that it holds with high probability, and Theorem 2.1 will follow.

Assume for the moment that $\mathcal{F}_{TX}$ is injective, which is equivalent to $\mathcal{F}_{TX}^*\mathcal{F}_{TX}$ being invertible. Later this will follow automatically with a certain probability, and in the continuous probability model it is even true almost surely if $N \geq |T|$ by Lemma 3.2 in [23], see also Section 3 in [1]. Then

$$\mathcal{F}_{TX}^\dagger \phi_j = (\mathcal{F}_{TX}^*\mathcal{F}_{TX})^{-1}\mathcal{F}_{TX}^*\phi_j.$$

Observe that

$$v^{(j)} := \mathcal{F}_{TX}^* \phi_j = (\langle \phi_j, \phi_k \rangle)_{k \in T} \in \ell^2(T), \quad j \in \Gamma \setminus T.$$

The entries of $v^{(j)}$ are given by

$$v_k^{(j)} = \sum_{\ell=1}^{N} e^{i(j-k)\cdot x_\ell}, \quad k \in T, \quad j \in \Gamma \setminus T. \tag{3.2}$$

As in [23] and [6] we introduce the operator

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

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

$$(H_0 d)_k = -\sum_{\ell=1}^{N} \sum_{\substack{j \in T \\ j \neq k}} d_j e^{i(j-k)\cdot x_\ell}. \tag{3.3}$$

The Frobenius norm of a matrix A is defined as $\|A\|_F^2 := \mathrm{Tr}(A^* A) = \sum_{j,k} |A_{j,k}|^2$.

Lemma 3.2. *Let $n \in \mathbb{N}$. If*

$$\|(N^{-1} H_0)^n\|_F < \kappa < 1 \tag{3.4}$$

then $\mathcal{F}_{TX}^* \mathcal{F}_{TX}$ is invertible. If for $j \in \Gamma \setminus T$ additionally, the estimate

$$\sum_{\nu=0}^{n-1} N^{-\nu-1} \|H_0^\nu v^{(j)}\|_2 < |T|^{-1/2} \left(1 + \sqrt{|T|} \frac{\kappa}{1-\kappa}\right)^{-1} =: a \tag{3.5}$$

is fulfilled, then it follows that $\|\mathcal{F}_{TX}^\dagger \phi_j\|_1 < 1$.

Proof: By the von Neumann series we can write

$$\left(I_T - \left(\frac{1}{N} H_0\right)^n\right)^{-1} = I_T + A_n \quad \text{with} \quad A_n := \sum_{r=1}^{\infty} (\frac{1}{N} H_0)^{rn} \tag{3.6}$$

and due to (3.4) this series converges with $\|A_n\|_F < \frac{\kappa}{1-\kappa}$. Moreover, the identity $(I - A)^{-1} = (I - A^n)^{-1}(I + A + \cdots + A^{n-1})$ yields

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

and thus, $\mathcal{F}_{TX}^* \mathcal{F}_{TX}$ is invertible. Furthermore, we obtain

$$\begin{aligned} \mathcal{F}_{TX}^\dagger \phi_j &= (N^{-1} \mathcal{F}_{TX}^* \mathcal{F}_{TX})^{-1} N^{-1} \mathcal{F}_{TX}^* \phi_j = (I_T - N^{-1} H_0)^{-1} (N^{-1} v^{(j)}) \\ &= (I_T + A_n) \sum_{\nu=0}^{n-1} \left(N^{-1} H_0\right)^\nu (N^{-1} v^{(j)}) = (I_T + A_n) \sum_{\nu=0}^{n-1} N^{-\nu-1} H_0^\nu v^{(j)}. \end{aligned}$$

for $j \in \Gamma \setminus T$. Taking the ℓ^1-norm and using the triangle inequality yields

$$\|F_{TX}^\dagger \phi_j\|_1 \;\le\; \|I_T + A_n\|_{\ell^1\to\ell^1} \|\sum_{\nu=0}^{n-1} N^{-\nu-1} H_0^\nu v^{(j)}\|_1 \;\le\; (1 + \|A_n\|_{\ell^1\to\ell^1}) \sum_{\nu=0}^{n-1} N^{-\nu-1} \|H_0^n v^{(j)}\|_1.$$

The operator norm of A_n from $\ell^1(T)$ to $\ell^1(T)$ is given by $\|A_n\|_{\ell^1\to\ell^1} = \max_{k\in T} \sum_{j\in T} |(A_n)_{j,k}|$ and by the Cauchy-Schwarz inequality it can be estimated as

$$\|A_n\|_{\ell^1\to\ell^1}^2 \;\le\; |T| \max_{k\in T} \sum_{j\in T} |(A_n)_{j,k}|^2 \;\le\; |T| \sum_{j,k} |(A_n)_{j,k}|^2 \;=\; |T| \|A_n\|_F^2.$$

Invoking once more the Cauchy-Schwarz inequality gives $\|H_0^\nu v^{(j)}\|_1 \le \sqrt{T} \|H_0^\nu v^{(j)}\|_2$, and altogether

$$\|\mathcal{F}_{TX}^\dagger \phi_j\|_1 \;\le\; (1 + \sqrt{|T|}\|A_n\|_F) \sqrt{|T|} \sum_{\nu=0}^{n-1} N^{-\nu-1} \|H_0^\nu v^{(j)}\|_2. \tag{3.7}$$

which is less than one by assumption (3.5). ∎

Altogether we can estimate the probability that condition (3.1) does *not* hold as follows,

$$\begin{aligned}
\mathbb{P}(\max_{j\in\Gamma\setminus T} \|\mathcal{F}_{TX}^\dagger \phi_j\|_1 \ge 1) \;&\le\; \mathbb{P}\left(\{\|(N^{-1}H_0)^n\|_F \ge \kappa\} \cup \left\{ \max_{j\in\Gamma\setminus T} \sum_{\nu=0}^{n-1} N^{-\nu-1} \|H_0^\nu v^{(j)}\|_2 \ge a \right\} \right) \\
&\le\; \mathbb{P}(\|(N^{-1}H_0)^n\|_F \ge \kappa) + \sum_{j\in\Gamma\setminus T} \mathbb{P}\left(\sum_{\nu=0}^{n-1} N^{-\nu-1} \|H_0^\nu v^{(j)}\|_2 \ge a \right).
\end{aligned} \tag{3.8}$$

In the second inequality we have used the union bound (also known as Boole's inequality). The probability that $\|(N^{-1}H_0)^n\|_F$ is larger than κ was already estimated in [23]. Indeed, in Section 3.3 of [23] (see also Lemma 3.3 and Remark (b) in Section 3.6 in [23]) it is stated that

$$\mathbb{P}(\|(N^{-1}H_0)^n\|_F \ge \kappa) \;\le\; \kappa^{-2} M G_{2n}(N/M). \tag{3.9}$$

(Recall that $M = |T|$ and G_{2n} was defined in (2.6).) Thus, it remains to estimate the second term in (3.8). To this end we proceed similarly as in [23, Section 3.2].

Let $\beta > 0$ and $K_\nu \in \mathbb{N}$, $\nu = 1, \dots, n$, such that

$$\sum_{\nu=1}^{n} \beta^{n/K_\nu} \;\le\; \left(1 + \sqrt{|T|}\frac{\kappa}{1-\kappa}\right)^{-1} \;=\; \sqrt{|T|}a.$$

Further let $j \in \Gamma \setminus T$. By the pigeonhole principle we obtain

$$\begin{aligned}
\mathbb{P}\left(\sum_{\nu=0}^{n-1} N^{-\nu-1} \|H_0^\nu v^{(j)}\|_2 \ge a \right) \;&\le\; \sum_{\nu=0}^{n-1} \mathbb{P}\left(N^{-\nu-1} \|H_0^\nu v^{(j)}\|_2 \ge \beta^{n/K_{\nu+1}} |T|^{-1/2} \right) \\
&= \sum_{\nu=0}^{n-1} \mathbb{P}\left(\|H_0^\nu v^{(j)}\|_2^{2K_{\nu+1}} \ge \beta^{2n} N^{2(\nu+1)K_{\nu+1}} |T|^{-K_{\nu+1}} \right) \\
&\le \sum_{\nu=0}^{n-1} \mathbb{E}\left[\|H_0^\nu v^{(j)}\|_2^{2K_{\nu+1}} \right] \beta^{-2n} N^{-2(\nu+1)K_{\nu+1}} |T|^{K_{\nu+1}}.
\end{aligned} \tag{3.10}$$

In the last inequality we have used Markov's inequality. So it remains to investigate the expected value of $\|H_0^\nu v^{(j)}\|_2^{2K_{\nu+1}}$.

Lemma 3.3. *Let $\nu \in \mathbb{N}_0$, $K \in \mathbb{N}$ and $j \in \Gamma \setminus T$. Then*

$$\mathbb{E}\left[\|H_0^\nu v^{(j)}\|_2^{2K}\right] \leq |T|^{2K\nu+K} F_{2K(\nu+1)}\left(\frac{N}{|T|}\right)$$

with F_n defined in (2.4).

Proof: Let $k_0 \in T$. Using (3.3) and formula (3.2) for $v^{(j)}$ we obtain after an elementary calculation

$$\begin{aligned}(H_0^\nu v^{(j)})_{k_0} &= (-1)^\nu \sum_{\ell_1,\ldots,\ell_\nu=1}^{N} \sum_{\substack{k_1,\ldots,k_\nu \in T \\ k_{r-1}\neq k_r, r=1,\ldots,\nu}} v_{k_\nu}^{(j)} \prod_{r=1}^{\nu} e^{i(k_r-k_{r-1})\cdot x_{\ell_r}} \\ &= (-1)^\nu \sum_{\ell_1,\ldots,\ell_\nu,\ell_{\nu+1}=1}^{N} \sum_{\substack{k_1,\ldots,k_\nu \in T \\ k_{r-1}\neq k_r, r=1,\ldots,\nu}} e^{i(j-k_\nu)\cdot x_{\ell_{\nu+1}}} e^{i(k_\nu-k_{\nu-1})\cdot x_{\ell_\nu}} \cdots e^{i(k_1-k_0)\cdot x_{\ell_1}}.\end{aligned}$$

(If $\nu = 0$ then the second summation symbol in the previous formula disappears.) Defining for simplicity of notation $k_{\nu+1}^{(p)} := j$ and $k_0^{(p)} := k_0$, $p = 1, 2$, we obtain

$$|(H_0^\nu v^{(j)})_{k_0}|^2 = \sum_{\substack{\ell_1^{(1)},\ldots,\ell_{\nu+1}^{(1)}=1 \\ \ell_1^{(2)},\ldots,\ell_{\nu+1}^{(2)}=1}}^{N} \sum_{\substack{k_1^{(1)},\ldots,k_\nu^{(1)} \in T \\ k_1^{(2)},\ldots,k_\nu^{(2)} \in T \\ k_{r-1}^{(p)} \neq k_r^{(p)}}} \left(\prod_{r=1}^{\nu+1} e^{i(k_r^{(1)}-k_{r-1}^{(1)})\cdot x_{\ell_r^{(1)}}}\right)\left(\prod_{r=1}^{\nu+1} e^{-i(k_r^{(2)}-k_{r-1}^{(2)})\cdot x_{\ell_r^{(2)}}}\right).$$

Summing over k_0 yields

$$\|H_0^\nu v^{(j)}\|_2^2 = \sum_{\substack{\ell_1^{(1)},\ldots,\ell_{\nu+1}^{(1)}=1 \\ \ell_1^{(2)},\ldots,\ell_{\nu+1}^{(2)}=1}}^{N} \sum_{k_0 \in T} \sum_{\substack{k_1^{(1)},\ldots,k_\nu^{(1)} \in T \\ k_1^{(2)},\ldots,k_\nu^{(2)} \in T \\ k_{r-1}^{(p)} \neq k_r^{(p)}, (p,r)\in[2]\times[\nu]}} \prod_{p=1}^{2}\prod_{r=1}^{\nu+1} \exp\left(i(-1)^p (k_r^{(p)} - k_{r-1}^{(p)})\cdot x_{\ell_r^{(p)}}\right).$$

Taking the K-th power gives

$$\|H_0^\nu v^{(j)}\|_2^{2K} = \sum_{\substack{\ell_1^{(1)},\ldots,\ell_{\nu+1}^{(1)}=1 \\ \vdots \\ \ell_1^{(2K)},\ldots,\ell_{\nu+1}^{(2K)}=1}}^{N} \sum_{\substack{k_0^{(1)},\ldots,k_\nu^{(1)} \in T \\ \vdots \\ k_0^{(2K)},\ldots,k_\nu^{(2K)} \in T \\ k_{r-1}^{(p)} \neq k_r^{(p)},\, (p,r)\in[2K]\times[\nu] \\ k_0^{(2p-1)} = k_0^{(2p)},\, p=1,\ldots,K}} \prod_{p=1}^{2K}\prod_{r=1}^{\nu+1} \exp\left(i(-1)^p (k_r^{(p)} - k_{r-1}^{(p)})\cdot x_{\ell_r^{(p)}}\right), \tag{3.11}$$

where similarly as above we agree on setting $k_{\nu+1}^{(p)} = j$ for $p = 1, \ldots, 2K$. Note that since $j \in \Gamma \setminus T$ the condition $k_r^{(p)} \neq k_{r-1}^{(p)}$ holds also for $r = \nu + 1$. For simpler notation we denote by $I(T, K, \nu)$ the set of indices $(k_r^{(p)})_{r=0,\ldots,\nu}^{p=1,\ldots,2K}$ satisfying all the conditions specified under the

second summation symbol in the previous formula. Using linearity of expectation we obtain

$$\mathbb{E}\left[\|H_0^\nu v^{(j)}\|_2^{2K}\right] = \sum_{\substack{\ell_1^{(1)},\dots,\ell_{\nu+1}^{(1)}=1 \\ \vdots \\ \ell_1^{(2K)},\dots,\ell_{\nu+1}^{(2K)}=1}}^{N} \sum_{(k_r^{(p)})\in I(K,\nu)} \mathbb{E}\left[\prod_{p=1}^{2K}\prod_{r=1}^{\nu+1} \exp\left(i(-1)^p(k_r^{(p)}-k_{r-1}^{(p)})\cdot x_{\ell_r^{(p)}}\right)\right]. \tag{3.12}$$

Let us investigate the expectation appearing in the previous expression. We would like to somehow use the stochastic independence of the random variables x_ℓ. However, we have to be careful since some of the indices $\ell_r^{(p)}$ in the expectation in (3.12) might coincide. As in [23] and [6] this is the point where set partitions enter the game. So before we continue with the proof we introduce some notation and background on set partitions.

Let Q be a finite set. A partition of Q is a set of subsets of Q called blocks such that each element of Q is contained in precisely one of the blocks. By $P(Q,t)$ we denote the set of all partitions of Q into precisely t blocks such that each block contains at least 2 elements. It is well known that the number of such partitions is given by the associated Stirling numbers of the second kind (see also (2.4), (2.5)), i.e., $|P(Q,t)| = S_2(|Q|,t)$, see e.g. [24]. Observe that $P(Q,t)$ is empty if $t \geq |Q|/2$. Here, we will encounter partitions of sets of the form

$$[K]\times[\nu] := \{(a,b), a\in\{1,\dots,K\}, b\in\{1,\dots,\nu\}\}.$$

where $|[K]\times[\nu]| = K\nu$ and $|P([K]\times[\nu],t)| = S_2(K\nu,t)$.

Now given a vector $(\ell_r^{(p)})_{r=1,\dots,\nu+1}^{p=1,\dots,2K} \in \{1,\dots,N\}^{2K(\nu+1)}$ we associate a partition $\mathcal{A}$ of $[2K]\times[\nu+1]$ such that $\ell_r^{(p)} = \ell_{r'}^{(p')}$ if and only if (p,r) and (p',r') are contained in the same block of the partition. This allows us to unambiguously write $\ell_A = \ell_r^{(p)}$ if $(p,r)\in A\in\mathcal{A}$. Clearly, $\mathcal{A}$ cannot have more than N blocks since the $\ell_r^{(p)}$ may take at most N different values. Since by construction of $\mathcal{A}$ all the ℓ_A, $A\in\mathcal{A}$, are different from each other, we are allowed to use the stochastic independence of the x_ℓ to write the expectation appearing in the sum of the right hand side of (3.12) as

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

Now we have to make a small distinction between the continuous and the discrete probability model for the random variables x_ℓ.

1. In case the x_ℓ have the uniform distribution on $[0,2\pi]^d$ we obtain

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

where δ denotes the Kronecker-Delta.

2. In case of the uniform distribution on $\frac{2\pi}{m}\mathbb{Z}_m^d$ we get

$$\begin{aligned}
&\mathbb{E}\left[\exp\left(i\sum_{(p,r)\in A}(-1)^p(k_r^{(p)}-k_{r-1}^{(p)})\cdot x_{\ell_A}\right)\right]\\
&=\frac{1}{m^d}\sum_{u\in\mathbb{Z}_m^d}\exp\left(i\sum_{(p,r)\in A}(-1)^p(k_r^{(p)}-k_{r-1}^{(p)})\cdot\frac{2\pi u}{m}\right)\\
&=\delta\left(\sum_{(p,r)\in A}(-1)^p(k_r^{(p)}-k_{r-1}^{(p)})\mod m\right).
\end{aligned}$$

Thus, the only difference to the continuous model is the calculation modulo m in the last expression.

We conclude that the expectation in (3.13) contributes to the sum in (3.12) if and only if

$$\sum_{(p,r)\in A}(-1)^p(k_r^{(p)}-k_{r-1}^{(p)})=0\quad\text{for all } A\in\mathcal{A},\tag{3.14}$$

where these sums have to be evaluated modulo m in case of the discrete probability model. Since we require that $k_r^{(p)}\neq k_{r-1}^{(p)}$ for all $r=1,\dots,\nu+1$ and $p=1,\dots,2K$, see (3.11), the above condition cannot be satisfied if $\mathcal{A}$ has a block A containing only one element. Thus, we only need to consider partitions having at least 2 elements in each block, i.e., the ones in $P([2K]\times[\nu+1],t)$. We denote by $B(\mathcal{A},T)$ the number of vectors $(k_r^{(p)})$ in $I(T,K,\nu)$ satisfying (3.14). Also observe that if $\mathcal{A}$ has t blocks then the number of vectors of indices $(\ell_A)_{A\in\mathcal{A}}\in\{1,\dots,N\}^t$ such that all ℓ_A are different is $N(N-1)\cdots(N-t+1)=\frac{N!}{(N-t)!}$.

Pasting all these pieces together yields

$$\mathbb{E}\left[\|H_0^\nu v^{(j)}\|_2^{2K}\right]=\sum_{t=1}^{\min\{K(\nu+1),N\}}\frac{N!}{(N-t)!}\sum_{\mathcal{A}\in P([2K]\times[\nu+1],t)}B(\mathcal{A},T).$$

Let us estimate $B(\mathcal{A},T)$. Recall that we required $k_0^{(2p-1)}=k_0^{(2p)}$, $p=1,\dots,K$ for a vector $(k_r^{(p)})_{r=0,\dots,\nu}^{p=1,\dots,2K}\in I(K,\nu)$, see (3.11). Hence, it may be identified with a vector in $T^{2K(\nu+1)-K}$. If $\mathcal{A}$ has t blocks then $(k_r^{(p)})$ is restricted to the t independent (!) linear conditions given in (3.14). It follows that the number of vectors in $I(T,K,\nu)$ satisfying (3.14) can be bounded by $|T|^{2K\nu+K-t}$ – whether we calculate modulo m or not. Recalling that $|P([2K]\times[\nu+1],t)|=S_2(2K(\nu+1),t)$ we finally obtain

$$\begin{aligned}
\mathbb{E}\left[\|H_0^\nu v^{(j)}\|_2^{2K}\right]&\le\sum_{t=1}^{\min\{K(\nu+1),N\}}\frac{N!}{(N-t)!}S_2(2K(\nu+1),t)\,|T|^{2K\nu+K-t}\\
&\le|T|^{2K\nu+K}\sum_{t=1}^{K(\nu+1)}S_2(2K(\nu+1),t)\left(\frac{N}{|T|}\right)^t=|T|^{2K\nu+K}F_{2K(\nu+1)}\left(\frac{N}{|T|}\right)
\end{aligned}$$

with F_n as defined in (2.4). ■

Let $\theta = N/|T| = N/M$. Using the previous lemma and (3.10) we conclude that

$$\mathbb{P}\left(\sum_{\nu=0}^{n-1} N^{-\nu-1}\|H_0^\nu v^{(j)}\|_2 \geq a\right) \leq \beta^{-2n}\sum_{\nu=0}^{n-1}|T|^{2(\nu+1)K_{\nu+1}}N^{-2(\nu+1)K_{\nu+1}}F_{2(\nu+1)K_{\nu+1}}\left(\frac{N}{|T|}\right)$$
$$= \beta^{-2n}\sum_{\nu=1}^{n} G_{2\nu K_\nu}(\theta).$$

Finally, together with (3.8) and (3.9) we obtain

$$\mathbb{P}\left(\mathcal{F}_{TX} \text{ not injective or } \max_{j\in\Gamma\setminus T}\|\mathcal{F}_{TX}^\dagger\phi_j\|_1 \geq 1\right) \leq \kappa^{-2}MG_{2n}(\theta) + (D-M)\beta^{-2n}\sum_{\nu=1}^{n} G_{2\nu K_\nu}(\theta).$$

Recall that we imposed the conditions $\kappa < 1$ and $\sum_{\nu=1}^n \beta^{n/K_\nu} \leq (1+\sqrt{M}\frac{\kappa}{1-\kappa})^{-1}$. Thus, by Lemma 3.1 we completed the proof of Theorem 2.1.

3.2 Proof of Corollary 2.2

The proof is similar to the one of Corollary 2.2 in [23]. Of course, the main issue is a good choice of the parameters $n, K_1, \ldots, K_n, \kappa, \beta$ in Theorem 2.1. Dependent on $\theta = N/M$, large enough, we let

$$n = n(\theta) := \left\lfloor \frac{\beta^3\theta}{8} \right\rfloor \geq 1. \tag{3.15}$$

Choose K_ν, $\nu = 1, \ldots, n$, to be the nearest integer to n/ν. It was proven in [23, Section 3.5] that in this case

$$(D-M)\beta^{-2n}\sum_{\nu=1}^{n} G_{2\nu K_\nu}(\theta) \leq 1/4(D-M)\theta 2^{-2n/3}$$

and the right hand side is less than $\epsilon/2$ provided

$$\frac{2\ln(2)}{3}n - \ln(\theta) \geq \ln(D-M) + \ln(\epsilon^{-1}) - \ln(2). \tag{3.16}$$

Further, let us choose κ such that

$$\frac{\kappa}{1-\kappa} = \frac{1}{24}M^{-1/2}.$$

Then condition (2.7) requires that

$$\sum_{\nu=1}^{n}\beta^{n/K_\nu} \leq \left(1+\sqrt{M}\frac{\kappa}{1-\kappa}\right)^{-1} = \frac{24}{25} = 0.96.$$

A simple numerical test shows that the choice $\beta = 0.47$ is valid for all $n \in \mathbb{N}$ and (3.15) yields $n = n(\theta) \approx \lfloor 0.013\,\theta \rfloor$. Recalling that $\theta = N/M$ it follows from (3.16) that there exists a constant C_1 such that $(D-M)\beta^{-2n}\sum_{\nu=1}^n G_{2\nu K_\nu}(\theta) \leq \epsilon/2$ provided

$$N \geq C_1 M \ln(D/\epsilon).$$

Further, it was shown in [23] that $G_{2n}(\theta) \leq (3n/\theta)^{n-1}$ if (3.15) holds. Since $M \geq 1$ our choice of κ implies $\kappa \geq \frac{1}{25} M^{-1/2}$. Thus, we have

$$M\kappa^{-2} G_{2n}(\theta) \leq 25^2 M^2 \left(\frac{3\beta^3}{8}\right)^{n-1}.$$

This term is less than $\epsilon/2$ if

$$(n-1)\ln\left(\frac{8}{3\beta^3}\right) \geq \ln(2 \cdot 25^2) + 2\ln(M) + \ln(\epsilon^{-1}).$$

The valid choice $\beta = 0.47$ yields $\ln(8/(3\beta^3)) \approx 3.2459$. We conclude that there exists a constant C_2 such that $M\kappa^{-2} G_{2n}(\theta) \leq \epsilon/2$ if

$$N \;\geq\; C_2 M \ln(M/\epsilon).$$

Choosing $C := \max\{C_1, C_2\}$ and using that $M \leq D$ concludes the proof of Corollary 2.2.

3.3 Proof of Corollary 2.3

We denote by success(X_1) the event that reconstruction is successful with the model that the sampling points $x_1, \ldots, x_N$ – collected in $X_1 = \{x_1, \ldots, x_N\}$ – are chosen independently from the uniform distribution on $\frac{2\pi}{m}\mathbb{Z}_m^d$, and by success$(X_2)$ the corresponding event when the sampling set X_2 is chosen uniformly at random from all subsets of $\frac{2\pi}{m}\mathbb{Z}_m^d$ of size N. It holds

$$\mathbb{P}(\text{success}(X_2)) \;=\; \mathbb{P}(\text{success}(X_1) |\; |X_1| = N).$$

Also observe that the probability of success certainly decreases if the sampling set X_1 contains less sampling points than N since then we have less information about f that could be used for the reconstruction. A basic property of conditional probabilities yields

$$\begin{aligned}
\mathbb{P}(\text{success}(X_1)) \;&=\; \sum_{k=1}^{N} \mathbb{P}(\text{success}(X_1) |\; |X_1| = k)\, \mathbb{P}(|X_1| = k) \\
&\leq \sum_{k=1}^{N} \mathbb{P}(\text{success}(X_1) |\; |X_1| = N)\, \mathbb{P}(|X_1| = k) \\
&= \mathbb{P}(\text{success}(X_1) |\; |X_1| = N) = \mathbb{P}(\text{success}(X_2)).
\end{aligned}$$

Thus, the probability of reconstruction when choosing the sampling set X_2 uniformly at random among all subsets of size N is larger than the probability of success(X_1) estimated in Theorem 2.1 and Corollary 2.3.

3.4 Proof of Theorem 2.5

The proof relies on estimating $\mathbb{E}[|\langle \phi_j, \phi_k \rangle|^{2n}]$ for $j \neq k$. In (3.2) we already computed

$$\langle \phi_j, \phi_k \rangle \;=\; \sum_{\ell=1}^{N} e^{i(j-k)\cdot x_\ell} \;=:\; \sigma_{j-k}. \tag{3.17}$$

Setting $k' = j - k \neq 0$ we get

$$|\sigma_{k'}|^{2n} \;=\; |\langle \phi_j, \phi_k\rangle|^{2n} \;=\; \sum_{\ell_1,\dots,\ell_{2n}=1}^{N} \;\prod_{r=1}^{2n} e^{i(-1)^r k'\cdot x_{\ell_r}}$$

Using linearity of expectation we obtain

$$\mathbb{E}\left[|\sigma_{k'}|^{2n}\right] \;=\; \sum_{\ell_1,\dots,\ell_{2n}=1}^{N} \mathbb{E}\left[\prod_{r=1}^{2n} e^{i(-1)^r k'\cdot x_{\ell_r}}\right]. \tag{3.18}$$

Similarly as in the proof of Theorem 2.1 we associate a partition $\mathcal{A}$ of $\{1,\dots,2n\}$ to a given vector $(\ell_r)_{r=1,\dots,2n} \in \{1,\dots,N\}^{2n}$ of indices such that $r, r' \in A \in \mathcal{A}$ if and only if $\ell_r = \ell_{r'}$. Then we may unambiguously write ℓ_A instead of ℓ_r. Using the independence of the random variables x_{ℓ_A} we get

$$\mathbb{E}\left[\prod_{r=1}^{2n} e^{i(-1)^r k'\cdot x_{\ell_r}}\right] \;=\; \prod_{A\in\mathcal{A}} \mathbb{E}\left[\exp\left(ik'\cdot(\sum_{r\in A}(-1)^r x_{\ell_r})\right)\right] \;=\; \prod_{A\in\mathcal{A}} \delta\left(\sum_{r\in A}(-1)^r\right). \tag{3.19}$$

The expectation in the second term is computed similarly as in the proof of Theorem 2.1 using that $k' = j - k \neq 0$. If the x_ℓ have the uniform distribution on $\frac{2\pi}{m}\mathbb{Z}_m^p$ then the sum in the last term initially has to be calculated modulo m. However, since A has at most n even and n odd elements calculating modulo m can be suppressed due to the assumption $n < m$.

Thus, the term in (3.19) contributes to the sum in (3.18) if and only if all blocks A of the partition $\mathcal{A}$ contain the same number of even and odd numbers. Let us denote the set of all such partitions of $\{1,\dots,2n\}$ with t blocks by $R(n,t)$. Clearly, $R(n,t)$ is empty if $t > n$ since then there would be a block with only 1 element. Also observe that the number of vectors of indices $(\ell_A)_{A\in\mathcal{A}} \in \{1,\dots,N\}^t$, $A \in \mathcal{A} \in R(n,t)$, with pairwise different entries is precisely $N!/(N-t)!$. We obtain

$$\mathbb{E}[|\sigma_{k'}|^{2n}] \;=\; \sum_{t=1}^{\min\{n,N\}} \sum_{\mathcal{A}\in R(n,t)} \frac{N!}{(N-t)!} \;\leq\; \sum_{t=1}^{n} |R(n,t)| N^t. \tag{3.20}$$

Let us determine the number $|R(n,t)|$ of partitions in $R(n,t)$ by developing a recursion formula. To this end consider the partitions of $\{1,\dots,2n+2\}$ in $R(n+1,t)$ and let A be the block containing the element $2n+2$. The overall number of elements in A may be $2s$ for s ranging between 1 and $n+1-(t-1) = n-t+2$ (each of the other $(t-1)$ blocks must contain at least 2 elements, so necessarily $2s + 2(t-1) \leq 2(n+1)$). For each s we may choose $s-1$ even numbers out of $\{2,4,\dots,2n\}$ (in addition to $2n+2$) and s odd numbers from $\{1,3,\dots,2n+1\}$ to be contained in A. Then the remaining numbers $\{1,\dots,2n+2\} \setminus A$ are partitioned into $t-1$ sets. The set of these remaining partitions is isomorphic to $R(n+1-s, t-1)$. Thus, the overall numbers of partitions in $R(n+1,t)$ is given by

$$\begin{aligned} |R(n+1,t)| \;&=\; \sum_{s=1}^{n-t+2} \binom{n}{s-1}\binom{n+1}{s} |R(n+1-s,t-1)| \\ &=\; \sum_{\ell=t-1}^{n} \binom{n}{\ell}\binom{n+1}{\ell} |R(\ell,t-1)|. \end{aligned} \tag{3.21}$$

In the second equality we used the substitution $\ell = n+1-s$ and the fact that $\binom{n}{n-\ell} = \binom{n}{\ell}$.

By [29, formula (23)] the recursion formula (3.21) is the same as the one for the Stirling type numbers $S^1(n,k)$ defined by the generating function in (2.10). Further, $R(n,1)$ contains only the trivial partition consisting of one block containing all the elements, i.e., $|R(n,1)| = 1$. By differentiating (2.10) with respect to x and setting $x = 0$ we see that also $S^1(n,1) = 1$ for all n. This implies that $|R(n,k)| = S^1(n,k)$. (This is the announced combinatorial interpretation of the numbers $S^1(n,k)$ that seemingly has not yet been observed.) Altogether we obtain

$$\mathbb{E}\left[|\langle \phi_j, \phi_k\rangle|^{2n}\right] \leq \sum_{t=1}^{n} S^1(n,t) N^t.$$

Using the union bound and Markov's inequality we further conclude for the coherence parameter μ defined in (2.9) that

$$\begin{aligned}
\mathbb{P}((2M-1)\mu \geq 1) &= \mathbb{P}((2M-1)N^{-1} \max_{j\neq k} |\langle \phi_j, \phi_k\rangle| \geq 1) = \mathbb{P}((2M-1)N^{-1} \max_{k'\in\Gamma'} |\sigma_{k'}| \geq 1) \\
&\leq \sum_{k'\in\Gamma'} \mathbb{P}(|\sigma_{k'}|^{2n} \geq (N/(2M-1))^{2n}) \leq \sum_{k'\in\Gamma'} \mathbb{E}[|\sigma_{k'}|^{2n}] \left(\frac{2M-1}{N}\right)^{2n} \\
&\leq D' \left(\frac{2M-1}{N}\right)^{2n} \sum_{t=1}^{n} S^1(n,k) N^k.
\end{aligned}$$

This concludes the proof of Theorem 2.5.

4 Implementation and Numerical Experiments

Algorithm 1 contains two costly computations. Step 4 multiplies the adjoint measurement matrix $\mathcal{F}_X^*$ with the current residual vector r_s. When drawing the sampling set from the lattice $\frac{2\pi}{m}\mathbb{Z}_m^d$ or from the interval $[0, 2\pi]^d$, we use a zero padded fast Fourier transform (FFT) or the nonequispaced FFT, cf. [22, 17], respectively. In both cases, the total costs of this step in one iteration is $\mathcal{O}(D \log D)$. Note furthermore, that if the maximum in step 4 occurs at several indices the algorithm chooses one of them. Step 5 solves in each iteration a least squares problem

$$\|\mathcal{F}_{T_s X} d_s - \mathrm{f}\|_2 \xrightarrow{d_s} \min.$$

A straightforward implementation yields costs $\mathcal{O}(MN^2)$ per iteration. Speed up for this computation is obtained by the QR factorization of $\mathcal{F}_{T_s X}$ obtained from the factorization of $\mathcal{F}_{T_{s-1} X}$, cf. [3, pp. 132], or by the use of the iterative algorithm LSQR, cf. [21], reducing the costs for solving one least squares problem to $\mathcal{O}(N^2)$ or $\mathcal{O}(MN)$, respectively. The latter assertion is due to a uniformly bounded condition number of $\mathcal{F}_{T_s X}$ and thus, a constant number of iterations within LSQR. (We will report on theoretical results concerning the boundedness of the condition numbers of $\mathcal{F}_{TX}$ in a subsequent contribution.) Clearly, if OMP succeeds, the algorithm takes M outer iterations. Two reasonable choices for stopping criteria are a maximum number of iterations (assuming an upper bound on the sparsity M is known) or a residual tolerance ε (or a combination of both). In any case the algorithm will do no more than N iterations. Assuming M outer iterations, a reasonable sparsity $M = \mathcal{O}(\sqrt{D})$, and

$N = \mathcal{O}(M \log D)$ sampling points, see also Corollary 2.2, Algorithm 1 (using LSQR) has a total cost of $\mathcal{O}(D^{1.5} \log D)$ arithmetic operations.

For the reader's convenience, we also provide an efficient and reliable implementation of the presented algorithm for the univariate case in MatLab$^{\text{TM}}$. Following the common accepted concept of *reproducible research*, all numerical experiments are included in our publicly available toolbox [18]. The toolbox comes with a simple version of the nonequispaced fast Fourier transform (NFFT). All numerical results were obtained on a Intel Pentium M with 1.60GHz, 512MByte RAM running OpenSUSE Linux kernel 2.6.13-15-default and MatLab7.1.0.183 (R14) Service Pack 3.

All examples testing the Basis Pursuit principle use the optimization tools of CVX [15], L1MAGIC [25], or MOSEK [20], respectively. If the vector of Fourier coefficients is assumed to be real valued, the ℓ^1-minimization problem is reformulated as a linear program, whereas for complex valued coefficients the corresponding second order cone problem is set up. While CVX and MOSEK handle both problems, the constraints have to be stored explicitly taking $\mathcal{O}(DN)$ bytes of memory and causing the drawback of not being able to use fast matrix multiplication algorithms like FFT or NFFT. On the other hand, L1MAGIC includes the use of function handles to avoid this memory bottleneck and reduces the computation time from $\mathcal{O}(DN)$ to $\mathcal{O}(D \log D)$ when multiplying with the matrix $\mathcal{F}_X$. Unfortunately, the solver for equality constraint ℓ^1-minimization of this package supports only the reformulation as linear program, i.e., real valued Fourier coefficients.

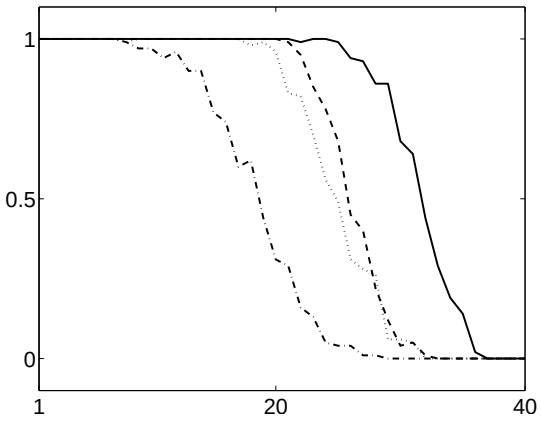
Subsequently, we compare our implementation of Algorithm 1 with different Basis Pursuit implementations for the univariate case. We use (pseudo-)random Fourier coefficients, where the real as well as the imaginary part is drawn from a normal distribution with mean zero and standard deviation one. The support $T \subset \Gamma$ of the Fourier coefficients is chosen uniformly at random from all the subsets of Γ of size M. The sampling points x_j are drawn uniformly from the interval $[0, 2\pi]$ for the continuous probability model, denoted by NFFT subsequently. Within the discrete probability model a subset of size N is chosen uniformly from all subsets of $\{0, \frac{2\pi}{D}, \dots, \frac{2\pi(D-1)}{D}\}$ with size N, denoted by FFT.

Example 4.1. *In our first example, we compare the ability of OMP and BP to reconstruct sparse trigonometric polynomials with complex valued coefficients on the set* $\Gamma = \{-25, \dots, 24\}$ *of possible indices, i.e., the dimension of* Π_Γ *is* $D = 50$. *We draw a set* T *of size* $M \in \{1, 2, \dots, 40\}$ *and* M *complex valued Fourier coefficients. Furthermore, we choose* $N = 40$ *sampling points within the discrete and within the continuous probability model. The samples* $(x_j, f(x_j))$, $j = 1, \dots, N$, *of the corresponding trigonometric polynomial and the set* Γ *are the input for the OMP and the BP algorithm. The OMP algorithm uses the updated QR factorization to solve the sequence of least squares problems, whereas the BP algorithm uses the MOSEK-package [20] to solve the second order cone problem. The output* $d_k \in \mathbb{C}$, $k \in \Gamma$, *of these algorithms is compared to the original vector of Fourier coefficients. Repeating the experiment* 100 *times for each level of sparsity* M, *we count how often each algorithm is able to reconstruct the given coefficients. Furthermore, the average CPU-time used by each algorithm with respect to the number of non-zero coefficients* M *is shown. The same experiment is done for the set* $\Gamma = \{-50, \dots, 49\}$.

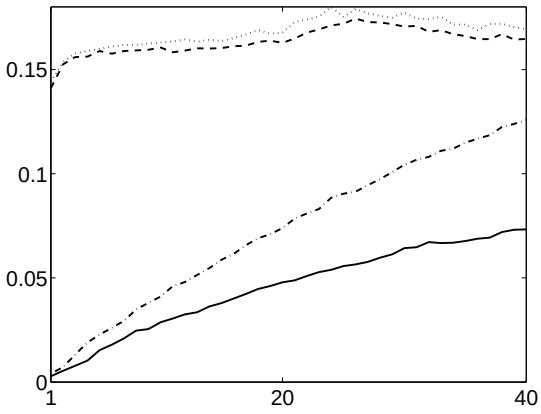
As readily can be seen from Figure 3, both OMP algorithms are faster than their BP counterparts. For the number of samples $N = 40$ *close to the dimension* $D = 50$, *the FFT-based OMP performs best, but both BP algorithms are able to reconstruct for sparsity levels when the NFFT-based OMP already fails. This effect is not present anymore and both OMP schemes*

perform at least as good as BP, if the dimension D is chosen larger.

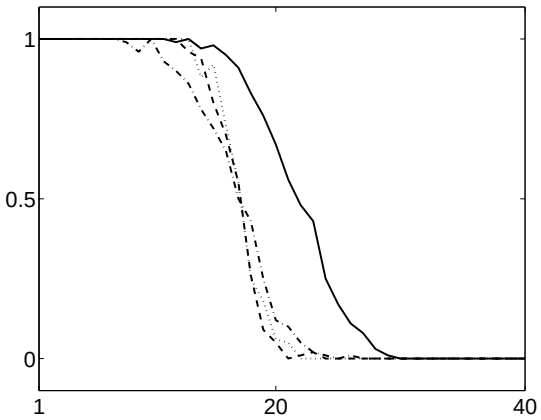
Also observe that the success rates are much better than the theoretical estimates provided in Theorem 2.1, see also Figure 2.



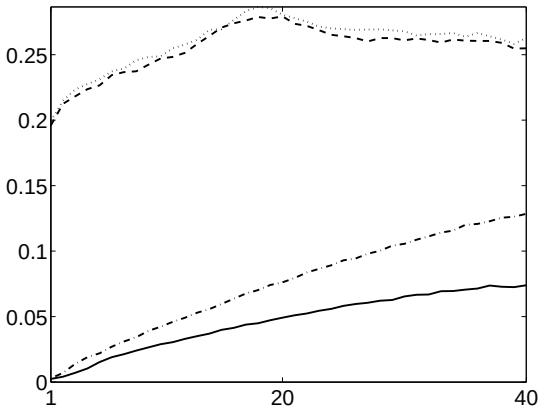
(a) Success rate, $D = 50$.



(b) Computation time $D = 50$.



(c) Success rate, $D = 100$.

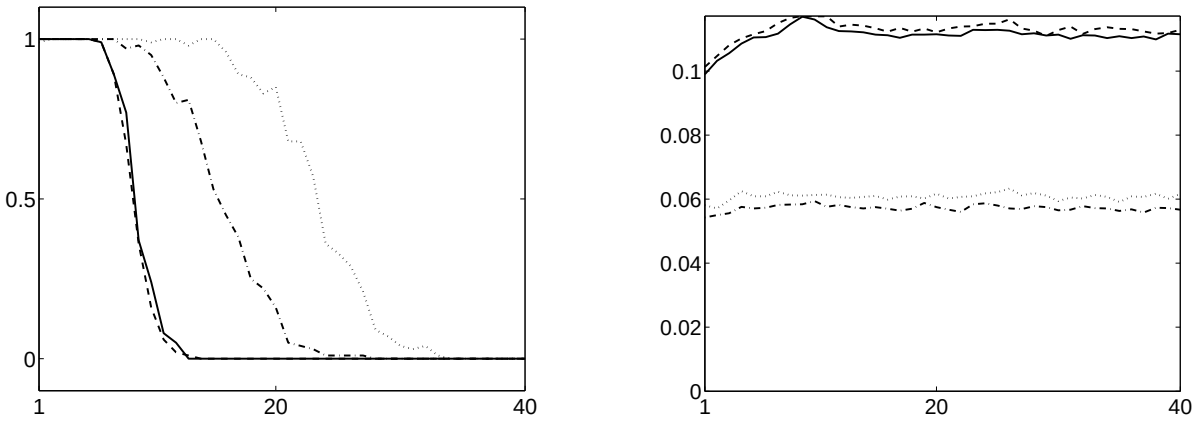


(d) Computation time $D = 100$.

Figure 3: Success rate of OMP, with FFT (solid) and NFFT (dash-dot), and BP, with FFT (dashed) and NFFT (dotted) with respect to an increasing number M of non-zero Fourier coefficients and a fixed number of samples $N = 40$ and dimension D of Π_Γ. Furthermore, we show the average computation time in seconds for each set of tested parameter.

Example 4.2. *We consider the Basis Pursuit principle exclusively. As in the previous example, the set of possible Fourier coefficients is* $\Gamma = \{-25, \dots, 24\}$*, whereas we use only* $N = 20$ *sampling points and choose real-valued or complex-valued Fourier coefficients, respectively. Again, we repeat the experiment* 100 *times for each level of sparsity, the results in Figure 4 reveal the following. Besides the easier implementation and a speed up of around two, the additional assumption, i.e., real valued coefficients, indeed saves roughly half of the needed samples to recover a sparse trigonometric polynomial.*

Example 4.3. *The last example that focuses on the success of reconstruction verifies Corollary 2.2 for OMP, i.e., the relation* $N \sim M \log(\frac{D}{\epsilon})$*, in the following way. For a fixed set* $\Gamma =$



(a) Success rate. (b) Computation time.

Figure 4: Basis pursuit from $N = 20$ samples, dimension $D = 50$. Success rate for complex-valued coefficients, with FFT (solid) and NFFT (dash-dot), and real-valued coefficients, with FFT (dashed) and NFFT (dotted) with respect to an increasing number M of non-zero coefficients. The chart on the right hand side shows again the average computation time.

$\{-2^9, \dots, 2^9 - 1\}$ *of possible indices* ($D = 1024$)*, we draw a support set* T *of increasing sizes* $M = 1, \dots, 40$*, complex-valued Fourier coefficients, and sampling sets of size* $N = 2.5M$*,* $N = 3M$*, and* $N = 3.5M$ *within the continuous probability model, respectively. For* 200 *runs of each experiment, we count the number of perfect reconstructions after exactly* M *steps. As Figure 5 (left) reveals, the success rate stays (almost) constant or might even increase slightly for an increasing number of non-vanishing coefficients if the ratio* $\theta = N/M$ *remains constant.*

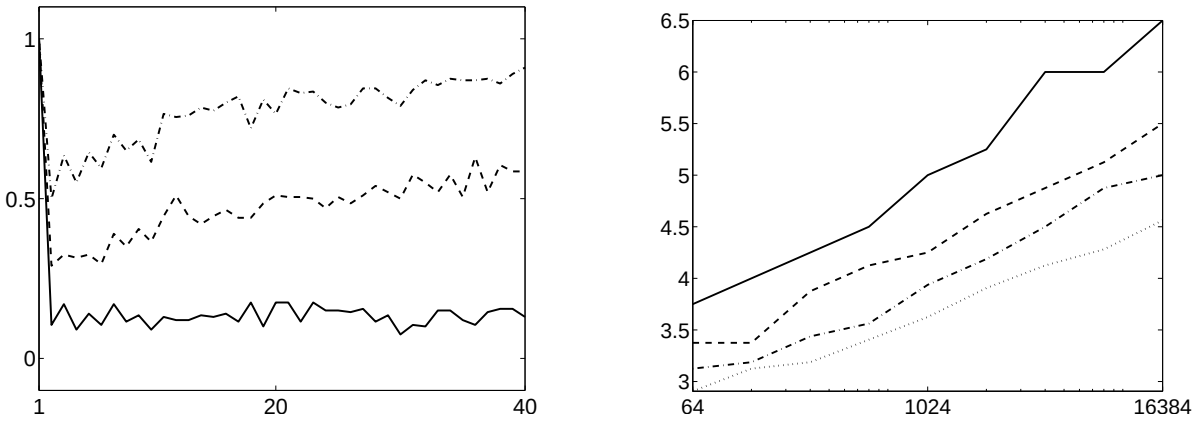
In the second part of this example, we are concerned with the dependence of this ratio $\theta = N/M$ *to reach a certain success rate when the dimension* D *varies. For an increasing number* $D = 2^6, 2^7, \dots, 2^{14}$ *of possible coefficients, we draw sets* T *of sizes* $M = 4, 8, 16, 32$ *and test for the* smallest *number* N *of (continuously drawn) samples, such that at least* 90% *(*180 *out of* 200*) of the runs result in a perfect recovery of the given Fourier coefficients. Figure 5 (right) confirms the relation* $\theta = C \log_2(D)$ *to reach a fixed success rate, whereas the constant* $C \leq \frac{2}{3}$ *even decreases mildly for a larger number* M *of non-zero coefficients.*

Example 4.4. *This example considers the computation time needed by each algorithm for an increasing dimension* D*, a dependent sparsity level* $M = \mathcal{O}(\sqrt{D})$*, and a number of samples* $N = \mathcal{O}(M \log D)$*. Constants are adjusted such that the used algorithms succeed in most cases in the reconstruction task; all methods, except* $L1MAGIC$*, are tested with complex coefficients.*

For the small scale experiment, we choose a dimension $D = 2^3, 2^4, \dots, 2^9$*, a sparsity level* $M = \lfloor \frac{1}{2}\sqrt{D} \rfloor$*, and a number of continuously drawn samples* $N = M(\log_2(D) - 2)$*. We use the OMP algorithm with the updated QR factorization and BP algorithms based on CVX, MOSEK, and L1MAGIC. The average computation time of* 10 *runs is shown in Figure 6 (left).*

Furthermore, we test for a dimension $D = 2^7, 2^8, \dots, 2^{17}$*, a sparsity level* $M = \lfloor \frac{1}{8}\sqrt{D} \rfloor$*, and a number of discrete drawn samples* $N = 2M \log_2 D$ *the following algorithms:*

1. *OMP with LSQR and explicitly stored matrices* $\mathcal{F}_{T_s X}$*,*
2. *OMP with LSQR and FFT-based multiplications with* $\mathcal{F}_{T_s X}$ *and its adjoint,*



(a) Success rate.

(b) 'Generalized oversampling factor' θ.

Figure 5: Numerical verification of Corollary 2.2 for OMP. Left: Rate of successful reconstruction with respect to number M of non-zero Fourier coefficients ($D = 2^{10}$) for three 'oversampling factors' $\theta = 2.5$ (solid), $\theta = 3$ (dashed), and $\theta = 3.5$ (dash-dot). Right: Ratio $\theta = N/M$, necessary to reach a success rate of 90 percent, with respect to the dimension D and fixed numbers of non-vanishing coefficients $M = 4$ (solid), $M = 8$ (dashed), $M = 16$ (dash-dot), and $M = 32$ (dotted).

3. BP using L1MAGIC and an explicitly stored matrix $\mathcal{F}_X$ for $D \leq 2^{12}$, and

4. BP using L1MAGIC and FFT-based multiplications with $\mathcal{F}_X$ and its adjoint.

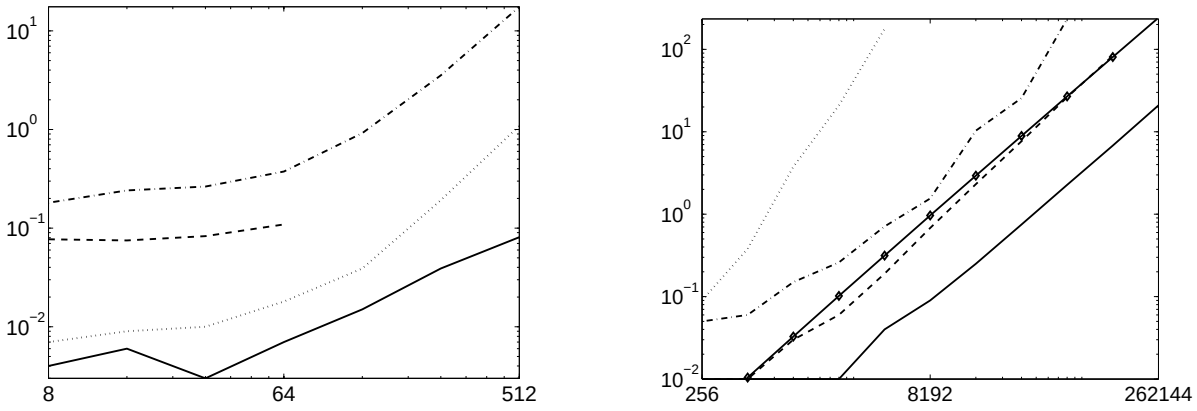
The FFT-based multiplications, denoted by implicit in Figure 6, need no additional memory, whereas storing the matrices explicitly needs $\mathcal{O}(D \log D)$ bytes for OMP and $\mathcal{O}(D^{1.5} \log D)$ bytes for BP.

Both OMP algorithms show a $\mathcal{O}(D^{1.5} \log D)$ time complexity, whereas the scheme with explicit storage of $\mathcal{F}_{T_s X}$ is a constant multiple faster. The Basis Pursuit algorithms are considerably slower in all cases. Moreover, the storage of the whole measurement matrix $\mathcal{F}_X$ results in large memory requirements.

5 Conclusions and future work

Our theoretical and numerical results show that both BP and OMP are well-suited for the problem of recovering sparse trigonometric polynomials from few random samples taken either on a grid (FFT) or from a continuous uniform distribution on the cube (NFFT). In practice however, OMP outperforms BP. Indeed, OMP is not only significantly faster and much easier to implement than BP, the numerical experiments show even a slightly higher rate of success of reconstruction. At first sight the latter observation might be surprising since OMP is a greedy algorithm which optimizes only locally at each step while BP is a global optimization scheme. It might be interesting (but also difficult) to find a mathematical explanation for this phenomenon.

We remark that both OMP and BP can be used to identify dominant frequencies from few samples – even if these frequencies are very high or if two (high) neighboring frequencies



(a) Small scale.

(b) Medium and large scale.

Figure 6: Computation time in seconds. Left: OMP with updated QR factorization (solid), BP using MOSEK (dashed), BP using CVX (dash-dot), BP using L1MAGIC(explicit) (dotted). Right: OMP with explicit LSQR (solid), OMP with implicit LSQR (dashed), BP using explicit L1MAGIC (dash-dot), BP using implicit L1MAGIC (dotted), and for comparison $\mathcal{O}(D^{1.5} \log D)$ (solid+diamond).

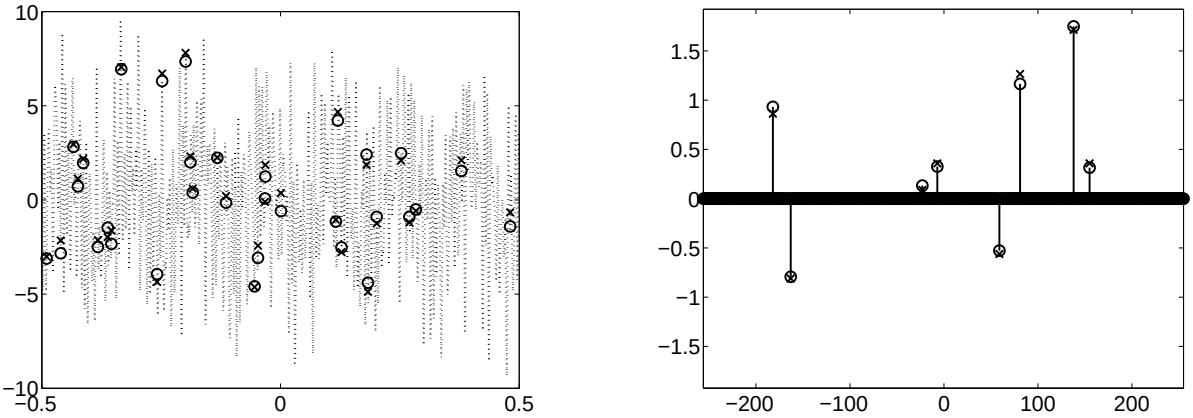
are present. Numerical tests showed for instance that the FFT based OMP can recover a signal consisting of 10 frequencies in $\{-2^{18}, \ldots, 2^{18} - 1\}$ – with $k_1 = 2^{18} - 1 = 262143$, $k_2 = 2^{18} - 2 = 262142$ being two of them – from 60 random samples.

In practice, signals are usually not sparse in a strict sense. However, they might still be well-approximated by sparse ones. For the case of BP in connection with FFT, Candes and Tao could prove in [7] that recovery is still possible with only small errors. Indeed, this is implied by the so called Uniform Uncertainty Principle for the Fourier Basis, see also the notion of restricted isometry constants in [8]. We plan to investigate such stability issues also for OMP, and for the NFFT case.

Further, stability under noise is important in practice. For BP in connection with the FFT this has been establish already by Candes, Romberg and Tao in [8]. For OMP we performed first numerical experiments in the following way. We corrupted the sample values with a significant amount of normal distributed noise. We observed that OMP usually finds the correct support set and makes only small errors on the coefficients, see Figure 7. Thus, it seems that also OMP is stable under noise – at least if the noise level is not very high. (Actually, we observed that for moderately higher noise as in our example in Figure 7, the reconstructed coefficient vector is significantly different from the original). We plan to address such issues more deeply in a future contribution.

Acknowledgments: The first author is grateful for partial support of this work by the German Academic Exchange Service (DAAD) and the warm hospitality during his stay at the Numerical Harmonic Analysis Group, University of Vienna. The second author is supported by an Individual Marie Curie Fellowship funded by the European Union under contract MEIF-CT-2006-022811. He also acknowledges funding by the European Union's Human Potential Programme, under contract HPRN–CT–2002–00285 (HASSIP).

We would like to thank Emmanuel Candes, Justin Romberg, Ingrid Daubechies and Jared Tanner for interesting discussions on the subject.



(a) Trig. polynomial and (noisy) samples.

(b) True and recovered coefficients.

Figure 7: Trigonometric polynomial (real part) as in Figure 1 and 30 samples (∘). The samples are disturbed by Gaussian distributed noise with variance 0.2 (×) (resulting here in a PSNR of 15.6dB). Nevertheless OMP reconstructs the true support set of the coefficients, and the true coefficients (∘) themselves with small error (×).

References

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

[2] A. Böttcher, D. Potts, Probability against condition number and sampling of multivariate trigonometric random polynomials, Preprint, 2006.

[3] Å. Björck, *Numerical Methods for Least Squares Problems*, SIAM, 1996.

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

[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] Compressed Sensing Resources, web-page, `http://www.dsp.ece.rice.edu/cs`.

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

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

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

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

[15] M. Grant, S. Boyd, Y. Ye, CVX: Matlab Software for Disciplined Convex Programming, Version 1.0 beta, `http://www.stanford.edu/~boyd/cvx`, 2006.

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

[17] S. Kunis, D. Potts, NFFT Softwarepackage, C subroutine library, `http://www.math.uni-luebeck.de/potts/nfft`, 2002 – 2006.

[18] S. Kunis, H. Rauhut, OMP-NFFT, MatLab-toolbox for orthogonal matching pursuit on sparse trig. polynomials, `http://www.tu-chemnitz.de/~skunis/software.php`, 2006.

[19] S. Mallat, Z. Zhang, Matching pursuit with time-frequency dictionaries, IEEE Trans. Sig. Proc., 41: 3397:3415, 1993

[20] mosek ApS, MOSEK optimization software, `http://www.mosek.com`.

[21] C. Paige, M. Saunders, LSQR: An algorithm for sparse linear equations and sparse least squares, ACM Trans. on Mathematical Software 8, 43 – 71, 1982.

[22] 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.

[23] H. Rauhut, Random Sampling of Sparse Trigonometric Polynomials, Preprint 2005.

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

[25] J. Romberg, E. Candes, L1MAGIC, MatLab toolbox for basis pursuit, `http://www.acm.caltech.edu/l1magic`, 2006.

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

[27] M. Rudelson, R. Vershynin, Sparse reconstruction by convex relaxation: Fourier and Gaussian measurements, Preprint, arXiv:math.NA/0602559, 2006.

[28] SeDuMi, optimization software, `http://sedumi.mcmaster.ca/`.

[29] J.-M. Sixdeniers, K.A. Penson, A.I. Salomon, Extended Bell and Stirling numbers from hypergeometric exponentiation, J. Integ. Seq. 4, Article 01.1.4, 2001.

[30] N.J.A. Sloane, On-Line Encyclopedia of Integer Sequences, published electronically at: `http://www.research.att.com/~njas/sequences`.

[31] J. Tropp, Greed is Good: Algorithmic Results for Sparse Approximation, IEEE Trans. Inform. Theory 50, 2231–2242, 2004.

[32] J. Tropp, Topics in Sparse Approximation, Ph.D. Thesis, University of Texas at Austin, 2004.

[33] 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.

Stefan Kunis
Fakultät für Mathematik
Technische Universität Chemnitz
Reichenhainer Str. 39
09107 Chemnitz, Germany
kunis@mathematik.tu-chemnitz.de

Holger Rauhut
NuHAG, Faculty of Mathematics
University of Vienna
Nordbergstr. 15
A-1090 Wien, Austria
holger.rauhut@univie.ac.at
WWW: `http://homepage.univie.ac.at/holger.rauhut`