Stability Results for Random Sampling of Sparse Trigonometric Polynomials

Holger Rauhut

September 18, 2006; revised April 4, 2008

Abstract

Recently, it has been observed that a sparse trigonometric polynomial, i.e. having only a small number of non-zero coefficients, can be reconstructed exactly from a small number of random samples using Basis Pursuit (BP) or Orthogonal Matching Pursuit (OMP). In the present article it is shown that recovery by a BP variant is stable under perturbation of the samples values by noise. A similar partial result for OMP is provided. For BP in addition, the stability result is extended to (non-sparse) trigonometric polynomials that can be well-approximated by sparse ones. The theoretical findings are illustrated by numerical experiments.

Key Words: random sampling, trigonometric polynomials, Orthogonal Matching Pursuit, Basis Pursuit, compressed sensing, stability under noise, fast Fourier transform, non-equispaced fast Fourier transform
AMS Subject classification: 94A20, 42A05, 15A52, 90C25

1 Introduction

Over the recent years compressed sensing has become a rapidly developing research field, see e.g. [1, 4, 8, 10, 30, 34]. In their seminal papers [4, 5, 6] Candès, Romberg and Tao observed that it is possible to recover sparse vectors, i.e., having only few non-vanishing coefficients, from a number of measurements that is small compared to the ambient dimension of the vector. As reconstruction method they promoted ℓ_1-minimization, also refered to as Basis Pursuit (BP) [7]. Their results apply in particular to recovery of a sparse vector from (random) samples of its discrete Fourier transform. In [28] the author extended their result to the situation where samples of the corresponding trigonometric polynomial are taken at random from the uniform (continuous) distribution on the cube, i.e., the samples are chosen "off the grid".

Another line of research suggests Orthogonal Matching Pursuit (OMP) as recovery method [15, 21, 33]. This is a greedy algorithm which is significantly faster than BP in practice. Partial results in [21] indicate that also OMP is able to recover a sparse trigonometric polynomial from few random samples. Moreover, numerical experiments suggest that OMP usually has a slightly higher probability of recovery success than BP - although BP has some theoretical advantages.

In practice, it is important that recovery methods are stable in the presence of noise on the measurements. Candès et al. showed in [5] that (a variant of) BP is indeed stable under a certain condition on the measurement matrix involving the so called restricted isometry constants. An estimation of these constants for the measurement matrix corresponding to random samples of the discrete Fourier transform was provided in [6] and [30]. In the present article we extend this estimate to the case of random samples at uniformly distributed points on the cube $[0, 2\pi]^d$.

We further provide partial results indicating that also OMP is stable under perturbation of the measurements by noise. Finally, numerical experiments reveal that the average reconstruction error of OMP is usually smaller than for (the variant of) BP in the presence of noise.

After the first submission of this manuscript, variants of OMP – Regularized Orthogonal Matching Pursuit (ROMP) [26, 25] and CoSaMP [24] – were introduced, that achieve similar theoretical recovery and stability guarantees as Basis Pursuit and are even slightly faster than OMP. Since the analysis of these algorithms is based on the restricted isometry constants our estimates for the Fourier type measurement matrix are useful for the analysis of ROMP and CoSaMP as well.

The paper is organized as follows. Section 2 gives some background on prior work, introduces notation and describes our problem. In Section 3 we present our main results concerning stability of a variant of BP, while Section 4 states stability theorems for OMP. Section 5 presents the proofs for BP, and Section 6 deals with the ones for OMP. The numerical experiments are detailed in Section 7. Finally, we conclude in Section 8 with a discussion.

2 Prior Work and Problem Statement

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

We will mainly deal with "sparse" trigonometric polynomials, i.e., we assume that the sequence of coefficients c_k is supported only on a small set $T \subset \Gamma$. However, a priori nothing is known about T apart from a maximum size. Thus, it is useful to introduce the (nonlinear) set $\Pi_\Gamma(M) \subset \Pi_\Gamma$ of all trigonometric polynomials whose Fourier coefficients are supported on a set $T \subset \Gamma$ satisfying $|T| \leq M$, $\Pi_\Gamma(M) = \bigcup_{T\subset\Gamma, |T|\leq M} \Pi_T$.

Our aim is to reconstruct an element $f \in \Pi_\Gamma(M)$ from sample values $f(x_1), \ldots, f(x_N)$, where the number N of sampling points $x_1, \ldots, x_N \in [0, 2\pi]^d$ is small compared to the dimension D (but, of course, larger than the sparsity M). As suggested by [4, 6, 15, 21, 28] we will study the behaviour of two reconstruction methods: Basis Pursuit (BP) and Orthogonal Matching Pursuit (OMP).

BP was much promoted by Donoho and his coworkers, see e.g. [7, 13]. It consists in solving the following ℓ^1-minimization problem

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

This task can be performed with convex optimization techniques [2]. Recently, much effort has been dedicated to the development of fast algorithms specialized to ℓ_1-minimization, see e.g. [9, 14, 18].

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 and/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 $\|r_s\| \leq \varepsilon$
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.

OMP is a greedy algorithm [23, 33], which selects a new element of the support set T in each step, see Algorithm 1. Its precise formulation uses the following notation. Let $X = (x_1, \ldots, x_N)$ be the sequence of sampling points. We denote by $\mathcal{F}_X$ the $N \times D$ matrix with entries

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

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., $\phi_k = \left(e^{ik\cdot x_\ell}\right)_{\ell=1}^N$. The restriction of $\mathcal{F}_X$ to the columns indexed by T is denoted by $\mathcal{F}_{TX}$. 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. For details on the implementation of OMP we refer to [21]. We only note that the fast Fourier transform (FFT) or the non-equispaced fast Fourier transform (NFFT), see e.g. [27] and the references therein, can be used for speed-ups of OMP.

Since it seems to be very hard to come up with deterministic recovery results we model the sampling points $x_1, \ldots, x_N$ as random variables. To this end we use two probability models.

(1) The sampling points $x_1, \ldots, x_N$ are independent random variables having the uniform distribution on the cube $[0, 2\pi]^d$.

(2) The sampling points $x_1, \ldots, x_N$ are independent random variables having the uniform distribution on the grid $\{0, \frac{2\pi}{q}, \ldots, 2\pi\frac{q-1}{q}\}^d$. Here, it is implicitly assumed that $\Gamma \subseteq \{0, 1, \ldots, q-1\}^d$.

Model (1) will also be refered to as the continuous model, while the second will be called "discrete". Observe that with model (2) it might happen with non-zero probability that some sampling points are selected more than once. To overcome this problem one might also choose the sampling set uniformly at random among all subsets of the grid $\{0, \frac{2\pi}{q}, \ldots, 2\pi\frac{q-1}{q}\}^d$ of size N. This model was actually used in [4, 6, 30]. However, for technical reasons we work with the model (2) here. Intuitively, moving from model (2) to its variant should actually improve the situation since always a maximum of information is used.

In [28] it was proven that BP is able to recover a sparse trigonometric polynomial from a rather small number of sample values.

Theorem 2.1. *Let $T \subset \Gamma$ with $|T| \leq M$. Choose $x_1, \ldots, x_N$ be random variables according to the probability models (1) or (2). Assume that*

$$N \geq CM \log(D/\epsilon). \tag{2.3}$$

Then with probability at least $1 - \epsilon$ both BP and OMP recover exactly all $f \in \Pi_\Gamma(M)$ with coefficients supported on T from the sample values $f(x_j), j = 1, \ldots, N$. The constant C is absolute.

The above theorem is non-uniform in the sense that for a single sampling set X recovery is guaranteed only for the given support set T (but for all Fourier coefficients supported on T). By Theorem 3.2 to be shown later it follows that this drawback can be removed, i.e., recovery by BP can be made fully uniform by introducing additional log factors to condition (2.3).

Recovery by OMP was studied theoretically and numerically in [21], although the theoretical results are only partial so far. At least the first step of OMP could be analyzed:

Theorem 2.2. *Let $f \in \Pi_\Gamma(M)$ with coefficients supported on T. Choose random sampling points $x_1, \ldots, x_N$ according to one of our two probability models. If*

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

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

The numerical experiments conducted in [21] suggest that also the further steps of OMP select elements of the true support T, so that after M steps the correct polynomial f is recovered. However, starting with the second step the theoretical analysis seems to be quite difficult due to subtle stochastic dependency issues.

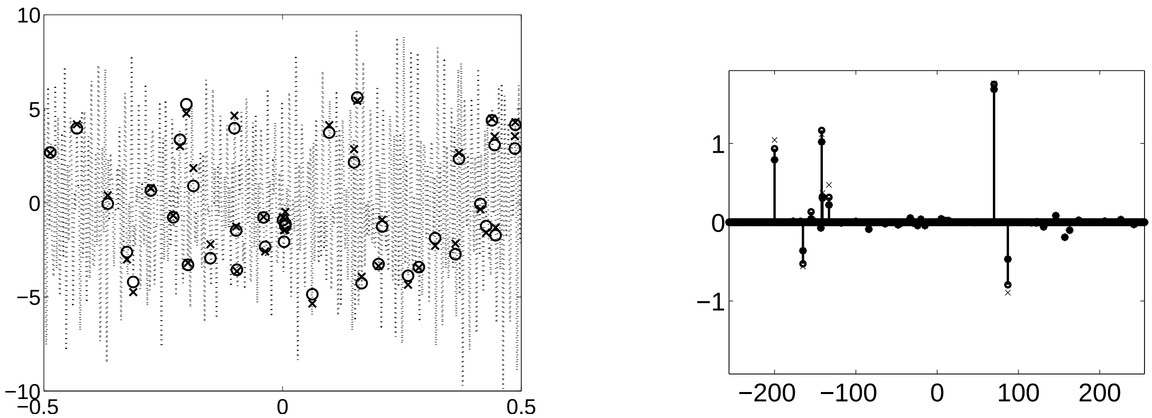
We note that the above theorem is non-uniform in the sense that the success probability is valid for the given polynomial, but it does not state that with high probability a single sampling set $\{x_1, \ldots, x_N\}$ is good for *all* sparse trigonometric polynomials. Such a uniform result was also provided in [21], which actually analyzes the full application of OMP, but requires significantly more samples.

Theorem 2.3. *Let $X = (x_1, \ldots, x_N)$ be chosen according to the continuous probability model (1) or the discrete model (2). Suppose that*

$$N \geq C(2M - 1)^2 \log(4D'/\epsilon), \tag{2.4}$$

where $D' := \#\{j - k : j, k \in \Gamma, j \neq k\} \leq D^2$. Then with probability at least $1 - \epsilon$ OMP recovers every $f \in \Pi_\Gamma(M)$. The constant satisfies $C \leq 4 + \frac{4}{3\sqrt{2}} \approx 4.94$. In case of the continuous probability model it can be improved to $C = 4/3$.

The above result is based on analysis of the coherence, see also below. It seems that condition (2.4) is actually optimal up to perhaps the constant C and the log-factor if one requires uniformity, i.e., recovery of *all* sparse trigonometric polynomials in $\Pi_\Gamma(M)$ from a single sampling set X, see [29]. In this regard, BP and OMP seem to be crucially different.



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

(b) True and recovered coefficients.

Figure 1: Left: Trigonometric polynomial (real part) of sparsity $M = 8$ and $N = 40$ samples ($\circ$). The samples are disturbed by noise η with $\|\eta\|_2 = 4$ ($\times$). Right: True coefficients ($\circ$), reconstruction by BP variant (3.1) ($*$), reconstruction by OMP ($\times$).

BP can give a uniform guarantee if the number of samples N scales linearly in the sparsity M (ignoring log-factors), see Theorem 2.1, Theorem 3.2 below and e.g. [6, 30], while OMP can give at most a non-uniform guarantee in this range, compare also [11, Section 7] and [15].

In this article we treat the question whether recovery by BP and OMP is stable if the sample values $f(x_j)$ are perturbed by noise. Additionally, for BP we consider also the case that f is not sparse in a strict sense, but can be well approximated by a sparse trigonometric polynomial.

In mathematical terms we assume that we observe the vector

$$y = (f(x_j))_{j=1}^N + \eta = \mathcal{F}_X c + \eta,$$

rather than $(f(x_j)) = \mathcal{F}_X c$, where the noise η satisfies $\|\eta\|_2 = (\sum_{\ell=1}^N \eta_\ell^2)^{1/2} \leq \sigma$ for some $\sigma \geq 0$. We will investigate whether the difference between the original coefficient vector and the one reconstructed by OMP or BP is small. For OMP we additionally ask whether the correct support set is recovered. Figure 1 provides a first illustration by showing an example of a reconstruction by the BP variant (3.1) and OMP from noisy samples.

In the sequel, $\|\cdot\|_{p\to q}$ will denote the operator norm from the sequence space ℓ^p into ℓ^q (on some index set), $\lfloor x \rfloor$ is the largest integer smaller or equal to x. Furthermore, C will always denote a generic constant, whose value might be different in each occurence.

3 Basis Pursuit

In the presence of noise it is useful to consider a slight variant of Basis Pursuit. Indeed, in [5] it is suggested to minimize the ℓ_1-norm of the coefficient vector c subject to the constraint that the residual error satisfies $\|\mathcal{F}_X c - y\|_2 \leq \sigma$, i.e., we solve

$$\min \|c\|_1 \quad \text{subject to } \|\mathcal{F}_X c - y\|_2 \leq \sigma. \tag{3.1}$$

Again this problem can be solved by convex optimization techniques [2]. Clearly, if $\sigma = 0$ then we are back to the original Basis Pursuit principle (2.1).

For the problem (3.1) quite general stability results were obtained by Candes, Romberg and Tao in [5], see also [8]. Their key concept is the following definition.

Definition 3.1. *The restricted isometry constant δ_M of a matrix A is the smallest number such that for all subsets T with $|T| \leq M$ it holds*

$$(1 - \delta_M)\|x\|_2^2 \;\leq\; \|A_T x\|_2^2 \;\leq\; (1 + \delta_M)\|x\|_2^2 \tag{3.2}$$

for all coefficients x supported on T. Here A_T denotes the restriction of A to the columns indexed by T.

In [5] the following theorem was proved. (Although it was originally stated only for the real-valued case the theorem together with its proof also holds for the complex-valued case.)

Theorem 3.1. *Assume that A is some matrix for which the restricted isometry constants satisfy*

$$\delta_{3M} + 3\delta_{4M} < 2.$$

Let $x \in \mathbb{C}^D$ and assume we have given noisy data $y = Ax + \eta$ with $\|\eta\|_2 \leq \sigma$. Denote by x_M the truncated vector corresponding to the M largest absolute values of x. Then the solution $x^\#$ to the problem

$$\min \|x\|_1 \text{ subject to } \|Ax - y\|_2 \leq \sigma$$

satisfies

$$\|x^\# - x\|_2 \;\leq\; C_1\sigma + C_2\frac{\|x - x_M\|_1}{\sqrt{M}}. \tag{3.3}$$

The constants C_1 and C_2 depend only on δ_{3M} and δ_{4M}.

Thus, recovery by the BP variant (3.1) is stable provided the restricted isometry constants are small. Note that the second term in (3.3) vanishes if x is sparse, i.e., has not more than M non-vanishing coefficients.

For our case this means that it is sufficient to provide conditions that ensure $\delta_{4M} \leq \delta$ for some small δ with high probability. (Note that for $\delta = 1/5$ the constants in the previous theorem are actually quite well-behaved, $C_1 \leq 12.04$ and $C_2 \leq 8.77$, see [5].)

Candès and Tao [6] provided such conditions for the discrete Fourier transform with a slightly different probability model than our discrete model (2). More recently, Rudelson and Vershynin came up with a more elegant and shorter solution to this problem [30]. It is possible to apply their technique also to our two probability models, notably the continuous one. This gives the following result.

Theorem 3.2. *Let $D = |\Gamma|$ and a sparsity M be given. Let $\epsilon, \delta \in (0, 1)$ and assume*

$$\frac{N}{\log(N)} \;\geq\; C\delta^{-2}M\log^2(M)\log(D)\log(\epsilon^{-1}). \tag{3.4}$$

Let the N sampling points $X = (x_1, \ldots, x_N)$ be chosen at random according to the model (1) or (2). Then with probability at least $1 - \epsilon$ the isometry constant of the matrix $N^{-1/2}\mathcal{F}_X$ satisfies

$$\delta_M \leq \delta. \tag{3.5}$$

The constant C is absolute.

The combination of Theorems 3.1 and 3.2 gives the following.

Corollary 3.3. *Let Γ with $|\Gamma| = D$, M, N and ϵ such that*

$$\frac{N}{\log(N)} \geq C_0 M \log^2(M) \log(D) \log(\epsilon^{-1}). \tag{3.6}$$

Choose $x_1, \ldots, x_N$ according to the probability model (1) or (2). Then with probability at least $1-\epsilon$ the following holds for all coefficient vectors $c \in \mathbb{C}^\Gamma$. Assume $y = \mathcal{F}_X c + \eta$ with $\|\eta\|_2 \leq \sigma$. Denote by c_M the truncated vector corresponding to the largest coefficients of c. Then the solution $c^\#$ to the minimization problem (3.1) satisfies

$$\|c^\# - c\|_2 \leq C_1 \frac{\sigma}{\sqrt{N}} + C_2 \frac{\|c - c_M\|_1}{\sqrt{M}}. \tag{3.7}$$

Remark 3.1.

(a) Choosing $\sigma = 0$ yields uniform exact recovery. Under condition (3.4) BP is able to reconstruct exactly *all* $f \in \Pi_\Gamma(M)$ from a single sampling set X.

(b) Note that condition (3.4) is satisfied if $N \geq C\delta^{-2} M \log^4(D) \log(\epsilon^{-1})$. Furthermore, (3.6) is probably not optimal. One may conjecture that $N = \mathcal{O}(M \log(D/\epsilon))$ or even $N = \mathcal{O}(M \log(D/(M\epsilon)))$ samples are enough, see also [30].

(c) With a discrete probability model (the variant of (2) outlined in Secion 2), Candès and Tao originally obtained a version of Theorem 3.2 (see [6, Lemma 4.3]) where for some parameter α and constant ρ the statement $\delta_M \leq c_0$ holds with probability at least $1 - CD^{-\rho/\alpha}$ under the condition $N \geq \alpha^{-1} M \log(D)^6$. Substituting $\epsilon = CD^{-\rho/\alpha}$ and solving for α yields the condition

$$N \geq C' M \log(D)^5 \log(\epsilon^{-1}). \tag{3.8}$$

It might be possible to adapt the original proof of Candès and Tao also to the continuous probability model (1) although this does not seem straightforward.

4 Orthogonal Matching Pursuit

In this section we consider the stability of OMP. Since we measure only noisy samples we cannot expect to have perfect recovery of a sparse signal, but at least we would like to obtain the true support of the sparse coefficient vector and only small deviations of their entries. We first provide the analogue of Theorem 2.2 for the noisy case. Unfortunately, we again have to restrict to the first iteration because it is still not clear how to deal with the subtle stochastic dependency issues arising in the analysis of the further iterations.

Theorem 4.1. *Let $f \in \Pi_\Gamma(M)$ with Fourier coefficients c. Let $N \in \mathbb{N}$ and $\tau, \epsilon \in (0,1)$ such that*

$$N \geq CM\tau^{-2} \log(D/\epsilon). \tag{4.1}$$

Further, let $\sigma > 0$ such that

$$\sigma \leq \frac{1-\tau}{4} \sqrt{\frac{N}{M}} \|c\|_2. \tag{4.2}$$

Choose the random sampling set $X = (x_1, \ldots, x_N)$ according to the probability model (1) or (2). Assume that we have given noisy samples $y = (f(x_\ell))_{\ell=1}^N + \eta = \mathcal{F}_X c + \eta$ with $\|\eta\|_2 \leq \sigma$. Then with probability exceeding $1-\epsilon$ OMP selects an element of the true support of c in the first step.
If after M steps OMP actually recovers the complete support of c then with probability exceeding $1-\epsilon$ the reconstructed coefficients $\tilde{c}$ satisfy

$$\|c - \tilde{c}\|_2 \leq \sqrt{\frac{2}{N}}\,\sigma. \tag{4.3}$$

From the proof of this Theorem one can deduce more precise information about the constant in condition (4.1). Indeed, N has to satisfy the two conditions

$$N \geq 17.88\, M\tau^{-2}\log(8D/\epsilon) \quad \text{and} \quad \left\lfloor \frac{N}{12eM} \right\rfloor \geq \ln(2(1-1/(4e))^{-1}M/\epsilon).$$

Note that $\|c\|_2 \geq \sqrt{M}\min_{j\in T}|c_j|$. Hence, condition (4.2) is satisfied if

$$\sigma \leq \frac{1-\tau}{4}\sqrt{N}\min_{j\in T}|c_j|. \tag{4.4}$$

One expects that this condition (with possibly a different constant) is sufficient that OMP selects an element of the true support T in every step. Hence, the noise level should not exceed the minimal absolute non-zero coefficient in order to have recovery of the correct support.

We note that our numerical experiments in Section 7 indicate that under condition (4.1) OMP actually selects elements of the true support T also in the further iterations and then (4.3) holds. However, we have not yet been able to carry through the corresponding theoretical analysis.

4.1 A uniform result

The result in the previous section is non-uniform. Let us state also a uniform recovery result for OMP extending Theorem 2.3 to the noisy situation.

Theorem 4.2. *Let the random sampling set $X = (x_1, \ldots, x_N)$ be chosen according to one of our probability models. Let $\tau, \epsilon \in (0,1)$ and $\sigma > 0$. Assume that*

$$N \geq C\tau^2(2M-1)^2\ln(4D'/\epsilon), \tag{4.5}$$

where $D' = \#\{j-k : j,k \in \Gamma, j \neq k\} \leq D^2$. Then with probability $1-\epsilon$ the following holds for all $f \in \Pi_\Gamma(M)$ whose Fourier coefficients satisfy

$$\min_{k\in\operatorname{supp} c}|c_k| > \frac{2\sigma}{(1-\tau)\sqrt{N}}. \tag{4.6}$$

If OMP is applied on the noisy samples $y = \mathcal{F}_X c + \eta$ with $\|\eta\|_2 \leq \sigma$, and stopped once the residual satisfies $\|r_s\| \leq \sigma$ then the true support of c is recovered and the reconstructed coefficient vector $\tilde{c}$ satisfies

$$\|c - \tilde{c}\|_2 \leq \frac{1}{\sqrt{N(1-\tau/2)}}\,\sigma.$$

The above result has the drawback that the number of samples required by (4.5) scales quadratically in the sparsity M rather than linearly as in (4.1). As in the noiseless case however, one cannot expect to come around the quadratic scaling if one requires uniformity, i.e., recovery by OMP of *all* $f \in \Pi_\Gamma(M)$ from a single sampling set X. Up to perhaps the log-factor condition (4.5) seems then to be optimal, see [29].

In contrast, BP gives a uniform guarantee if the number of samples is only linear in the sparsity up to some log-factors, see Theorem 3.2. Thus, under this requirement, BP seems to be the method of choice. However, for certain applications it might be enough to have a non-uniform guarantee and then OMP is a good alternative considering that it is usually significantly faster and much easier to implement, see also Section 7.

5 Proof of Theorem 3.2

We mainly follow the ideas in [30]. Condition (3.2) for $N^{-1/2}\mathcal{F}_X$ is equivalent to

$$\sup_{|T|\leq M} \|I_T - N^{-1}\mathcal{F}_{TX}^*\mathcal{F}_{TX}\|_{2\to 2} \;=\; \delta_M,$$

and we have to prove that this inequality holds for $\delta_M \leq \delta$ with high probability. We denote by $z_\ell \in \mathbb{C}^\Gamma$ the vector

$$z_\ell \;=\; (e^{-ik\cdot x_\ell})_{k\in\Gamma} \tag{5.1}$$

and by z_ℓ^T its truncation to the index set $T \subset \Gamma$. For vectors y, z we define a rank one operator by $(y \otimes z)(x) = \langle x, y\rangle z$. We note that

$$(z_\ell^T \otimes z_\ell^T)(c) \;=\; \langle c, z_\ell^T\rangle z_\ell^T \;=\; \left(\sum_{j\in T} c_j e^{i(j-k)\cdot x_\ell}\right)_{k\in T}.$$

Observe that we can write $\mathcal{F}_{TX}^*\mathcal{F}_{TX} \;=\; \sum_{\ell=1}^N z_\ell^T \otimes z_\ell^T$. Thus, we have to show that

$$\sup_{|T|\leq M} \left\| I_T - \frac{1}{N}\sum_{\ell=1}^N z_\ell^T \otimes z_\ell^T \right\|_{2\to 2} \leq \delta \tag{5.2}$$

with probability at least $1-\epsilon$. To this end we consider the expectation of the above expression. Further, we introduce an auxiliary matrix norm,

$$|||A||| \;=\; |||A|||_M \;:=\; \sup_{|T|\leq M} \|A_{T,T}\|_{2\to 2}$$

where $A_{T,T}$ denotes the submatrix of a matrix A consisting of the columns and rows indexed by T. The left hand side of (5.2) can be written as

$$X_N \;:=\; \sup_{|T|\leq M} \left\| I_T - \frac{1}{N}\sum_{\ell=1}^N z_\ell^T \otimes z_\ell^T \right\|_{2\to 2} \;=\; |||I - \frac{1}{N}\sum_{\ell=1}^N z_\ell \otimes z_\ell||| \;=\; |||\sum_{\ell=1}^N N^{-1}(I - z_\ell \otimes z_\ell)|||.$$

The random matrices $N^{-1}(I - z_\ell \otimes z_\ell)$, $\ell = 1, \ldots, N$, are stochastically independent. Moreover, it is easy to see that for both probability models (1) and (2) $\mathbb{E}[z_\ell \otimes z_\ell] = I$ and $I - z_\ell \otimes z_\ell$ is symmetric. Then by standard symmetrization techniques, see e.g. [22, Lemma 6.3], we have

$$\begin{aligned}\mathbb{E}X_N &= \mathbb{E} \sup_{|T|\le M} \left\| I_T - \frac{1}{N}\sum_{\ell=1}^N z_\ell^T \otimes z_\ell^T \right\|_{2\to 2} = \mathbb{E}\left[\left\| I - \frac{1}{N}\sum_{\ell=1}^N z_\ell \otimes z_\ell \right\| \right] \\ &\le 2\mathbb{E}\left[\left\| \frac{1}{N}\sum_{\ell=1}^N \epsilon_\ell z_\ell \otimes z_\ell \right\| \right] = 2\mathbb{E} \sup_{|T|\le M} \left\| \frac{1}{N}\sum_{\ell=1}^N \epsilon_\ell\, z_\ell^T \otimes z_\ell^T \right\|_{2\to 2}, \end{aligned} \tag{5.3}$$

where the ϵ_ℓ are independent symmetric random variables taking values in $\{-1,+1\}$, also jointly independent of the x_ℓ. Now the core of the proof is the following lemma due to Rudelson and Vershynin [30, Lemma 3.5].

Lemma 5.1. *Let $z_1,\dots,z_N$, $N \le D$, be (fixed) vectors in $\mathbb{C}^D$ with uniformly bounded entries, $\|z_\ell\|_\infty \le 1$. Then*

$$\mathbb{E} \sup_{|T|\le M} \left\| \sum_{\ell=1}^N \epsilon_\ell\, z_\ell^T \otimes z_\ell^T \right\|_{2\to 2} \le K(M,N,D) \sup_{|T|\le M} \left\| \sum_{\ell=1}^N z_\ell^T \otimes z_\ell^T \right\|_{2\to 2}^{1/2}$$

where

$$K(M,N,D) = C_0\sqrt{M}\log(M)\sqrt{\log(D)}\sqrt{\log(N)}.$$

We remark that the elegant proof of this lemma uses entropy methods, in particular, Dudley's inequality [22, Theorem 11.17] for the maximum of a Gaussian process.

Now, as in [30], we denote $E = \mathbb{E}[X_N]$. Using (5.3), taking the expectation only with respect to the variables ϵ_ℓ, applying Lemma 5.1 and Hölder's inequality we obtain

$$\begin{aligned} E &\le \frac{2K(M,N,D)}{\sqrt{N}} \mathbb{E} \sup_{|T|\le M} \left\| \frac{1}{N}\sum_{\ell=1}^N z_\ell^T \otimes z_\ell^T \right\|_{2\to 2}^{1/2} \\ &\le \frac{2K(M,N,D)}{\sqrt{N}} \left(\mathbb{E} \sup_{|T|\le M} \left\| I_T - \frac{1}{N}\sum_{\ell=1}^N z_\ell^T \otimes z_\ell^T \right\|_{2\to 2} + 1 \right)^{1/2} = \frac{2K(M,N,D)}{\sqrt{N}}\sqrt{E+1}. \end{aligned}$$

It follows that $E \le \theta$ provided

$$\frac{2K(M,N,D)}{\sqrt{N}} \le \frac{\theta}{\sqrt{1+\theta}}. \tag{5.4}$$

To finish the proof we need to show that the random variable on the left hand side of (5.2) does not deviate much from its expectation. Inspired by [3] we proceed differently as in [31] and use the following version of Talagrand's concentration inequality [32] proved by Klein and Rio in [19].

Theorem 5.2. *Let $Y_1,\dots,Y_N$ be a sequence of independent random variables with values in some Polish space X. Let $\mathcal{F}$ be a countable collection of real-valued measurable and bounded functions f on X with $\|f\|_\infty \le B$ for all $f \in \mathcal{F}$. Let Z be the random variable*

$$Z = \sup_{f\in\mathcal{F}} \sum_{\ell=1}^N f(Y_\ell).$$

Assume $\mathbb{E}f(Y_\ell) = 0$ *for all* $\ell = 1, \ldots, N$ *and all* $f \in \mathcal{F}$. *Let* $\sigma^2 := \sup_{f\in\mathcal{F}} \sum_{\ell=1}^N \mathbb{E}f(Y_\ell)^2$. *Then for* $t \geq 0$

$$\mathbb{P}(Z \geq \mathbb{E}Z + t) \leq \exp\big(-\frac{t}{4B}\log\big(1 + 2\log\big(1 + \frac{Bt}{2B\mathbb{E}Z + \sigma^2}\big)\big)\big).$$

In order to apply the Theorem, we observe that

$$\begin{aligned}
X_N &= \sup_{|T|\leq M} \|I_T - \frac{1}{N}\sum_{\ell=1}^N z_\ell^T \otimes z_\ell^T\| \\
&= \sup_{|T|\leq M} \sup_{v\in\mathbb{C}^T, \|v\|_2\leq 1} \sup_{w\in\mathbb{C}^T, \|w\|_2\leq 1} \big|\frac{1}{N}\sum_{\ell=1}^N \langle (I_T - z_\ell^T \otimes z_\ell^T)v, w\rangle\big| \\
&= \sup_{(v,w)\in S_M} \big|\frac{1}{N}\sum_{\ell=1}^N \langle (I - z\otimes z)v, w\rangle\big|,
\end{aligned}$$

where

$$S_M = \{(v,w) \in \mathbb{C}^D : \|v\|_2, \|w\|_2 \leq 1; \operatorname{supp} v = \operatorname{supp} w = T \text{ for some } T \text{ with } |T| \leq M\}.$$

Defining

$$f_{v,w}(z) \;=\; \frac{1}{N}\langle (I - z\otimes z)v, w\rangle.$$

we obtain

$$X_N = \sup_{(v,w)\in S_M} |\sum_{\ell=1}^N f_{v,w}(z_\ell)|.$$

Clearly, $\mathbb{E}f_{v,w}(z_\ell) = N^{-1}\langle \mathbb{E}(I - z_\ell \otimes z_\ell)v, w\rangle = 0$. Furthermore, for $(v,w) \in S_M$ and $z = (e^{ik\cdot x})_{k\in\Gamma}$ we have

$$\begin{aligned}
|f_{v,w}(z)| &= \frac{1}{N}\big|\sum_{j,k\in T, j\neq k} v_j e^{i(j-k)x}\overline{w_k}\big| \leq N^{-1}\sum_{j,k\in T, j\neq k} |v_j||w_k| \\
&\leq N^{-1}\sum_{j\in T}\sum_{k\in T} |v_k||w_{\sigma_j(k)}| = N^{-1}\sum_{j\in T}\langle |v|, |w^{(\sigma_j)}|\rangle \leq N^{-1}M\|v\|_2\|w\|_2 \leq \frac{M}{N},
\end{aligned} \tag{5.5}$$

where $\{(k, \sigma_j(k)), j, k \in T\}$ is a reparametrization of $T \times T$ such that $\sigma_j(T) = T$, i.e., σ_j is a suitable permutation; and $w^{(\sigma_j)}$ denotes the corresponding vector of reordered entries of w. Above we used the Cauchy Schwarz inequality in the fifth step. We deduced $\|f_{v,w}\|_\infty \leq M/N$ for all $(v,w) \in S_M$.

Next, for $(v,w) \in S_M$, we compute

$$\begin{aligned}
\mathbb{E}|f_{v,w}(z_\ell)|^2 &= N^{-2}\mathbb{E}\big|\sum_{j,k\in T, j\neq k} v_j e^{i(j-k)\cdot x_\ell}\overline{w_k}\big|^2 \\
&= N^{-2}\sum_{j,k\in T, j\neq k}\;\sum_{j',k'\in T, j'\neq k'} v_j\overline{v_{j'}}\,\overline{w_k}w_{k'}\mathbb{E}[e^{i(j-k-j'+k')\cdot x_\ell}].
\end{aligned}$$

Since x is uniformly distributed on $[0,2\pi]^d$ or on $\frac{2\pi}{m}\mathbb{Z}_m^d$ we have $\mathbb{E}[e^{i(j-k-j'+k')\cdot x_\ell}] = \delta_{j',j-k+k'}$ and, hence,

$$\begin{aligned}\mathbb{E}|f_{v,w}(z_\ell)|^2 &= N^{-2} \sum_{j,k\in T, j\neq k} \sum_{k'\in T} v_j \overline{v_{j-k+k'}} w_{k'} \overline{w_k} \leq N^{-2}\|v\|_2^2 \sum_{k,k'\in T} w_{k'}\overline{w_k} \\ &\leq N^{-2}\|v\|_2^2\, |T|\, \|w\|_2^2 \leq M/N^2.\end{aligned}$$

In the second step we applied the Cauchy Schwarz inequality and in the third step a similar estimate as in (5.5). Hence,

$$\sigma^2 = \sup_{(v,w)\in S_M} \sum_{\ell=1}^N \mathbb{E}|f_{(v,w)}(z_\ell)|^2 \leq M/N.$$

Theorem 5.2 applies to real-valued functions f. Hence, we split into real and imaginary parts $f^r_{v,w} = \mathrm{Re}(f_{v,w})$, $f^i_{v,w} = \mathrm{Im}(f_{v,w})$. Then the estimates above apply also to these functions, i.e., $\|f^r_{v,w}\|_\infty, \|f^i_{v,w}\|_\infty \leq \frac{M}{N}$ and $\sigma_r^2, \sigma_i^2 \leq \frac{M}{N}$.

Denote $Z^r = \sup_{(v,w)\in S_M} \sum_{\ell=1}^N f^r_{v,w}(z_\ell)$ and similarly define Z^i. Since $f_{v,-w} = -f_{v,w}$ we have $Z^r = \sup_{(v,w)\in S_M} |\sum_{\ell=1}^N f_{v,w}(z_\ell)|$. By the union bound

$$\mathbb{P}(X_N \geq \delta) = \mathbb{P}(|Z^r|^2 + |Z^i|^2 \geq \delta^2) \leq \mathbb{P}\big(Z^r \geq \frac{\delta}{\sqrt{2}}\big) + \mathbb{P}\big(Z^i \geq \frac{\delta}{\sqrt{2}}\big).$$

Now assume $\mathbb{E}X_N \leq \delta/2$, which by (5.4) will be satisfied provided $2K(M,N,D)/\sqrt{N} \leq \frac{\delta/2}{\sqrt{1+\delta/2}}$, in particular, if

$$2K(M,N,D)/\sqrt{N} \leq \frac{\delta}{2\sqrt{3/2}} = \frac{\delta}{\sqrt{6}}.$$

Setting $t = \frac{\delta}{\sqrt{2}} - \frac{\delta}{2} = \frac{\sqrt{2}-1}{2}\delta =: c\delta$ in Theorem 5.2 we obtain

$$\mathbb{P}(X_N \geq \delta) \leq 2\exp\left(-\frac{t}{4M/N}\log(1 + 2\log(1+\frac{t}{\delta+1}))\right) = 2e^{-c_0(\delta)N/M},$$

where $c_0(\delta) = c\delta\log(1+2\log(1+\frac{c\delta}{\delta+1}))$. In other words, $X_N \leq \delta$ with probability at least $1-\epsilon$ provided $N \geq c_0(\delta)^{-1}M\log(2/\epsilon)$ and $2K(M,N,D)/\sqrt{N} \leq \delta/\sqrt{6}$. Note that $C_2\delta^{-2} \geq c_0(\delta)^{-1}$ for all $\delta \in (0,1)$ where $C_2 = c_0(1)^{-1} = \frac{2}{\sqrt{2}-1}\log(1+2\log(1+\frac{\sqrt{2}-1}{4}))^{-1} \approx 26.84$. With the definition of $K(M,N,D)$ we deduce that $\delta_M \leq \delta$ with probability at least $1-\epsilon$ provided

$$\frac{N}{\log(N)} \geq C_1\delta^{-2}M\log^2(M)\log(D) \quad \text{ and } \quad N \geq C_2\delta^{-2}M\log(2\epsilon^{-1}).$$

Both conditions are satisfied once

$$\frac{N}{\log(N)} \geq C\delta^{-2}M\log^2(M)\log(D)\log(\epsilon^{-1})$$

for some suitable constant C. This finishes the proof of Theorem 3.2.

6 Proofs for Orthogonal Matching Pursuit

6.1 Proof of Theorem 4.1

The proof is an extension of the one in [21]. We will use the following result from [17] on the eigenvalues of a submatrix $\mathcal{F}_{TX}$, which is based on the analysis in [28, Lemma 3.3 and Section 3.3].

Theorem 6.1. *Let T of size $|T| = M$ and let $x_1, \ldots, x_N$ be i.i.d. random variables that are uniformly distributed over $[0, 2\pi]^d$ or over the grid $\frac{2\pi}{m}\mathbb{Z}_m^d$. Choose $\epsilon, \delta \in (0, 1)$ and assume*

$$\left\lfloor \frac{\delta^2 N}{3eM} \right\rfloor \geq \ln(c(\delta)M/\epsilon), \tag{6.1}$$

where $c(\delta) = (1 - \delta^2/e)^{-1} \leq (1 - e^{-1})^{-1} \approx 1.582$. Then with probability at least $1 - \epsilon$ the minimal and maximal eigenvalue of $\mathcal{F}_{TX}^* \mathcal{F}_{TX}$ satisfy

$$1 - \delta \leq \lambda_{\min}(N^{-1}\mathcal{F}_{TX}^*\mathcal{F}_{TX}), \quad \text{and} \quad \lambda_{\max}(N^{-1}\mathcal{F}_{TX}^*\mathcal{F}_{TX}) \leq 1 + \delta. \tag{6.2}$$

Further, we need the following concentration inequality proved in [21].

Lemma 6.2. *Assume that c is a vector supported on T. Further, assume that the sampling set X is chosen according to one of our two probability models. Then for $j \notin T$ and $t > 0$ it holds*

$$\mathbb{P}\left(|N^{-1}\langle \mathcal{F}_{TX}c, \phi_j\rangle| \geq t\right) \leq 4\exp\left(-N\frac{t^2}{4\|c\|_2^2 + \frac{4}{3\sqrt{2}}\|c\|_1 t}\right).$$

Now we can turn to the proof of Theorem 4.1. (Orthogonal) Matching Pursuit selects an element of the support $\operatorname{supp} c =: T$ in the first iteration if

$$\max_{j \notin T} |N^{-1}\langle \phi_j, \mathcal{F}_{TX}c + \eta\rangle| < \max_{k \in T} |N^{-1}\langle \phi_k, \mathcal{F}_{TX}c + \eta\rangle|. \tag{6.3}$$

By the triangle inequality and Cauchy-Schwarz (note also that $\|\phi_k\|_2 = \sqrt{N}$) this will be satisfied if

$$\max_{j \notin T} |N^{-1}\langle \phi_j, \mathcal{F}_{TX}c\rangle| \leq \|N^{-1}\mathcal{F}_{TX}^*\mathcal{F}_{TX}c\|_\infty - \frac{2}{\sqrt{N}}\|\eta\|_2.$$

Assume for the moment that $\lambda_{\min}(N^{-1}\mathcal{F}_{TX}^*\mathcal{F}_{TX}) \geq 1 - \delta$ for some $\delta \in (0, 1)$. (The probability that this happens can be estimated by Theorem 6.1.) This yields

$$\|N^{-1}\mathcal{F}_{TX}^*\mathcal{F}_{TX}c\|_\infty \geq M^{-1/2}\|N^{-1}\mathcal{F}_{TX}^*\mathcal{F}_{TX}c\|_2 \geq M^{-1/2}(1 - \delta)\|c\|_2.$$

Thus, (6.3) is satisfied if

$$\max_{j \notin T} |N^{-1}\langle \phi_j, \mathcal{F}_{TX}c\rangle| \leq \frac{1 - \delta}{\sqrt{M}}\|c\|_2 - \frac{2}{\sqrt{N}}\|\eta\|_2. \tag{6.4}$$

Assuming further that

$$\|\eta\|_2 \leq \frac{(1 - \delta)(1 - \tau)}{2}\sqrt{\frac{N}{M}}\|c\|_2 \tag{6.5}$$

condition (6.4) becomes true if

$$\max_{j \notin T} |N^{-1}\langle \phi_j, \mathcal{F}_{TX} c\rangle| \leq \frac{1-\delta}{\sqrt{M}} \tau \|c\|_2.$$

By the concentration inequality in Lemma 6.2 the probability that the above inequality does not hold can be estimated by

$$\begin{aligned}
\mathbb{P}\left(\max_{j \notin T} |N^{-1}\langle \phi_j, \mathcal{F}_{TX} c\rangle| \leq \frac{1-\delta}{\sqrt{M}} \tau \|c\|_2\right) &\leq \sum_{j \notin T} \mathbb{P}\left(|N^{-1}\langle \phi_j, \mathcal{F}_{TX} c\rangle| \leq \frac{1-\delta}{\sqrt{M}} \tau \|c\|_2\right) \\
&\leq 4D \exp\left(-\frac{N}{M} \frac{(1-\delta)^2 \tau^2 \|c\|_2^2}{4\|c\|_2^2 + \frac{4}{3\sqrt{2}} \|c\|_1 M^{-1/2} (1-\delta) \tau \|c\|_2}\right) \\
&\leq 4D \exp\left(-\frac{N}{M} \frac{(1-\delta)^2 \tau^2}{4 + \frac{4}{3\sqrt{2}} (1-\delta)}\right). \qquad (6.6)
\end{aligned}$$

In the last line we used the Cauchy-Schwarz inequality, $\|c\|_1 \leq \sqrt{M}\|c\|_2$. Now we choose $\delta = 1/2$. Then condition (6.5) becomes (4.2) and

$$\mathbb{P}\left(\max_{j \notin T} |N^{-1}\langle \phi_j, \mathcal{F}_{TX} c\rangle| \leq \frac{1-\delta}{\sqrt{M}} \tau \|c\|_2\right) \leq 4D \exp\left(-\frac{N}{M} \frac{\tau^2}{16 + \frac{8}{3\sqrt{2}}}\right).$$

The latter term is less than $\epsilon/2$ if

$$N \geq C M \tau^{-2} \log(8D/\epsilon)$$

with $C = 16 + \frac{8}{3\sqrt{2}} \approx 17.88$. Furthermore, by Theorem 6.1 our initial assumption that $\lambda_{\min}(N^{-1}\mathcal{F}_{TX} * \mathcal{F}_{TX}) \geq 1 - \delta = 1/2$ fails with probability at most $\epsilon/2$ if

$$\left\lfloor \frac{N}{12eM} \right\rfloor \geq \ln(2(1 - e^{-1}/4)^{-1} M/\epsilon).$$

Altogether, the probability that OMP does not select an element of T in the first step is less than ϵ if

$$N \geq C M \tau^{-2} \log(D/\epsilon)$$

for some suitable constant C.

Now consider the final statement of the Theorem, i.e., assume that OMP has reconstructed the true support T after M steps. Then the reconstructed coefficients are given by $\tilde{c} = \mathcal{F}_{TX}^{\dagger}(\mathcal{F}_{TX} c + \eta)$ where $\mathcal{F}_{TX}^{\dagger}$ denotes the pseudo-inverse of $\mathcal{F}_{TX}$. Observe that $\mathcal{F}_{TX}^{\dagger}\mathcal{F}_{TX} c = c$. Hence

$$\|\tilde{c} - c\|_2 = \|\mathcal{F}_{TX}^{\dagger}\eta\|_2 \leq \|\mathcal{F}_{TX}^{\dagger}\| \|\eta\|_2 = \sqrt{\lambda_{\min}(\mathcal{F}_{TX}^* \mathcal{F}_{TX})^{-1}} \|\eta\|_2 \leq \frac{1}{\sqrt{N(1-\delta)}} \sigma = \sqrt{\frac{2}{N}} \sigma,$$

where we used Theorem 6.1 once more.

6.2 Proof of Theorem 4.2

The proof of the uniform recovery result is based on the coherence parameter, which measures the maximum correlation between distinct columns of a matrix $A = (\psi_1 | \dots | \psi_D)$, i.e.,

$$\mu := \max_{j \neq k} |\langle \psi_j, \psi_k \rangle|.$$

Based on μ the following theorem due to Donoho, Elad and Temlyakov [12, Theorem 4.1] analyzes the performance of OMP in the presence of noise.

Theorem 6.3. *Assume that A has coherence μ. Suppose that $y = Ac + \eta$ with only M coefficients of c being nonzero and $\|\eta\|_2 \leq \sigma$. Suppose $(2M-1)\mu < 1$ and*

$$\sigma < \frac{1-(2M-1)\mu}{2} \min_{k \in \operatorname{supp} c} |c_k|.$$

If we run OMP until the residual satisfies $\|r_s\| \leq \sigma$ then the true support of c has been recovered, and consequently OMP has done M iterations. Furthermore, the error between the reconstructed coefficients $\tilde{c}$ and the original coefficients satisfies

$$\|c - \tilde{c}\|_2 \leq \frac{1}{\sqrt{1-(M-1)\mu}} \sigma.$$

In [21] the following estimate of the coherence of $\mathcal{F}_X$ was proven.

Lemma 6.4. *Let the random sampling set $X = (x_1, \dots, x_N)$ be chosen according to one of our probability models and let μ be the coherence of the random matrix $N^{-1/2}\mathcal{F}_X$. Then*

$$\mathbb{P}(\mu > t) \leq 4D' \exp\left(-N \frac{t^2}{4 + \frac{4}{3\sqrt{2}}t}\right),$$

where $D' = \#\{j-k : j,k \in \Gamma, j \neq k\} \leq D^2$.

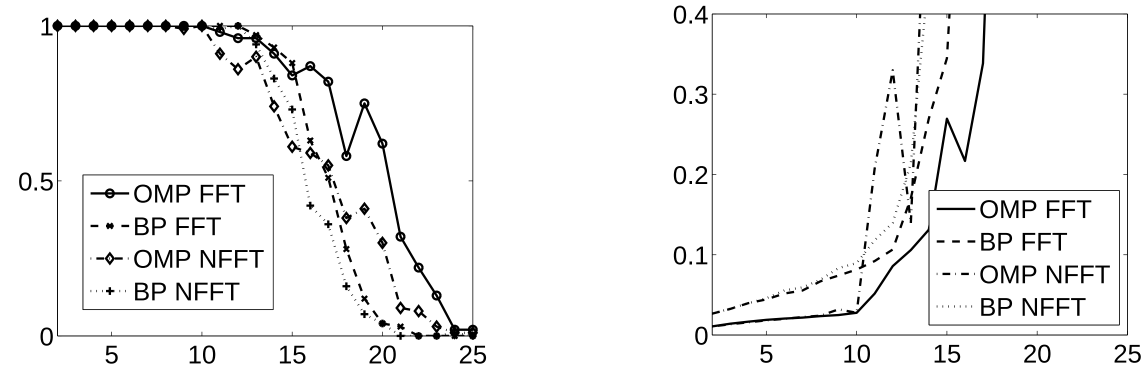
Remark 6.1. In case of the continuous probability model the previous estimate can be slightly improved to [21]

$$\mathbb{P}(\mu > t) \leq (1-\kappa)^{-1} D' e^{-N\kappa t^2}, \qquad \kappa \in (0,1).$$

Now the proof of Theorem 4.2 is a mere application of the above statements. Note that $y = \mathcal{F}_X c + \eta = N^{-1/2}\mathcal{F}_X(\sqrt{N}c) + \eta$. Thus, setting $t = \frac{\tau}{2M-1}$ in the lemma, solving for N and using Theorem 6.3 with $c' = \sqrt{N}c$ shows the assertion.

7 Numerical Experiments

To illustrate the theoretical results we also conducted numerical experiments. We choose a number of samples N, the noise level σ, the sparsity M and an (even) dimension D and set $\Gamma = \{-D/2+1, \dots, D/2\}$. Then we repeat the following reconstruction experiment 100 times. We choose a subset T uniformly at random among all subsets of size M. Then we randomly select the real part and imaginary part of the coefficients c_k on T from a standard normal distribution. The sampling points $x_1, \dots, x_N$ are randomly drawn either from the uniform distribution on $[0, 2\pi]$ (probability model (1), labelled NFFT in the plots) or uniformly among



(a) Recovery success rate of the true support.

(b) Avarage error.

Figure 2: Recovery of sparse trigonometric polynomials in dimension $D = 256$ from $N = 50$ noisy samples, $\|\eta\|_2 = \sigma = 0.4$. The sparsity M is varied.

all subsets of $\{0, \frac{2\pi}{D}, \dots, \frac{2\pi(D-1)}{D}\}$ of size N (a slight variation of the probability model (2) preventing that some of the sampling points coincide, labelled FFT). The perturbed sampling points are given by $y_\ell = \sum_{k \in T} c_k e^{ik \cdot x_\ell} + \eta_\ell$, $\ell = 1, \dots, N$, where the noise vector η is chosen uniformly at random on the sphere with radius σ in $\mathbb{C}^N$, i.e. $\|\eta\|_2 = \sigma$.

Then we solve the ℓ_1-minimization problem (3.1) (with the chosen σ) and run OMP (with precisely M iterations), respectively, and compute the error between the reconstructed vector $\tilde{c}$ and the original vector c for both methods. Also we test whether the correct support has been recovered.

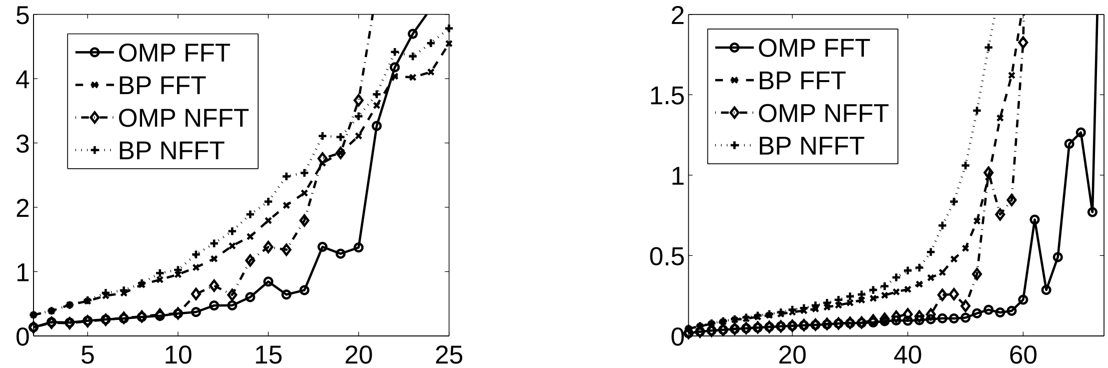
Figures 2 and 3 show the results for varying sparsity, while in Figure 4 the noise level σ is varied. These plots indicate that the BP variant and OMP are both stable under noise as predicted by the theoretical results. Figure 4 suggests that the correct support set can be recovered even when the noise level reaches the order of the ℓ_2-energy of the samples of the signal. Moreover, OMP usually performs slightly better than BP. In fact, OMP yields a smaller avarage reconstruction error and also reconstructs more often the correct support - despite that fact that theoretically BP gives a uniform recovery guarantee while OMP does not. This might be due to the fact that OMP forces the reconstruction to be M-sparse while BP may result in larger support sets. Furthermore, OMP is much faster than BP (by a factor between 10 and 200 in the examples). For a more detailed comparison of the computation times we refer to [21].

The Matlab toolbox CVX [16] was used for solving (3.1). The examples (including the OMP algorithm) are part of the Matlab toolbox [20], which is available online.

8 Discussion

We presented theoretical and numerical results concerning the stability of recovery of sparse trigonometric polynomials with (a variant of) Basis Pursuit and Orthogonal Matching Pursuit. The (non-uniform) recovery Theorem 4.1 for OMP, however, is only partial so far. It remains open to analyze theoretically the further iterations after the first step.

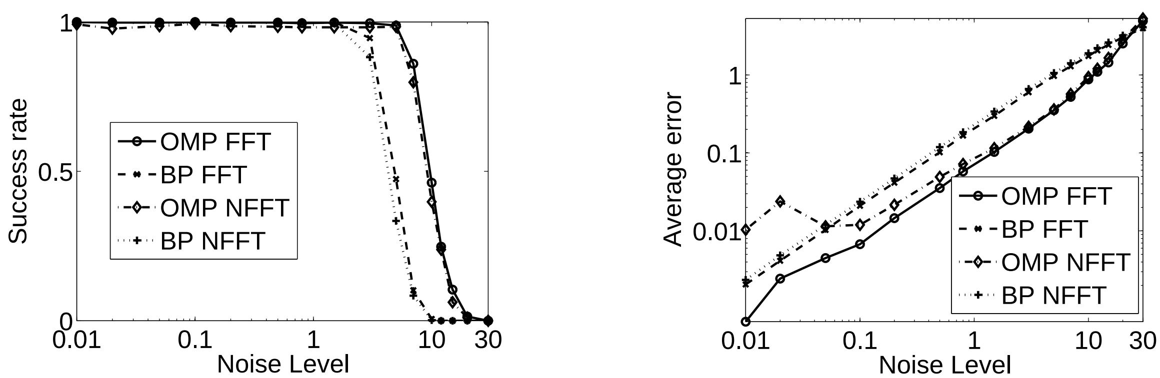
BP has the advantage of giving a uniform guarantee of recovery success, i.e., a single



(a) Average error, $\sigma = 5$, $D = 256$, $N = 50$

(b) Average error, $\sigma = 2$, $D = 512$, $N = 150$

Figure 3: Recovery of sparse trigonometric polynomials for different sets of parameters. The sparsity is varied.



(a) Recovery success, $D = 256$, $N = 50$, $M = 10$

(b) Average error, $D = 256$, $N = 50$, $M = 10$

Figure 4: Recovery of sparse trigonometric polynomials for different sets of parameters. The noise level σ is varied. (For comparison, the average ℓ_2-norm of the vector of samples of the unperturbed polynomial is approximately 39.4).

sampling set X may be sufficient to recover *all* sparse trigonometric polynomials, while it seems that OMP is only able to provide non-uniform recovery results at reasonably small ratio of the number of samples to the sparsity [29]. (But note the results for variants of OMP in [26, 25, 24].) In practice, however, a non-uniform guarantee might be sufficient and indeed our numerical experiments show that OMP even slightly outperforms BP on generic (=random) signals.

Corollary 3.3 concerning BP covers also the case that the coefficient vector c is not sparse in a strict sense. In this case it estimates the approximation error of the reconstruction by the approximation error with M-terms. In principle, we might also apply Theorems 4.1 and 4.2 for OMP to the non-sparse case by letting $\eta = \mathcal{F}_X c_{\Gamma \setminus T}$, i.e., by treating the contribution of the (small) coefficients outside T as noise. However, for most situations conditions (4.4) and (4.6) on the magnitude of the coefficients become then unrealistic. Roughly speaking they would imply that the smallest coefficient of c in T is significantly larger than the ℓ_2-norm of the coefficients outside T. So a thorough treatment of the non-sparse case for OMP is still open.

OMP is usually faster (and easier to implement) than BP in practice, and the numerical results even indicate that OMP is slightly more stable. So in most practical situations one would probably prefer to use OMP despite its lack of giving a uniform recovery guarantee when the number of samples is only linear in the sparsity.

Acknowledgements

The author would like to thank Stefan Kunis, Ingrid Daubechies, Joel Tropp, Rémi Gribonval, Pierre Vandergheynst, Karin Schnass and Roman Vershynin for enlightening discussions. Parts of the manuscript were written while visiting the group of Pierre Vandergheynst at EPFL in Lausanne. He thanks its members for their warm hospitality. His stay was funded by the European Union's Human Potential Program under contract HPRN-CT-2002-00285 (HASSIP). The author is currently supported by an Individual Marie Curie fellowship from the European Union under contract MEIF-CT-2006-022811.

References

[1] D. Baron, M. B. Wakin, M. F. Duarte, S. Sarvotham, and R. G. Baraniuk. Distributed compressed sensing. *preprint*, 2005.

[2] S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge Univ. Press, 2004.

[3] E. Candès and J. Romberg. Sparsity and incoherence in compressive sampling. *Inverse Problems*, 23(3):969–985, 2007.

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

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

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

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

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

[9] I. Daubechies, M. Fornasier, and I. Loris. Accelerated projected gradient methods for linear inverse problems with sparsity constraints. *J. Fourier Anal. Appl.*, to appear.

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

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

[12] D. L. Donoho, M. Elad, and V. N. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Trans. Inform. Theory*, 52(1):6–18, 2006.

[13] D. L. Donoho and J. Tanner. Sparse nonnegative solutions of underdetermined linear equations by linear programming. *Proc. Nat. Acad. Sci.*, 102(27):9446–9451, 2005.

[14] M. A. T. Figueiredo, R. D. Nowak, and S. Wright. Gradient projection for sparse reconstruction: Application to compressed sensing and other inverse problems. *IEEE J. Sel. Top. Signal Proces.*, 1(4):586–598, 2007.

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

[16] M. Grant, S. Boyd, and Y. Ye. CVX: Matlab software for disciplined convex programming, Version 1.1. Available: `http://www.stanford.edu/~boyd/cvx`, 2008.

[17] K. Gröchenig, B. Pötscher, and H. Rauhut. Learning trigonometric polynomials from random samples and exponential inequalities for eigenvalues of random matrices. *preprint*, 2007.

[18] S. Kim, K. Ksh, M. Lustig, S. Boyd, and D. Gorinevsky. A method for large-scale l1-regularized least squares problems with applications in signal processing and statistics. *IEEE J. Sel. Top. Signal Proces.*, 4(1):606–617, 2007.

[19] T. Klein and E. Rio. Concentration around the mean for maxima of empirical processes. *Ann. Probab.*, 33(3):1060–1077, 2005.

[20] S. Kunis and H. Rauhut. OMP4NFFT, Matlab toolbox. Available: `http://www.tu-chemnitz.de/~skunis/software.php`, 2006.

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

[22] M. Ledoux and M. Talagrand. *Probability in Banach spaces. Isoperimetry and processes.*, volume 23 of *Ergebnisse der Mathematik und ihrer Grenzgebiete. 3. Folge.* Springer-Verlag, Berlin, Heidelberg, NewYork, 1991.

[23] S. G. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Trans. Signal Process.*, 41(12):3397–3415, 1993.

[24] D. Needell and J. A. Tropp. CoSamp: Iterative signal recovery from incomplete and inaccurate samples. *Preprint*, 2008.

[25] D. Needell and R. Vershynin. Signal recovery from incomplete and inaccurate measurements via regularized orthogonal matching pursuit. *Preprint*, 2007.

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

[27] D. Potts, G. Steidl, and M. Tasche. *Fast Fourier transforms for nonequispaced data: A tutorial*, chapter 12, pages 247 – 270. Birkhäuser, 2001.

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

[29] H. Rauhut. On the impossibility of uniform sparse reconstruction using greedy methods. *Sampl. Theory Signal Image Process.*, to appear.

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

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

[32] M. Talagrand. New concentration inequalities in product spaces. *Invent. Math.*, 126(3):505–563, 1996.

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

[34] J. Zou, A. C. Gilbert, M. Strauss, and I. Daubechies. Theoretical and experimental analysis of a randomized algorithm for sparse Fourier transform analysis. *J. Comput. Phys.*, 211:572–595, 2005.