Well-Posedness of Sparse Frequency Estimation

Benedikt Diederichs*

May 21, 2019

Abstract

The problem of estimating the frequencies of an exponential sum has been studied extensively over the last years. It can be understood as a sparse estimation problem, as it strives to identify the sparse representation of a signal using exponentials. In this paper, we are interested in its intrinsic stability properties. We derive a bound very similar to the restricted isometry property. Conditional well-posedness follows: Any exponential sum with samples close to the unknown ground truth has close frequencies as well, provided that it satisfies our model assumptions. The most important assumption is that the frequencies are well-separated. Furthermore, we show that the presented bound is sharp and gives rise to improved estimates of condition numbers of certain Vandermonde matrices.

Key words. super-resolution, sparse estimation, frequency analysis, exponential sum

AMS subject classifications. 65T40, 42C15, 15A18

1 Introduction

Spectral analysis of a given signal is one of the most ubiquitous problems in signal processing. Depending on the problem at hand, one can choose from a large collection of well-studied methods.

In this paper, we are interested in one of the most important special cases, where the signal satisfies the harmonic model, i.e.,

$$f(x) = \sum_{y \in Y^f} c_y e^{2\pi i y x}, \quad c_y \in \mathbb{C} \setminus \{0\}.$$  

Here, $Y^f \equiv Y \subset \mathbb{T} = \mathbb{R}/\mathbb{Z} \simeq [0, 1)$ is the finite set of frequencies of $f$, with coefficients (or amplitudes) $c_y \in \mathbb{C} \setminus \{0\}$. Note that we enumerate $c^f \equiv c = (c_y)_{y \in Y} \in \mathbb{C}^Y$ by the set $Y$, which is convenient if one does not want to fix an enumeration of $Y$.

Such an $f$ is then sampled at integers $k = 0, \ldots, N$, and we wish to estimate $Y$ and the coefficients $c$. Typically, the available samples are corrupted by noise.

This problem has been studied for a long time, and many relevant applications are known, see for example the textbook [16]. It reemerged over the last years for a number of reasons. The first is that using the harmonic model allows to overcome

*University of Passau and Fraunhofer IIS Research Group Knowledge Based Image Processing, Passau, Germany
resolution limits of generic methods, i.e., super-resolution is possible. And while that 
is well known, new ideas are further expanding the possibilities in applications.

More interesting for this paper is the connection to compressed sensing. As \( f \) is 
sparse in the frequency domain (its spectrum is a sum of Dirac deltas), the frequency 
estimation problem can be seen as finding a sparse representation of \( f \) from an un-
countable, highly correlated dictionary consisting of all functions \( x \mapsto e^{2\pi i x y} \). Often, 
the problem is discretized, see for example [11] [7]. An alternative is an approach using 
TV-minimization, as promoted in [4] [5] [8] and others.

In contrast to the existing literature, we are interested in the intrinsic stability 
properties of the continuous problem. We prove that for any two exponential sums 
satisfying our model assumption, close samples imply close frequencies. The model 
assumptions include most notably a separation condition of the frequencies. That 
bound can be interpreted as a continuous analog of the restricted isometry property in 
compressed sensing and is valid regardless of a solution method.

As the connection to compressed sensing is of central importance for the paper, 
we give an informal explanation and motivation in the next section. After that, we 
state and prove the main results in Section 3. In the last two sections we present some 
corollaries: In Section 4 we give lower bounds for singular values of Vandermonde 
matrices and in Section 5 we finally state the well-posedness of the problem and a 
posteriori error bounds.

To keep notation simple, we use \( a \lesssim b \), which means that \( a \leq C b \), where \( C \) is 
a positive constant, independent of any quantities of interest (on which \( a, b \) might 
depend). Further, \( a \sim b \) means \( a \lesssim b \) and \( b \lesssim a \). On the other hand, \( a \approx b \) is used as 
an informal notation for \( a \) and \( b \) being close.

We call an vector \( s \)-sparse, if all but \( s \) of its entries are zero. The Fourier transform 
is defined by

\[
\mathcal{F}f(w) = \hat{f}(w) = \int_{\mathbb{R}} f(x)e^{-2\pi i xw} \, dx.
\]

2 Connection to Sparse Estimation

We start by formulating an abstract sparse estimation problem. We are interested in 
identifying

\[
x = \sum_{\theta \in Y} c_{\theta} h(\theta) \in \mathcal{H}.
\]

Here, \( Y \subset \Theta \) is a finite subset of \( \Theta \subset \mathbb{R}^d \), the set of valid parameters. \( c_{\theta} \in \mathbb{C} \setminus \{0\} \) 
are the coefficients, \( \mathcal{H} \) is a Banach space and \( h : \Theta \to \mathcal{H} \) is a continuous and injective 
function.

Given are a finite number of measurements of the form

\[
b_j = a_j(x) + \eta_j, \quad j = 1, \ldots, m,
\]

where \( a_j \in \mathcal{H}' \) are elements of the dual space and \( \eta_j \) is noise, corrupting the measurements.

The frequency estimation problem fits into that framework by choosing

\[
h : \mathbb{T} \to C(I), \quad h(y) = e^{2\pi iy},
\]

2
where $I \subset \mathbb{R}$ is an interval containing all sampling points. $a_j$ are point evaluations at the $j$th sampling point and $C(I)$ is the Banach space of continuous functions from $I$ to $\mathbb{C}$.

To solve this problem, one typically discretizes $\Theta$ by replacing it with $\tilde{\Theta} = \{\tilde{\theta}_1, \ldots, \tilde{\theta}_N\}$. Letting $a_{jk} = a_j(\tilde{\theta}_k)$ and $A = (a_{jk})$ and assuming $\theta_j \in \tilde{\Theta}$, we arrive at the typical problem of compressed sensing, namely

$$Ax = b,$$

where $b = (b_1, \ldots, b_m) \in \mathbb{C}^m$. We identified $x$ with the vector $x \in \mathbb{C}^N$, $x_k = c_j$ if $\tilde{\theta}_k = \theta_j$ and zero otherwise. Now as $N \gg m$ is expected, the problem is underdetermined. However, under suitable conditions on $A$, we can exploit the sparsity of $x$ to recover $(c_{\theta_j}, \theta_j)_{j=1}^s$.

A necessary condition to do so is that if we have given a second sparse vector $y \in \mathbb{C}^N$, we have to have $Ax \approx Ay \Rightarrow x \approx y$. This is satisfied if the matrix $A$ satisfies for a $\delta \in [0, 1)$ the lower bound

$$(1 - \delta)\|z\|_2^2 \leq \|Az\|_2^2 \quad \text{for all 2s-sparse vectors } z,$$

which is just the lower bound of the well-known restricted isometry property. It can be directly applied, as $x - y$ is 2s-sparse.

One simple consequence of (1) is that if all non-zero coefficients of $x$ and $y$ have a modulus of at least $\rho > 0$, we have that

$$\|A(x - y)\|_2^2 < (1 - \delta)\rho^2 \Rightarrow \text{supp } x = \text{supp } y.$$

Combined with a noise model, one can easily derive a posteriori error estimates. Of course all these considerations do not help too much in actually solving the problem $Ax = b$ with $x$ sparse, which is due to its non-convex nature quite difficult. However, without a bound like (1), the problem is ill-posed and its solution non-unique or not stable. See [9] for a thorough introduction to compressive sensing.

In practice, discretizing $\Theta$ poses a difficult problem. When discretizing too coarsely, the assumption $\theta_j \in \tilde{\Theta}$ might be far from true. On the other hand if one discretises $\Theta$ very finely, many columns of $A$ will be almost equal and (1) will not hold. There are possibilities to circumvent these problems. Though not of interest for our approach, we give two examples. An adaptive discretization strategy is possible, see for example [11]. Another alternative is to discretize $\Theta$ finely, and then stick to subspaces where all $\tilde{\theta}_j$’s are not too close. That approach is called structured compressed sensing, see for example [7].

Here, however, we strive to prove similar stability properties without having to discretize $\Theta$ first. Due to the continuity of $h$ it is not possible to stably distinguish between $h(\theta)$ and $h(\theta + \varepsilon)$; indeed, if $h$ is differentiable, we expect that

$$\|h(\theta) - h(\theta + \varepsilon)\| \sim |\varepsilon|, \quad \text{for } \varepsilon \text{ small}.$$

The natural idea is therefore to replace a sparse sum by a well-separated sum. Then one can ask whether it is possible to estimate $\{\theta_1, \ldots, \theta_s\}$ stably, when there is a separation distance $q \in \mathbb{R}_{>0}$ such that $|\theta_j - \theta_k| \geq q$ for all $j \neq k$. 

3
On the first glance, it seems that we have to prove a bound like (1). Such a bound would be of the form
\[ \|c\|_2^2 \lesssim \left\| \sum_{\theta \in Y} c_\theta h(\theta) \right\|_2^2 \] (3)
uniformly over all \( Y \subset \Theta \) satisfying \( |\theta - \theta'| \geq q \) for all \( \theta \neq \theta' \in Y \). This, however, is not sufficient to guarantee that we are able to distinguish two different well-separated families. Let \( Y' = \{ \theta'_1, \ldots, \theta'_s \} \) be a second set with \( q \) separated \( \theta_j \)s. We cannot hope to get a bound like (3), as some \( \theta'_j \) could be arbitrarily close to a \( \theta_k \).

The next best thing is that a good matching between \( Y \) and \( Y' \) must exist if their samples are very close. Clearly, for every \( \theta \in Y \) there is at most one \( \theta' \in Y' \) with \( |\theta - \theta'| < q/2 \), which we call \( n(\theta) \). We denote by \( Y_1 \) all \( y \in Y \) with such a match and by \( Y_3 \) all elements from \( Y \cup Y' \) without a match (the enumeration is chosen to be consistent with later use). Then a bound like
\[ \sum_{\theta \in Y_3} |c_\theta|^2 + \sum_{\theta \in Y_1} \left( |\theta - n(\theta) - c_\theta + c_{n(\theta)}|^2 \right) \lesssim \left\| \sum_{\theta \in Y} c_\theta h(\theta) - \sum_{\theta' \in Y'} c_{\theta'} h(\theta') \right\|_2^2 \]
might be possible. If the right-hand side is sufficiently small and all coefficients have a modulus of at least \( c_{\min} > 0 \), we can deduce that \( Y_3 = \emptyset \) and a one-to-one matching between \( Y \) and \( Y' \) exists. That can be interpreted as the continuous analog of (2). Furthermore, if the right-hand side becomes very small, we can conclude that \( \theta \approx n(\theta) \) and \( c_\theta \approx c_{n(\theta)} \).

To prove such a bound for the frequency estimation problem is the main objective of this paper. The bound is presented in Theorem 3.6 and its implication for the well-posedness is given in Corollary 5.1. An example of an a posteriori bound is given in Corollary 5.3. Clearly, such a bound always implies a lower bound of the corresponding sampling matrix, which is a Vandermonde matrix in our case. We state these bounds explicitly in Section 4.

Such bounds do not give a tractable method to determine \( Y \). They only indicate that a stable estimation is possible. However, for the problem at hand, many very efficient methods exist. Some of them, like ESPRIT [18], do not need a discretization of \( Y \), as they estimate \( Y \) by an eigenvalue problem.

The results presented here are not directly related to the popular approach of using TV-minimization to solve the sparse frequency estimation problem. We are interested in proving intrinsic stability of the problem, independent of the solution method. On the other hand, TV-minimization is one particular technique. Unfortunately, it is not possible to directly reprove results from TV-minimization using the bounds presented here. The reason is that one cannot easily prove that TV-minimization yields measures satisfying our model assumptions. We refer to [8] for a good overview, including stability and convergence results.

## 3 Localizing Functions and Stability

We introduce some notation. Let
\[ S = \left\{ \sum_{y \in Y} c_y e^{2\pi i y} : c_y \in \mathbb{C} \setminus \{0\}, \ Y \subset [0, 1) \text{ finite} \right\} . \]
For an $f \in \mathcal{S}$, we call $Y^f$ the set of its frequencies and $c^f \in \mathbb{C}^{Y^f}$ the corresponding coefficients, which we enumerate by $Y^f$. We measure the distance of two frequencies $y, y' \in [0, 1)$ by the wrap-around distance

$$|y - y'|_T = \min_{k \in \mathbb{Z}} |y - y' - k|$$

and for a finite set $Y \subset [0, 1)$, we define its separation by

$$\text{sep} Y = \min_{y, y' \in Y \atop y \neq y'} |y - y'|_T$$

and collect for a $q > 0$ all well-separated exponential sums in

$$\mathcal{S}(q) = \{ f \in \mathcal{S} : \text{sep} Y^f \geq q \}.$$ 

As a first result, we prove bounds of the form

$$\|c^f\|_2^2 \lesssim B \sum_{k=A}^{B} |f(k)|^2 \lesssim \|c^f\|_2^2 \quad \forall f \in \mathcal{S}(q).$$

Such results are well-known, particular sharp lower bounds are proven by Moitra in \[17\] and by Aubel and Bölcskei in \[1\]. A sharp upper bound was already given by Selberg \[19\]. We use basically the same approach as Moitra, however improving it.

At the core of Moitra’s proof are particular functions, that were already constructed by Selberg. These functions $\psi_{A,B,q} \in L^1(\mathbb{R})$, $A, B \in \mathbb{Z}$, $A < B$, $q \in \mathbb{R}_{>0}$ satisfy the following properties:

(P1) $\psi_{A,B,q} \leq \chi_{[A,B]}$,

(P2) $\text{supp} \mathcal{F} \psi_{A,B,q} \subset [-q,q]$,

(P3) $\mathcal{F} \psi_{A,B,q}(0) = \int_{\mathbb{R}} \psi_{A,B,q}(x) \, dx = B - A - q^{-1}$.

Note that (P2) implies that $\psi_{A,B,q}$ is an entire function.

These functions $\psi_{A,B,q}$ are useful, because they allow to estimate a function localized in the spatial domain by something localized in the frequency domain, thus “cheating” the uncertainty principle. We call any function satisfying [P1] and [P2] $q$-localizing functions. For the special case $B - A \in q\mathbb{Z}$, the functions $\psi_{A,B,q}$ are extremal, in the sense that $\int_{\mathbb{R}} \chi_{[A,B]} - \psi_{A,B,q}$ is minimized. The extremal functions in the other cases are known as well, a result due to Littmann, see \[15\] Theorem 5.2.

The localizing function lie in the Paley-Wiener space

$$\text{PW} = \{ f \in L^2(\mathbb{R}) : \text{supp} \mathcal{F} f \subset [-1, 1] \}.$$ 

Furthermore, the following version of Poisson summation will prove useful, which follows from the more general Poisson formula given in \[20\] p. 69.

**Theorem 3.1 (Poisson Summation Formula).** For any $f \in \text{PW} \cap L^1(\mathbb{R})$ it holds true that

$$\sum_{k \in \mathbb{Z}} f(k) = \sum_{k \in \mathbb{Z}} \hat{f}(k).$$
With these ingredients, we can give an improved version of Moitra’s lower bound.

**Theorem 3.2.** Let \( f \in \mathcal{S}(q) \) and \( A, B \in \mathbb{Z}, A < B \). Then

\[
\left( B - A + 2 - \frac{1}{q} \right) \|c^f\|_2^2 \leq \sum_{k=A}^{B} |f(k)|^2 \leq \left( B - A + \frac{1}{q} \right).
\]

**Proof.** The upper bound is already due to Selberg, see [19]. For the lower bound, note that

\[
\sum_{k=A}^{B} |f(k)|^2 \geq \sum_{k \in \mathbb{Z}} \psi_{A-1,B+1,q}(k) |f(k)|^2 = \sum_{y,y' \in Y^f} c_y \overline{c}_{y'} \sum_{k \in \mathbb{Z}} \psi_{A-1,B+1,q}(k) e^{2\pi i (y-y')k} = \sum_{y,y' \in Y^f} c_y \overline{c}_{y'} \hat{\psi}_{A-1,B+1,q}(y-y') = \left( B - A + 2 - \frac{1}{q} \right) \|c^f\|_2^2,
\]

where we used Poisson summation and the fact that \( \psi_{A-1,B+1,q}(A-1) = \psi_{A-1,B+1,q}(B+1) \leq 0 \) (in fact, equality holds) due to (P1) and the continuity of \( \psi_{A-1,B+1,q} \). \( \square \)

**Remark.** Moitra used \( \psi_{A,B,q} \) instead of \( \psi_{A-1,B+1,q} \), resulting in the constant \( (B - A - q^{-1}) \). Aubel and Bölcskei improved the constant to \( (B - A + \frac{3}{2} - q^{-1}) \). Furthermore, they discuss the more general case of frequencies in the unit disc. The constant \( (B - A + 2 - \frac{1}{q}) \) is sharp in the following sense: For \( q = (B - A + 2)^{-1} \) the space \( \mathcal{S}(q) \) contains a linear subspace of dimension \( B - A + 2 \) and it is therefore clear that there is a \( f \in \mathcal{S}(q) \) vanishing on \( k = A, \ldots, B \). Thus, for \( q = (B - A + 2)^{-1} \) the best lower bound is zero.

The proof of Theorem 3.2 is worth a short reflection. A \( q \)-localizing function \( \psi \) gives rise to a sesquilinear form

\[
(\cdot, \cdot)_{\psi} : \mathcal{S} \times \mathcal{S} \to \mathbb{C}
\]

satisfying the following properties:

1. For two exponentials, we have

\[
(e^{2\pi iy}, e^{2\pi iy'})_{\psi} = \hat{\psi}(y-y').
\]

In particular, it is local in the sense that whenever \( |y-y'|_{T} \geq q \), we have that

\[
(e^{2\pi iy}, e^{2\pi iy'})_{\psi} = 0.
\]

2. It minorizes sampling at \( A, \ldots, B \), i.e., for all \( f \in \mathcal{S} \) we get

\[
(f, f)_{\psi} \leq \sum_{k=A}^{B} |f(k)|^2.
\]

One can use these properties to estimate other constellations of frequencies. If for an \( f \in \mathcal{S} \) we have that

\[
f = f_1 + \cdots + f_R
\]
such that all \( f_r \in S \) have mutually \( q \)-separated frequencies in the sense that
\[
\text{dist}(Y^{f_j}, Y^{f_k}) \geq q, \quad \text{for all } j \neq k,
\]
we obtain the lower bound
\[
\sum_{k=A}^{B} |f(k)|^2 \geq \sum_{r=1}^{R} (c^{fr})^* \left( \hat{\psi}(y - y') \right)_{y,y' \in Y^{f_r}} c^{fr} \geq \lambda_{\text{min}} ||c'||^2.
\]
Here, \( \lambda_{\text{min}} \) denotes the smallest eigenvalue of any of the \( R \) matrices \( \left( \hat{\psi}(y - y') \right)_{y,y' \in Y^{f_r}} \).

For example if every \( f_j \) corresponds to a cluster of up to \( L \) frequencies, we have to find a \( \psi \) such that
\[
\left( \hat{\psi}(y - y') \right)_{y,y' \in Y} \tag{4}
\]
is positive definite for any \( Y \subset [0,1) \), \( |Y| \leq L \) of interest, e.g., of a certain inner separation. Then we need to bound the smallest eigenvalue uniformly over all these sets.

For general values of \( L \) that seems to be a difficult problem. However, in the case \( L = 2 \), this is easy, as for \( Y = \{y,y'\} \) the matrix (4) has eigenvalues \( \hat{\psi}(0) \pm |\hat{\psi}(y - y')| \).

Clearly, it is necessary for \( \hat{\psi} \) to have a global maximum in zero. Then, an estimate of the form
\[
\hat{\psi}(0) - |\hat{\psi}(y - y')| \gtrsim |y - y'|^{1/2}
\]
gives rise to a very sharp bound. The localizing functions used above unfortunately do not have a global maximum at zero. We construct an alternative. A useful tool is a formula going back to Jageman and Fogel [12], which allows for Hermite interpolation on \( \mathbb{Z} \) in the Paley-Wiener space PW.

**Theorem 3.3.** For any \( f \in \text{PW} \) the following representation holds true:
\[
f(x) = \frac{\sin^2(\pi x)}{\pi^2} \sum_{k \in \mathbb{Z}} \left( \frac{f(k)}{(x-k)^2} + \frac{f'(k)}{x-k} \right).
\]

We are now able to construct a suitable localizing function.

**Lemma 3.4.** The function \( \phi \in \text{PW} \cap L^1(\mathbb{R}) \) defined by
\[
\phi(x) = \frac{\sin^2(\pi x)}{\pi^2} \left( \frac{2}{3} \left( \frac{1}{x} - \frac{1}{x-3} \right) + \frac{1}{(x-1)^2} + \frac{1}{(x-2)^2} \right)
\]
satisfies \( \phi \leq \chi_{[0,3]} \) and \( \hat{\phi}(0) = 2 \).

**Proof.** We start with \( \phi \leq \chi_{[0,3]} \). The claim obviously holds true for \( x \in \mathbb{Z} \), which we exclude in the following. Verifying \( \phi(x) \leq 0 \) for \( x \leq 0 \) is a direct calculation that we omit here. The case \( x \geq 3 \) follows by symmetry.

The inequality \( \phi(x) \leq 1 \) for \( x \in [0,3] \) is slightly more complicated. Note that
\[
\phi(x) \leq 1 = \frac{\sin^2(\pi x)}{\pi^2} \sum_{k \in \mathbb{Z}} \frac{1}{(x-k)^2},
\]

7
which is equivalent to
\[ \frac{2}{3} \left( \frac{1}{x} + \frac{1}{3 - x} \right) \leq \sum_{k \in \mathbb{Z}, k \neq 1/2} \frac{1}{(x - k)^2}. \] 

Now as the trapezoidal rule overestimates convex functions, we can estimate for \( K \in \mathbb{Z} \) and \( x < K \) that
\[ \sum_{k \geq K} \frac{1}{(x - k)^2} \geq \frac{1}{2} \left( \frac{1}{x - K} \right)^2 + \int_K^\infty \frac{1}{(x - y)^2} \, dy = \frac{1}{2} \left( \frac{1}{x - K} \right)^2 + \frac{1}{K - x} \]
and (6) follows by applying this twice. \( \hat{\phi}(0) = 2 \) results from Poisson summation. \( \square \)

**Remark.** A few remarks are in order.

1. It follows by Poisson summation that \( \phi \) maximizes \( \hat{\phi}(0) \) over all \( f \in PW \) satisfying \( f \leq \chi_{[0,3]} \). \( \hat{\phi}(0) \) is not unique, other values for \( \phi'(0) \) than \( \frac{2}{3} \) are possible (namely, any number between \( \left[ \frac{2}{3}, \frac{4}{3} \right] \)). Similar constructions work for any minorants of \( \chi_{[A,B]} \), \( A,B \in \mathbb{Z} \). In particular, it follows that if \( A + 1 = B \), the constant zero minorant is optimal.

2. The non-uniqueness is certainly already known, however difficult to track down. For example, in [10, p. 289] a similar result is cited for the majorants and attributed to unpublished work of Selberg.

3. The construction originally used by Selberg and by Moitra corresponds to choosing \( \phi'(0) = 1 \).

(5) is now the result of a direct calculation.

**Lemma 3.5.** \( \phi \) satisfies
\[ \hat{\phi}(0) - |\hat{\phi}(w)| \geq \begin{cases} \pi^2 w^2, & \text{for } |w| \in [0, \frac{1}{3}] \\ \pi^2/9, & \text{for } |w| \in [\frac{1}{3}, 1] \end{cases}. \]

**Proof.** The Fourier transform of \( \phi \) is given by
\[ \hat{\phi}(w) = (1 - w) \left( e^{-2\pi i w} + e^{-4\pi i w} \right) + \frac{1}{3\pi i} \left( 1 - e^{-6\pi i w} \right), \quad \text{for } w \in [0, 1] \]
and \( \hat{\phi}(w) = \overline{\hat{\phi}(-w)} \) for \( w \in [-1, 0] \).

We estimate \( |\hat{\phi}| \) by first noting that for \( w \in [0, \frac{1}{3}] \) we have
\[ |\hat{\phi}(w)| = \left| e^{3\pi i w} \hat{\phi}(w) \right| = \left| (1 - w) \left( e^{\pi i w} + e^{-\pi i w} \right) + \frac{1}{3\pi i} \left( e^{3\pi i w} - e^{-3\pi i w} \right) \right| \]
\[ = (1 - w)2\cos(\pi w) + \frac{2}{3\pi} \sin(3\pi w). \]

Next, we use
\[ \cos(x) \geq 1 - \frac{x^2}{2} \quad \text{and} \quad \sin(x) \geq x - \frac{x^3}{6} \quad \forall x \in \mathbb{R}_{\geq 0} \]
and obtain 
\[ |\hat{\phi}(w)| \geq (1 - w) \left( 2 - \pi^2 w^2 \right) + 2w - 3\pi^2 w^3 = 2 - \pi^2 w^2 - 2\pi^2 w^3. \]

This results in 
\[ \hat{\phi}(0) - |\hat{\phi}(w)| \geq \pi^2 w^2 \quad \text{for } w \in \left[ 0, \frac{1}{3} \right]. \]

It is not difficult to show that \( \hat{\phi}(0) - |\hat{\phi}(w)| \geq \pi^2/9 \) for \( w \in \left[ \frac{1}{3}, 1 \right] \). We omit the computation here.

Now we are able to state the main theorem.

**Theorem 3.6.** Let \( N \in \mathbb{N} \) be given. Further, let \( f \in \mathcal{S} \) be such that \( Y_f \) can be decomposed in three disjoint sets \( Y_1, Y_2, Y_3 \) satisfying:

1. For all \( y, y' \in Y_j \) with \( y \neq y' \) we have that \( |y - y'|_T \geq \frac{3}{N+1} \).
2. For all \( y \in Y_1 \) there is exactly one \( y' \in Y_2 \) with \( |y - y'|_T < \frac{3}{N+1} \). We denote \( y' \) by \( n(y) \).
3. For all \( y \in Y_3 \) and all \( y' \in Y_1 \cup Y_2 \) we have that \( |y - y'|_T \geq \frac{3}{N+1} \).

Then the following bound holds true:

\[
\frac{N}{2} |f(k)|^2 \geq \frac{2}{3} (N + 1) \left( \sum_{y \in Y_1} |c_y|^2 + \sum_{y \in Y_1} \left[ \frac{N + 1}{3} \left| e^{-\pi \imath y} c_y + e^{-\pi \imath \pi n(y)} c_n(y) \right|^2 
\right.
\left. + \frac{\pi^2 (N + 1)^2}{2 \cdot 3^5} |y - n(y)|^2 \left| e^{-\pi \imath y} c_y - e^{-\pi \imath \pi n(y)} c_n(y) \right|^2 \right].
\]

Proof. We dilate \( \phi \) to fit our data. Let 
\[
\phi_N(x) = \phi \left( \frac{3x}{N + 1} \right) \leq \chi_{[0,N+1]}(x), \quad \hat{\phi}_N(w) = \frac{N + 1}{3} \hat{\phi} \left( \frac{(N + 1)w}{3} \right).
\]

Note that \( \text{supp} \hat{\phi}_N \subset \left[ -\frac{3}{N+1}, \frac{3}{N+1} \right] \). Again, we use \( \phi_N \) to localize the sum:

\[
\sum_{k=1}^{N} |f(k)|^2 \geq \sum_{k \in \mathbb{Z}} \phi_N(k) |f(k)|^2 = \sum_{y, y' \in Y'} c_y c_{y'} \hat{\phi}_N(y - y') = \hat{\phi}_N(0) \sum_{y \in Y_3} |c_y|^2 + \sum_{y \in Y_1} c_y^* A_y c_y
\]

with 
\[
c_y = \begin{pmatrix} c_y \\ c_n(y) \end{pmatrix} \quad \text{and} \quad A_y = \begin{pmatrix} \hat{\phi}_N(0) & \hat{\phi}_N(n(y) - y) \\ \hat{\phi}_N(n(y) - y) & \hat{\phi}_N(0) \end{pmatrix}.
\]

Next, we consider the eigenvalue decomposition of \( A_y \). We have the two eigenvectors 
\[
v_1 = \frac{1}{\sqrt{2}} \left( 1 \quad e^{-\imath \arg \hat{\phi}_N(n(y) - y)} \right) \quad \text{and} \quad v_2 = \frac{1}{\sqrt{2}} \left( e^{-\imath \arg \hat{\phi}_N(n(y) - y)} \quad 1 \right).
\]
Letting $\theta_y = \text{Arg} \hat{\phi}_N(n(y) - y)) = 3\pi i (n(y) - y)$, we get
\[
2c_y^* A_y c_y = \left( \hat{\phi}_N(0) + |\hat{\phi}_N(n(y) - y)| \right) |c_y + e^{-i\theta_y} c_{n(y)}|^2 + \left( \hat{\phi}_N(0) - |\hat{\phi}_N(n(y) - y)| \right) |c_y - e^{-i\theta_y} c_{n(y)}|^2
\]
Plugging that into (7) and using Lemma 3.6 we obtain
\[
\sum_{k=1}^{N} |f(k)|^2 \geq \frac{2}{3}(N + 1) \sum_{y \in Y_3} |c_y|^2 + \sum_{y \in Y_3} \left[ \frac{N + 1}{3} |c_y + e^{-i\theta_y} c_{n(y)}|^2 + \frac{\pi^2(N + 1)^3}{54} \min \{ (N + 1)^2 |y - n(y)|^2, 1 \} |c_y - e^{-i\theta_y} c_{n(y)}|^2 \right].
\]
Using $|y - n(y)| < \frac{3}{N+1}$ gives the result. \(\square\)

As the bound presented in Theorem 3.6 is rather technical, we discuss it and give a few special cases and variations. Note that we did not try to optimize the constants. Slightly sharper results are certainly possible.

**Sharpness:** The estimate is sharp for $|y - n(y)| \to 0$ up to constants. To see that, we consider
\[
f_{\tau}(x) = 1 - e^{2\pi i \tau x}.
\]
A direct estimate gives for $|\tau| \ll N^{-1}$
\[
\sum_{k=1}^{N} |f(k)|^2 \sim \sum_{k=1}^{N} |2\pi i \tau k|^2 \sim \tau^2 N^3,
\]
which is the exact asymptotic Theorem 3.6 gives as well. Clearly, for $\|c\|_2 \to 0$ the estimate is sharp as well.

The estimate gives the expected result for the signs of the coefficients as well. Namely, if $c_y \approx -c_{n(y)}$ the critical cancellation occurs, as the first term of the second sum vanishes. If one is interested in an estimate independent of $\text{Arg} c$, one obtains
\[
\sum_{k=1}^{N} |f(k)|^2 \geq \frac{2}{3}(N + 1) \sum_{y \in Y_3} |c_y|^2 + \frac{\pi^2(N + 1)^3}{2 \cdot 3^5} \sum_{y \in Y_3} |y - n(y)|^2 (|c_y|^2 + |c_{n(y)}|^2). \quad (8)
\]

**Symmetric Samples:** Quite often, in particular when $N$ is odd, one is interested in estimating
\[
\sum_{k=1}^{N} \left| f \left( k - \frac{N + 1}{2} \right) \right|^2.
\]
Due to symmetry, one obtains the estimate
\[
\sum_{k=1}^{N} \left| f \left( k - \frac{N + 1}{2} \right) \right|^2 \geq \frac{2}{3}(N + 1) \sum_{y \in Y_3} |c_y|^2 + \sum_{y \in Y_3} \left[ \frac{N + 1}{3} |c_y + c_{n(y)}|^2 + \frac{\pi^2(N + 1)^3}{2 \cdot 3^5} |y - n(y)|^2 |c_y - c_{n(y)}|^2 \right]. \quad (9)
\]
Here, we do not need to modulate the coefficients $c_{n(y)}$ and $c_y$. 

10
Choice of localizing function: Different choices for $\phi$ are possible. Assume we pick a 1-localizing function $\phi \leq \chi_{[0,A]}$. Then, after dilation, we can localize up to $A/(N+1)$. As discussed earlier, $A > 1$ is necessary. Using the extremal construction outlined above for $A = 2$ instead of $A = 3$, gives worse rates, as it can be shown that it only satisfies

$$\hat{\phi}(0) - |\phi(w)| \gtrsim |w|^3.$$ 

Note that we cannot hope to obtain a bound as in Theorem 3.6, if $q < \frac{2}{N+1}$, as then counterexamples exist: $(N+1)/2$ pairs of very close frequencies would result in $|Y^f| = N + 1$ and such an exponential sum can vanish on $N$ points.

In [6] the function $\phi \leq \chi_{[-2,2]}$ was used, covering the symmetric case. Due to that, the assumptions on $q$ are a slightly stronger there.

4 Singular Value Estimates of Vandermonde Matrices

Both Theorem 3.2 and Theorem 3.6 directly give estimates of the smallest singular values of Vandermonde matrices. For the reader’s convenience, we state them here explicitly.

We denote the Vandermonde matrix with frequencies $y_1, \ldots, y_M \in [0,1)$ by

$$V_N(y_1, \ldots, y_M) := \begin{pmatrix}
1 & e^{2\pi i y_1} & \cdots & e^{2\pi i y_M} \\
1 & e^{2\pi i y_2} & \cdots & e^{2\pi i y_M} \\
\vdots & \vdots & \ddots & \vdots \\
e^{2\pi i (N-1)y_1} & e^{2\pi i (N-1)y_2} & \cdots & e^{2\pi i (N-1)y_M}
\end{pmatrix} \in \mathbb{C}^{N \times M}.
$$

Using

$$V_N(y_1, \ldots, y_M)c = \left(\sum_{j=1}^{M} c_j e^{2\pi i y_j k}\right)_{k=0,\ldots,N-1},$$

Theorem 3.2 gives the following:

**Corollary 4.1.** Let $y_1, \ldots, y_M \in [0,1)$ be $q$-separated frequencies. Then the smallest singular value $\sigma_{\min}$ of $V_N(y_1, \ldots, y_M)$ satisfies

$$\sigma_{\min}^2 \geq \left( N + 1 - \frac{1}{q} \right).$$

Analogously, [8] gives rise to the following result for Vandermonde matrices with pairwise colliding nodes.

**Corollary 4.2.** Let $Y = \{y_1, \ldots, y_M\} \subset [0,1)$ and $Y' = \{y'_1, \ldots, y'_{M'}\} \subset [0,1)$ be two $q$-separated sets of frequencies, where $q \geq \frac{3}{N+1}$. Assume further that for each $y \in Y$ there is at most one $y' \in Y'$ such that $|y - y'| \leq q$. Let $\tau$ be the smallest of these distances, i.e., $\tau = \text{dist}_{T}(Y,Y')$. Then the smallest singular value $\sigma_{\min}$ of $V_N(y_1, \ldots, y_M, y'_1, \ldots, y'_{M'})$ satisfies

$$\sigma_{\min}^2 \geq \frac{\pi^2}{2 \cdot 3^5 (N + 1)^3} \tau^2.$$
assuming that \( \tau < \frac{3}{N+1} \) and
\[
\sigma_{\min}^2 \geq \frac{2}{3} (N+1)
\]
on-otherwise.

The corollary is a significant improvement over similar estimates, very recently presented in [13]. Bounding singular values of Vandermonde matrices with multiple tightly clustered sets of frequencies has attracted some attention recently, as it gives insights when recovery of frequencies even without separation might be possible. Corollary 4.2 is (up to small improvements in the constants) sharp for the case of clusters of two nodes. For results covering clusters of multiple nodes, see [2,14].

Finally, note that Theorem 3.6 gives some information on the geometry of the singular spaces associated with small singular values. Indeed, it confirms the intuition that close frequencies should have coefficients that sum to zero.

5 Well-posedness and A Posteriori Error Estimates

Next, we apply Theorem 3.6 to obtain conditional well-posedness of the frequency estimation problem. To simplify notation, we stick to the case of symmetric samples \( f(k), k = -N, \ldots, N \).

Assume that we are given \( f,g \in S(2q) \). Then for every \( y \in Y^f \) there is at most one \( y' \in Y^g \) with \( |y - y'|_\tau < q \). Therefore, we are in position to apply Theorem 3.6.

**Corollary 5.1.** Let \( f,g \in S(2q) \) and \( N \in \mathbb{N}_{>0} \) with \( q \geq \frac{3}{2N+2} \) be given. Assume further that all coefficients of \( f \) and \( g \) have a modulus of at least \( c_{\text{min}} \in \mathbb{R}_{>0} \). If
\[
\sum_{k=-N}^{N} |f(k) - g(k)|^2 < \frac{4N + 4c_{\text{min}}^2}{3},
\]
then for every \( y \in Y^f \) there is exactly one \( n(y) \in Y^g \) with \( |y - n(y)|_\tau < \frac{3}{2N+2} \).

Furthermore, the following estimate holds true:
\[
\sum_{y \in Y^f} \left[ \frac{N+1}{3} |c_y - c_{n(y)}|^2 + \frac{2c_{\text{min}}^2(N+1)^3}{35} |y - n(y)|^2 |c_y + c_{n(y)}|^2 \right] \leq \sum_{k=-N}^{N} |f(k) - g(k)|^2.
\]

**Proof.** We invoke [9] to see that \( Y_3 = \emptyset \), i.e., for all \( y \in Y^f \) there has to be a \( n(y) \). The claim follows by using that \( |y - n(y)|_\tau < \frac{3}{2N+2} \). \( \square \)

That estimate can now easily be used to obtain a posteriori error estimates. Assume that we are given noisy samples \( \hat{f}(k) = f(k) + \eta_k \) and use any algorithm we want to obtain a candidate \( g \). Now, if both fit our model, i.e., \( f,g \in S(2q) \), Corollary 5.1 applies. However, as we do not know \( f(k) \), we cannot use Corollary 5.1 directly. Instead, we combining it with a noise model to get an a posteriori estimate.

As the noise model depends on the application, we give a prototypical result, using Gaussian noise.

**Lemma 5.2.** Let \( \mathbf{v} \in \mathbb{C}^K \) and \( \eta = (\eta_k)_k \in \mathbb{C}^K \), with
\[
\eta_k = X_{k,1} + iX_{k,2},
\]
then
\[
\sum_{k=-N}^{N} |f_k - g_k|^2 \leq \sum_{k=-N}^{N} |\hat{f}_k|^2.
\]
where $X_{k,j}$ are independent Gaussian random variables with mean zero and variance $\sigma^2$. Let $\delta \in (0, 1)$. Then with probability
\[ 1 - e^{-K^{(1+\delta)/2}} - 2e^{-K^\delta/8} \]
it holds true that
\[ \|v\|_2 \leq \left\| v + \eta \right\|_2^2 - 2K\sigma^2 \left( 1 + \sqrt{2} \right)\sigma K^{(1+\delta)/4}. \]

Proof. Let
\[ \|v + \eta\|_2^2 = \|v\|_2^2 + \sigma^2 Y + 2\sigma \|v\|_2 Z, \]
where $Z$ is a standard Gaussian random variable and $Y$ is a $\chi^2$-distributed with $2K$ degrees of freedom. (5) is equivalent to
\[ \|v + \eta\|_2^2 + \sigma^2 Z - \sigma^2 Y = (\|v\|_2^2 + \sigma Z)^2. \]
We estimate
\[ \|v\|_2 \leq \left( \sigma^2 Z^2 - \sigma^2 Y + \|v + \eta\|_2^2 \right)^{1/2} + \sigma |Z| \]
\[ \leq \left| \|v + \eta\|_2^2 - 2K\sigma^2 \right|^{1/2} + 2\sigma |Z| + \sigma |Y - 2K|^{1/2}, \]
where we used the subadditivity of the square root.

Next, we use two tail estimates. For $Z$ we have that
\[ \Pr(|Z| \geq t_1) \leq e^{-\frac{t_1^2}{2}} \quad \text{for all } t > 0, \]
see for example [9], Proposition 7.5. And $Y$ is concentrated around $\mathbb{E}Y = 2K$:
\[ \Pr(|Y - 2K| \geq 2Kt_2) \leq 2e^{-Kt_2^2/4} \quad \text{for all } t_2 \in (0, 1), \]
see [3], Theorem 2.57. Now we choose
\[ t_1 = K^{(1+\delta)/4}, \quad t_2 = \frac{K^{(\delta-1)/2}}{\sqrt{2}} \]
and by the union bound we obtain with probability
\[ 1 - e^{-K^{(1+\delta)/2}} - 2e^{-K^\delta/8} \]
that
\[ \|v\|_2 \leq \left| \|v + \eta\|_2^2 - 2K\sigma^2 \right|^{1/2} + (2 + \sqrt{2})\sigma K^{(1+\delta)/4}. \]
\[ \Box \]

Combining these results, we obtain the following error estimate.

**Corollary 5.3.** Assume that $f, g \in S(2q)$ with $q \geq \frac{3}{2N+2}$. Let
\[ f = (f(k))_{k=-N,...,N} \in \mathbb{C}^{2N+1}, \quad g = (g(k))_{k=-N,...,N} \in \mathbb{C}^{2N+1}, \quad \eta = (\eta(k))_{k=-N,...,N} \in \mathbb{C}^{2N+1}. \]
Here,
\[ \eta_k = X_{k,1} + iX_{k,2}, \]

13
where $X_{k,j}$ are independent Gaussian random variables with mean zero and variance $\sigma^2$. Let $\delta \in (0, 1)$. Then with probability

$$1 - e^{-(2N+1)(1+\delta)/2} - 2e^{-(2N+1)^\delta/8}$$

it holds true that if

$$\left( \| f + \eta - g \|^2_2 - 2(2N + 1)\sigma^2 \right)^{1/2} + (2 + \sqrt{2})\sigma(2N + 1)^{(1+\delta)/4}$$

we have that for every $y \in Y^f$ there is exactly one $n(y) \in Y^g$ with $\| y - n(y) \|_T < \frac{3}{2N+2}$.

Furthermore, the following estimate holds true:

$$\sum_{y \in Y^f} \left[ \frac{N + 1}{3} | c_y - c_{n(y)} |^2 + \frac{2\pi^2(N + 1)^3}{3^3} | y - n(y) |^2 \right] \leq \left( \| f + \eta - g \|^2_2 - 2(2N + 1)\sigma^2 \right)^{1/2} + (2 + \sqrt{2})\sigma(2N + 1)^{(1+\delta)/4}$$

(10)

Remark. The case of bounded instead of Gaussian noise is easier. We skip the details.

We close this paper by giving a small example, how the bound can be applied.

Example 5.4. We consider the exponential sum $f$ with frequencies

$$Y^f = \{ 0.1, 0.3, 0.6, 0.9 \}, \quad c = [1.1, -1.1, 2, 2],$$

where we enumerate the frequencies in order of their size. Now we sample $f$ at $-20, \ldots, 20$ and add Gaussian noise as described in Corollary 5.3 of variance $\sigma^2$. Next, we apply ESPRIT to obtain an estimate $\tilde{f}$. We calculate the left-hand side of (10), called error, and the right-hand side, called error estimator. Further, we calculate the sampling distance, which is $\| f - \tilde{f} \|^2_2$. The results for different choices of $\sigma$ are presented in Figure 7. For each choice of $\sigma$, fifty instances were calculated and the largest error is shown.

We pick $\delta = 0.9$, resulting in a probability of a little bit more than 94%, that the estimate can be applied. For $\sigma = 1$ the premise of Corollary 5.3 is actually not satisfied. We can observe in Figure 7 that the given bound seems to be reasonably sharp.

Acknowledgments

The author was supported by the Bavarian Ministry of Economy by means of the research project “Big Picture”.

The results of the paper are extensions of results obtained in the author’s phd thesis [6], written under supervision of Prof. Armin Iske. The author would like to thank Prof. Iske for his constant support and encouragements. Furthermore, the author is indebted to Prof. Tomas Sauer for helpful remarks on the manuscript.
Figure 1: Comparison of the error and the error estimate at different noise levels. x-axis: Standard deviation of the noise.

References

[1] Céline Aubel and Helmut Bölcskei. Vandermonde matrices with nodes in the unit disk and the large sieve. *Applied and Computational Harmonic Analysis*, 2017.

[2] Dmitry Batenkov, Laurent Demanet, Gil Goldman, and Yosef Yomdin. Stability of partial fourier matrices with clustered nodes. *arXiv preprint arXiv:1809.00658*, 2018.

[3] Bernard Bercu, Bernard Delyon, and Emmanuel Rio. *Concentration inequalities for sums and martingales*. Springer, 2015.

[4] Emmanuel J Candès and Carlos Fernandez-Granda. Super-resolution from noisy data. *Journal of Fourier Analysis and Applications*, 19(6):1229–1254, 2013.

[5] Emmanuel J Candès and Carlos Fernandez-Granda. Towards a mathematical theory of super-resolution. *Communications on Pure and Applied Mathematics*, 67(6):906–956, 2014.

[6] Benedikt Diederichs. *Sparse Frequency Estimation: Stability and Algorithms*. PhD thesis, Universität Hamburg, 2018.

[7] Marco F Duarte and Richard G Baraniuk. Spectral compressive sensing. *Applied and Computational Harmonic Analysis*, 35(1):111–129, 2013.

[8] Vincent Duval and Gabriel Peyré. Exact support recovery for sparse spikes deconvolution. *Foundations of Computational Mathematics*, 15(5):1315–1355, 2015.

[9] Simon Foucart and Holger Rauhut. *A mathematical introduction to compressive sensing*. Springer, 2013.

[10] Sidney West Graham and Jeffrey D Vaaler. A class of extremal functions for the fourier transform. *Transactions of the American Mathematical Society*, 265(1):283–302, 1981.
[11] Lei Hu, Zhiguang Shi, Jianxiong Zhou, and Qiang Fu. Compressed sensing of complex sinusoids: An approach based on dictionary refinement. *IEEE Transactions on Signal Processing*, 60(7):3809–3822, 2012.

[12] D Jagerman and L Fogel. Some general aspects of the sampling theorem. *IRE transactions on information theory*, 2(4):139–146, 1956.

[13] Stefan Kunis and Dominik Nagel. On the condition number of vandermonde matrices with pairs of nearly-colliding nodes. *arXiv preprint arXiv:1812.08645*, 2018.

[14] Weilin Li and Wenjing Liao. Stable super-resolution limit and smallest singular value of restricted fourier matrices. *arXiv preprint arXiv:1709.03146*, 2017.

[15] Friedrich Littmann. Quadrature and extremal bandlimited functions. *SIAM Journal on Mathematical Analysis*, 45(2):732–747, 2013.

[16] Dimitris G Manolakis, Vinay K Ingle, and Stephen M Kogon. *Statistical and adaptive signal processing: spectral estimation, signal modeling, adaptive filtering, and array processing*. McGraw-Hill Boston, 2000.

[17] Ankur Moitra. Super-resolution, extremal functions and the condition number of vandermonde matrices. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 821–830. ACM, 2015.

[18] Robert Roy, Arogyaswami Paulraj, and Thomas Kailath. Esprit—a subspace rotation approach to estimation of parameters of cisoids in noise. *IEEE transactions on acoustics, speech, and signal processing*, 34(5):1340–1342, 1986.

[19] Jeffrey D Vaaler. Some extremal functions in fourier analysis. *Bulletin of the American Mathematical Society*, 12(2):183–216, 1985.

[20] Antoni Zygmund. *Trigonometric series: Vol. II, 2nd edn. (reprinted with corrections and some additions)*. Cambridge University Press, London/New York, 1968.