A Hybrid Scattering Transform for Signals with Isolated Singularities

Michael Perlmutter
Dept. of Mathematics
University of California, Los Angeles
Los Angeles, CA, USA
perlmutter@math.ucla.edu

Jieqian He
Dept. of CMSE
Michigan State University
East Lansing, MI, USA
hejieqia@msu.edu

Mark Iwen
Dept. of Mathematics
Dept. of CMSE
Michigan State University
East Lansing, MI, USA
iwenmark@msu.edu

Matthew Hirn
Dept. of CMSE
Dept. of Mathematics
Michigan State University
East Lansing, MI, USA
mhirn@msu.edu

***Abstract*—The scattering transform is a wavelet-based model of Convolutional Neural Networks originally introduced by S. Mallat. Mallat's analysis shows that this network has desirable stability and invariance guarantees and therefore helps explain the observation that the filters learned by early layers of a Convolutional Neural Network typically resemble wavelets. Our aim is to understand what sort of filters should be used in the later layers of the network. Towards this end, we propose a two-layer hybrid scattering transform. In our first layer, we convolve the input signal with a wavelet filter transform to promote sparsity, and, in the second layer, we convolve with a Gabor filter to leverage the sparsity created by the first layer. We show that these measurements characterize information about signals with isolated singularities. We also show that the Gabor measurements used in the second layer can be used to synthesize sparse signals such as those produced by the first layer.**

***Index Terms*—scattering transforms, wavelets, sparsity, deep learning, time-frequency analysis**

I. Introduction

The wavelet scattering transform is a mathematical model of Convolutional Neural Networks (CNNs) introduced by S. Mallat [3]. Analogously to the feed-forward portion of a CNN, it produces a latent representation of an input signal via an alternating sequence of filter convolutions and nonlinearities. It differs, most notably, by using predesigned wavelet filters rather than filters learned from data.

Using predefined filters allows for rigorous analysis and helps us understand why a deep nonlinear network is better than a wide, shallow, linear network with the same number of parameters. Ideally, a feed-forward network should produce a representation which is sufficiently descriptive for downstream tasks, but also stable to deformations such as translations. Linear networks are typically unable to do both and often must discard high-frequency information to achieve stability. Mallat's analysis in [3] shows that the scattering transform, on the other hand, captures high-frequency information via wavelets and then pushes it down to lower, more stable, frequencies using a nonlinear activation function. Thus, the nonlinear structure enables the network to stably capture high-frequency information.

J.H. was supported by the National Science Foundation (grant #1845856). M.I. was supported in part by NSF DMS 1912706. M.H. acknowledges funding from the National Institutes of Health (grant #R01GM135929), the National Science Foundation (CAREER award #1845856), and the Department of Energy (grant #DE-SC0021152).

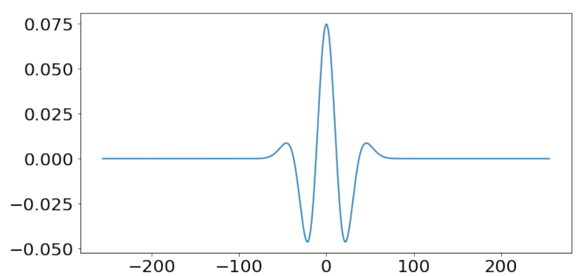


Fig. 1: Wavelet filter used in the first layer

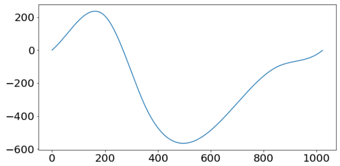
The scattering transform also helps us understand which filters are useful for effectively encoding information. While the optimal choice is task dependent, wavelets are often a good choice since natural images are typically sparse in the wavelet basis and as discussed above, they are able to capture high-frequency information. Moreover, and perhaps most importantly, the filters learned in the early layers of CNNs typically resemble wavelets.

This paper focuses on the choice of filters for later layers of the network. In particular, we propose a two-layer hybrid scattering model. In the first layer, we use a wavelet convolution to sparsify the input. Then, we use a Gabor type filter to leverage this sparsity.

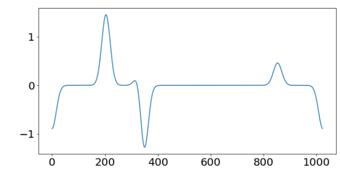
For simplicity, we assume that the input $y(t)$ is a piecewise polynomial whose knots are located at points $\{u_i\}_{i=1}^k$ $u_i < u_{i+1}$. We shall also assume that each of its piecewise components $y_i(t)$ has degree at most m. We let ψ be a mother wavelet with $\text{supp}(\psi) \subseteq [-1, 1]$ and let

$$\psi_\ell(t) = \frac{1}{2^\ell}\psi\left(\frac{t}{2^\ell}\right).$$

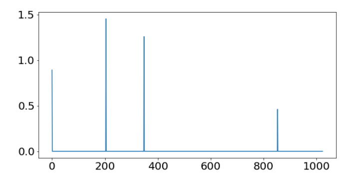
We will assume that ψ has $m+1$ vanishing moments, which implies that $\psi_\ell \star y_i(t) = 0$ (see e.g. [2]). It follows that $\text{supp}(\psi_\ell \star y)$ is contained in $\cup_{i=1}^k [u_i - 2^\ell, u_i + 2^\ell]$. To further



(a) piecewise polynomial $y(t)$

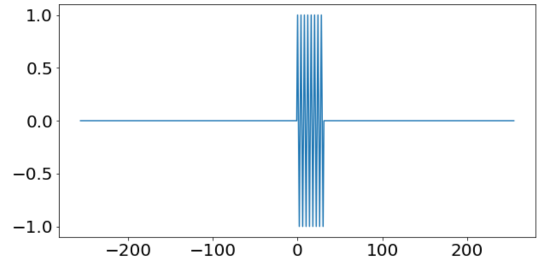


(b) Convolution against wavelet ψ_ℓ

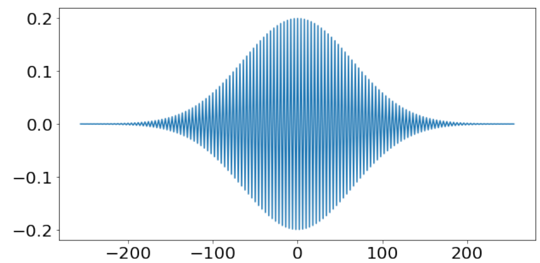


(c) Sparse Signal from Max Pooling

Fig. 2: Wavelets sparsify piecewise polynomials on the interval $[0, 1024h]$.



(a) Indicator function window



(b) Gaussian window

Fig. 3: Gabor filters used in the second layer (real parts)

promote sparsity, we next apply a max-pooling operator:

$$MP_\ell z(t) = \begin{cases} z(t) & \text{if } z(t) = \max_{t' \in [t_i - 2^\ell, t_i + 2^\ell] \cap h\mathbb{Z}} z(t') \\ 0 & \text{otherwise} \end{cases}.$$

As summarized in the following theorem, this yields a linear combination of Dirac delta functions.

Theorem 1. *Assume that* $2^{\ell+1} \leq \min_{i \neq i'} |u_i - u_{i'}|$. *Then,*

$$MP_\ell(|\psi_\ell \star y|)(t) = \sum_{j=1}^{k} a_j \delta_{v_j}(t).$$

for some $a_1, \ldots, a_k > 0$, $v_j \in [u_j - 2^\ell, u_j + 2^\ell], 1 \leq j \leq k$.

In our second layer, rather than another wavelet, we use a Gabor filter

$$g_{s,\xi} = w\left(\frac{t}{s}\right) \mathbb{e}^{\mathbb{i}\xi t}, \tag{1}$$

where the parameters s and ξ determine the scale and central frequency and the window function w is supported on an interval of unit length. Next, we take the L^p norm for some integer $p \geq 1$. As a result, we obtain translation invariant *hybrid scattering coefficients*

$$\|g_{s,\xi} \star MP_\ell(|\psi_\ell \star y|)\|_p.$$

By design, these measurements are invariant to translations, reflections, and global sign changes. We aim to investigate the ability of our measurements to characterize y up to these natural ambiguities. The wavelet-modulus is known to be a powerful signal descriptor [4]. Therefore, in light of Theorem 1, we shall analyze the ability of the measurements

$$f_\xi(s)[x] := \|g_{s,\xi} \star x\|_p \tag{2}$$

to characterize signals of the form

$$x(t) = \sum_{j=1}^{k} a_j \delta_{v_j}(t). \tag{3}$$

For such a signal, we will let $\vec{a}$ be the vector] defined by $\vec{a} = (a_1, \ldots, a_k)$ and let $\|\vec{a}\|_p$ denote its ℓ^p norm.

To supplement our theory, we will show that the measurements (2) can be used to reconstruct a sparse signal of the form (3) up to translations, reflections and global sign changes in Section VI.

We will show that our measurements characterize the support set $\{v_j\}_{j=1}^k$. For $i < j$, we let $\Delta_{i,j} = v_j - v_i$ and consider the difference set

$$\mathcal{D}(x) := \{\Delta_{i,j} : 1 \leq i < j \leq k\}.$$

We will assume that x is *collision free*, i.e., that $\Delta_{i,j} \neq \Delta_{i',j'}$ except for when $(i, j) = (i', j')$ and that $\mathcal{D}(x)$ is contained in a fine grid, $h\mathbb{Z} = \{hn : n \in \mathbb{Z}\}$ for some $h > 0$. Under these assumptions, it is known [1], [5] that the support set $\{v_j\}_{j=1}^k$ is determined (up to reflection and translation) by $\mathcal{D}(x)$ except for in the case where $k = 6$ and the $\{v_j\}_{j=1}^6$ belong to a specific parametric family. (See Theorem 1 of [5] for full details. For the remainder of this work, we will assume that $\{v_j\}_{j=1}^k$ does not belong to this family and therefore the support set $\{v_j\}_{j=1}^k$ is determined by $\mathcal{D}(x)$.) This motivates the following theorem which shows that the measurements (2) uniquely determine $\mathcal{D}(x)$.

Theorem 2. *Let $p \geq 1$ be an integer and let $w(t) = 1_{[0,1]}(t)$. Then for almost every ξ, the function*

$$f_\xi(s) = \|g_{s,\xi} \star x\|_p$$

is piecewise linear. Morover, the set of its isolated singularities is exactly the support set $\mathcal{D}(x)$.

Theorem 4 shows that selecting a single random frequency and enough scales $\{s_j\}$ such that there is one s_j in between each element of $\mathcal{D}(x)$ allows us detect the location of each point of $\mathcal{D}(x)$ by evaluating $f_\xi(s)$ at each of the s_j (up to a precision corresponding to the density of the scales). The next result shows that the amplitudes a_j can also be recovered with $\mathcal{O}(p)$ randomly chosen frequencies. Thus, the measurements (2) characterize sparse signals up to natural ambiguities.

Theorem 3. *Let $w(t) = 1_{[0,1]}(t)$ and, let*

$$x(t) = \sum_{j=1}^{k} a_j \delta_{v_j}(t)$$

be a sparse signal of the form (3). Let $\xi_1, \ldots, \xi_L$ be i.i.d. standard normal random variables, where L is assumed to be at least $p+2$ if p is even and at least $4p+2$ if p is odd. Then the following uniqueness result holds almost surely:

Let

$$\widetilde{x}(t) = \sum_{j=1}^{k} \widetilde{a}_j \delta_{\widetilde{v}_j}(t).$$

Suppose that $\|\vec{a}\|_p = \|\vec{\widetilde{a}}\|_p$, that $\mathcal{D}(\widetilde{x}) = \mathcal{D}(x)$, and that

$$\partial_s^2 f_{\xi_\ell}[x](d) = \partial_s^2 f_{\xi_\ell}[\widetilde{x}](d)$$

for all $d \in \mathcal{D}(x)$ and all $1 \leq \ell \leq 4p+2$.

Then we have that $\vec{a} = \pm\vec{\widetilde{a}}$, and therefore $\widetilde{x}(t)$ is equivalent to $x(t)$ up to translation, reflection, and global sign change.

II. Generalized Exponential Polynomials

In this section, we will introduce some notation and state some lemmas that are needed in order to prove Theorems 2 and 3. For the proof of the lemmas in this section, please see section V.

We let $\mathcal{E}$ denote the set of functions that can be written as

$$p(\theta) = \sum_{k=1}^{N} \alpha_k \mathrm{e}^{\mathrm{i}\gamma_k \theta} \tag{4}$$

where $N \geq 1$, $\alpha_k, \gamma_k \in \mathbb{R}$, $\alpha_k \neq 0$, and $\gamma_1 < \gamma_2 < \ldots < \gamma_N$. Since the γ_k are allowed to be arbitrary (possibly negative or irrational) real numbers, we call these functions *generalized exponential polynomials*. For $p \in \mathcal{E}$, we refer to γ_N as the degree of p. We let $\mathcal{E}(d)$ refer to the set of all $p \in \mathcal{E}$ with degree$(p) = d$, and let $\mathcal{E}_0(d)$ denote the set of $p \in \mathcal{E}(d)$ such that $\gamma_1 = 0$.

The following lemma shows that each $p \in \mathcal{E}$ has a unique representation as the sum of exponentials, and that therefore, the degree of p is well defined.

Lemma 1. *Let $p, q \in \mathcal{E}$, with*

$$p(\theta) = \sum_{k=1}^{N} \alpha_k \mathrm{e}^{\mathrm{i}\gamma_k \theta} \text{ and } q(\theta) = \sum_{k=1}^{N'} \beta_k \mathrm{e}^{\mathrm{i}\eta_k \theta}.$$

Then $p = q$ if and only if $N = N'$ and for all $k = 1, \ldots, N$ $\alpha_k = \beta_k$ and $\gamma_k = \eta_k$.

Lemma 1 implies that if $p \in \mathcal{E}(d_1)$ and $q \in \mathcal{E}(d_2)$, then

$$pq \in \mathcal{E}(d_1 + d_2). \tag{5}$$

In particular, if $p \in \mathcal{E}_0(d)$

$$|p|^2 = p\bar{p} \in \mathcal{E}(d + 0) = \mathcal{E}(d). \tag{6}$$

Furthermore, if $d_2 \leq d_1$ then

$$(p + q) \in \mathcal{E}(d_1), \tag{7}$$

except, of course, if $d_1 = d_2$ and the lead coefficients of p and q are negatives of one another.

The next several lemmas will be needed in the proofs of Theorems 2 and 3.

Lemma 2. *For $i = 1, 2, 3, 4$ let $p_i \in \mathcal{E}_0(d_i)$ assume that $d_1 > d_2, d_3, d_4$. Then the set of points θ such that*

$$|p_1(\theta)|^p + |p_2(\theta)|^p = |p_3(\theta)|^p + |p_4(\theta)|^p \tag{8}$$

has measure zero.

Lemma 3. *Let $p \geq 1$ be an odd integer, and let $a, b, c, d, C \in \mathbb{R}$, $a, b, c, d \neq 0$. Let $p(\theta) = a + b\mathrm{e}^{\mathrm{i}\theta}$, and $q(\theta) = c + d\mathrm{e}^{\mathrm{i}\theta}$. If there are more than $4p$ distinct $\theta \in [0, 2\pi]$ such that*

$$|p(\theta)|^p - |q(\theta)|^p = C,$$

then $ab = cd$ and $a^2 + b^2 = c^2 + d^2$.

Lemma 4. *Let $p \geq 1$ be an integer and let $a, b, c, d, C \in \mathbb{R}, \gamma > 0, \kappa \neq 0, \pm 1$. Then the set of θ such that*

$$\left|a + b\mathrm{e}^{\mathrm{i}\theta} + c\mathrm{e}^{\mathrm{i}(\gamma+1)\theta}\right|^p - \left|\kappa a + \frac{1}{\kappa} b\mathrm{e}^{\mathrm{i}\theta} + \kappa c\mathrm{e}^{\mathrm{i}(\gamma+1)\theta}\right|^p = C \tag{9}$$

has measure zero.

III. The proof of Theorem 2

Before proving Theorem 2 we will first prove a preliminary result which shows, even without the assumption that $x(t)$ is collision free, that $f_\xi(s)$ is a peicewise linear function whose set of knots is contained in $\mathcal{D}(x)$. This result is based on the observation that we may write

$$f_\xi(s) = \sum_{i<j} \alpha_{i,j}(s)\beta_{i,j}(\xi).$$

where for each $i < j$,

$$\beta_{i,j}(\xi) \coloneqq \sum_{\ell=i}^{j} a_\ell \mathrm{e}^{\mathrm{i}\xi\Delta_{i,\ell}} \tag{10}$$

is a function that only depends on ξ and $\alpha_{i,j}(s)$ is piecewise linear function of s whose singularities are contained in $\mathcal{D}(x)$.

Specifically, we prove the following theorem. We emphasize that this result does not assume that $x(t)$ is collision free, which is why for $d \in \mathcal{D}(x)$ there might be multiple i,j such that $\Delta_{i,j} = d$.

Theorem 4. *Let $p \geq 1$ be an integer, and assume $w(t) = 1_{[0,1]}(t)$. For $i \leq j$, $\beta_{i,j}(\xi)$ be as in (10). Then, for every fixed ξ, the function $f_\xi(s) = \|g_{s,\xi} * x\|_p^p$ is piecewise linear, and $\partial_s^2 f_\xi(s)$ is a grid-free sparse signal whose support is contained in $\mathcal{D}(x)$. Specifically,*

$$\partial_s^2 f_\xi(s) = \sum_{d \in \mathcal{D}(x)} \left(\sum_{\Delta_{i,j}=d} c_{i,j}(\xi) \right) \delta_d, \tag{11}$$

where

$$c_{i,i+1}(\xi) = |\beta_{i,i+1}(\xi)|^p - |\beta_{i+1,i+1}(\xi)|^p - |\beta_{i,i}(\xi)|^p \tag{12}$$

and for $j \geq i+2$

$$c_{i,j}(\xi) = |\beta_{i,j}(\xi)|^p + |\beta_{i+1,j-1}(\xi)|^p - |\beta_{i+1,j}(\xi)|^p - |\beta_{i,j-1}(\xi)|^p. \tag{13}$$

Proof. We first note that

$$\begin{aligned} |(g_{s,\xi} \star x)(t)| &= \left| \sum_{i=1}^k a_i g_{s,\xi}(t - v_i) \right| \\ &= \left| \sum_{i=1}^k a_i \mathbb{e}^{\mathbb{i}\xi(t-v_i)} 1_{[v_i, v_i+s]}(t) \right| \\ &= \left| \sum_{i=1}^k a_i \mathbb{e}^{-\mathbb{i}\xi v_i} 1_{[v_i, v_i+s]}(t) \right|. \end{aligned}$$

For $I \subseteq \{1, \ldots, k\}$, let $R_I(s)$ be the set of t for which $a_i \mathbb{e}^{-\mathbb{i}\xi v_i} 1_{[v_i, v_i+s]}(t)$ is nonzero if and only if $i \in I$, i.e.,

$$R_I(s) = \{t : t \in [v_i, v_i + s]\ \forall i \in I, t \notin [v_i, v_i+s]\ \forall i \notin I\}.$$

Then, since $w(t) = 1_{[0,1]}(t)$ it is clear that for $t \in R_I$,

$$|(g_{s,\xi} \star x)(t)| = \left| \sum_{i \in I} a_i \mathbb{e}^{-\mathbb{i}\xi v_i} \right| =: y_I(\xi).$$

Therefore,

$$f_\xi(s) = \|(g_{s,\xi} \star x)(t)\|_p^p = \sum_{I \subseteq \{1, \ldots k\}} |y_I(\xi)|^p |R_I(s)|, \tag{14}$$

where $|R_I(s)|$ denotes the Lebesgue measure of $R_I(s)$. We will show that for all $I \subseteq \{1, \ldots, k\}$, $|R_I(s)|$ is piecewise linear function whose knots are contained in $\mathcal{D}(x)$.

First, we note that $R_I(s) = \emptyset$ unless I has the form $\{i, i+1, \ldots, j-1, j\}$ for some $i \leq j$. Therefore,

$$f_s(\xi) = \sum_{i=1}^k \sum_{j=i}^k |\beta_{i,j}(\xi)|^p |R_{i,j}(s)|, \tag{15}$$

where $R_{i,j} := R_{\{i,\ldots,j\}}$. and, as in (10), $\beta_{i,j}(\xi)$ is given by

$$|\beta_{i,j}(\xi)| = \left| \sum_{\ell=i}^j a_\ell \mathbb{e}^{\mathbb{i}\xi \Delta_{i,\ell}} \right| = \left| \sum_{\ell=i}^j a_\ell \mathbb{e}^{\mathbb{i}\xi v_\ell} \right|.$$

Now, turning our attention to $R_{i,j}(s)$, we observe by definition that a point t is in $R_{i,j}(s)$ if and only if it satisfies the following three conditions:

$$\begin{aligned} v_\ell \leq\, & t \leq v_\ell + s \quad \text{for all } i \leq \ell \leq j, \\ & t > v_{i-1} + s, \text{ and} \\ & t < v_{j+1}. \end{aligned}$$

Therefore, letting $\vee(a,b)$ and $\wedge(a,b)$ denote $\min\{a,b\}$ and $\max\{a,b\}$, we see

$$\begin{aligned} R_{i,j}(s) &= [v_j, v_i + s] \cap [v_{i-1} + s, v_{j+1}] & (16) \\ &= [v_j \vee (v_{i-1} + s), (v_i + s) \wedge v_{j+1}], & (17) \end{aligned}$$

and therefore

$$|R_{i,j}(s)| = ((v_i + s) \wedge v_{j+1}) - (v_j \vee (v_{i-1} + s))$$

if the above quantity is positive and zero otherwise. It follows from (16) that $|R_{i,j}(s)|$ is a piecewise linear function, and that $\partial_s^2 |R_{i,j}(s)|$ is given by

$$\partial_s^2 |R_{i,j(S)}| = \delta_{\Delta_{i,j}}(s) + \delta_{\Delta_{i-1,j+1}}(s) - \delta_{\Delta_{i-1,j}}(s) - \delta_{\Delta_{i,j+1}}(s). \tag{18}$$

We note that in order for this equation to be valid for all $1 \leq i < j \leq k$, we identify v_0 and v_{k+1} with $-\infty$ and ∞, and therefore, $\delta_{\Delta_{0,j}}$ $\delta_{\Delta_{i-1,k+1}}$ are interpreted as being the zero function since the domain of f is $(0, \infty)$. Likewise $\delta_{\Delta_{i,i}} = \delta_0$ is interpreted as the zero function in the above equation.

Combining (18) with (15) implies that $\partial_s^2 f_\xi(s)$ is a sparse signal with support contained in $\mathcal{D}(x)$, and for $d \in \mathcal{D}(x)$,

$$\partial_s^2 f_\xi(d) = \sum_{\Delta_{i,j}=d} c_{i,j}(\xi)$$

as desired. □

Before we prove Theorem 2, we note the following example which shows that, in general, the support of $\partial_s^2 f_\xi(s)$ may be a proper subset of $\mathcal{D}(x)$.

Example 1. *If $p = 2$ and*

$$x(t) = \delta_1(t) + \delta_2(t) + \delta_3(t) - \delta_4(t),$$

then $2 \in \mathcal{D}(x)$, but

$$\partial_s^2 f_\xi(2) = 0.$$

Proof. For this choice of x, there are two pairs (i,j) such that $\Delta_{i,j} = 2$, namely $(1,3)$ and $(2,4)$. Therefore, by Theorem 4,

$$\begin{aligned} \partial_s^2 f_\xi(2) = &\left(|y_{1,3}(\xi)|^2 + |y_{2,2}(\xi)|^2 - |y_{1,2}(\xi)|^2 - |y_{2,3}(\xi)|^2 \right) \\ &+ \left(|y_{2,4}(\xi)|^2 + |y_{3,3}(\xi)|^2 - |y_{2,3}(\xi)|^2 - |y_{3,4}(\xi)|^2 \right). \end{aligned}$$

Inserting $(a_1, a_2, a_3, a_4) = (1, 1, 1, -1)$, $\Delta_{i,i+1} = 1$, and $\Delta_{i,i+2} = 2$ into (10) implies that

$$\begin{aligned} \partial_s^2 f_\xi(2) = &\left(|1 + \mathbb{e}^{\mathbb{i}\xi} + \mathbb{e}^{2i\xi}|^2 + 1 - |1 + \mathbb{e}^{\mathbb{i}\xi}|^2 - |1 + \mathbb{e}^{\mathbb{i}\xi}|^2 \right) \\ &+ \left(|1 + \mathbb{e}^{\mathbb{i}\xi} - \mathbb{e}^{2i\xi}|^2 + 1 - |1 + \mathbb{e}^{\mathbb{i}\xi}|^2 - |1 - \mathbb{e}^{\mathbb{i}\xi}|^2 \right) \\ = &\ |1 + \mathbb{e}^{\mathbb{i}\xi} + \mathbb{e}^{2i\xi}|^2 + |1 + \mathbb{e}^{\mathbb{i}\xi} - \mathbb{e}^{2i\xi}|^2 \\ &+ 2 - 3|1 + \mathbb{e}^{\mathbb{i}\xi}|^2 - |1 - \mathbb{e}^{\mathbb{i}\xi}|^2 \\ = &\ 0. \end{aligned}$$

The last inequality follows from repeatedly applying the the trigonometric identities $\sin^2(\theta)+\cos^2(\theta) = 1$ and $\cos(\theta) = \cos(2\theta)\cos(\theta) + \sin(2\theta)\sin(\theta)$. ☐

We shall now prove Theorem 2.

The Proof of Theorem 2. By assumption, $x(t)$ is collision free. Therefore, for all $d \in \mathcal{D}(x)$, there is a unique i,j such that $\Delta_{i,j} = d$, and so, by (11), it suffices to show that $c_{i,j}(\xi) \neq 0$ for all $i<j$ and for almost every $\xi \in \mathbb{R}$, where as in (12) and for $j \geq i+$ (13)

$$c_{i,i+1}(\xi) = |\beta_{i,i+1}(\xi)|^p - |\beta_{i+1,i+1}(\xi)|^p - |\beta_{i,i}(\xi)|^p,$$

and for $j \geq i+2$,

$$c_{i,j}(\xi) = |\beta_{i,j}(\xi)|^p + |\beta_{i+1,j-1}(\xi)|^p - |\beta_{i+1,j}(\xi)|^p - |\beta_{i,j-1}(\xi)|^p,$$

where

$$\beta_{i,j}(\xi) = \left|\sum_{k=i}^{j} a_k \mathbb{e}^{-\mathbb{i}\xi\Delta_{i,k}}\right|.$$

Observe that $\beta_{i,j}(\xi)$ are generalized exponential Laurent polynomials of the form introduced in Section II, and in particular, $\beta_{i,j} \in \mathcal{E}_0(\Delta_{i,j})$. Therefore, when $j \geq i+2$, it follows from Lemma 2 that $c_{i,j}(\xi)$ vanishes on a set of measure zero since if $c_{i,j}(\xi) = 0$ we have

$$|\beta_{i,j}(\xi)|^p + |\beta_{i+1,j-1}(\xi)|^p = |\beta_{i+1,j}(\xi)|^p + |\beta_{i,j-1}(\xi)|^p.$$

In the case where $j = i+1$, we see that

$$c_{i,i+1}(\xi) = |a_i + a_{i+1}\mathbb{e}^{-\mathbb{i}\xi\Delta_{i,i+1}}|^p - |a_i|^p - |a_{i+1}|^p,$$

For any ξ such that $c_{i,i+1}(\xi) = 0$, we see that $\xi\Delta_{i,i+1}$ is a solution to

$$\left|a_i + a_{i+1}\mathbb{e}^{\mathbb{i}\theta}\right|^2 - (|a_i|^p + |a_{i+1}|^p)^{2/p} = 0.$$

Thus, $c_{i,i+1}(\xi)$ vanishes on a set of measure zero since the left-hand side of the above equation is a trigonometric polynomial. ☐

IV. The Proof of Theorems 3

Proof. Let $\xi_1, \xi_2, \ldots, \xi_L$ be i.i.d. standard normal random variables. Since x is collision free, with probability one, each of the $\xi_\ell\Delta_{i,i+1}(x)$ are distinct modulo 2π, i.e.

$$\xi_\ell\Delta_{i,i+1}(x) \not\equiv \xi_{\ell'}\Delta_{i',i'+1}(x) \mod 2\pi \tag{19}$$

for all $1 \leq i,i' \leq k-1$ and $1 \leq \ell,\ell' \leq L$, except when $(i,\ell) = (i',\ell')$. For the rest of the proof we will assume this is the case.

Let

$$\widetilde{x}(t) = \sum_{j=1}^{k} \widetilde{a}_j \delta_{\widetilde{v}_j}(t)$$

be a signal $\mathcal{D}(\widetilde{x}) = \mathcal{D}(x)$, $\|\vec{a}\|_p = \|\vec{\widetilde{a}}\|_p$, and $\partial_s^2 f_{\xi_\ell}[x](d) = \partial_s^2 f_{\xi_\ell}[\widetilde{x}](d)$ for all $d \in \mathcal{D}(x)$ and for all $1 \leq \ell \leq L-1$. Note that $\widetilde{x}(t)$ depends on $\xi_1, \ldots, \xi_{L-1}$, but is independent of ξ_L. By assumption that $x(t)$ and $\widetilde{x}(t)$ are collision free (and also, as discussed in the Section I, we assume that we are not in the special case where $k = 6$ and the $\{v_j\}_{j=1}^k$ belong to a special parametrized family). Therefore, the fact that $\mathcal{D}(x) = \mathcal{D}(\widetilde{x})$ implies that the support sets of x and $\widetilde{x}$ are equivalent up to translation and reflection, so we may assume without loss of generality that $\Delta_{i,j}(x) = \Delta_{i,j}(\widetilde{x}) =: \Delta_{i,j}$ for all $1 \leq i \leq j \leq k$.

We will show that $\vec{\widetilde{a}}$ must be given by

$$\widetilde{a}_i = \begin{cases} \frac{1}{c}a_i & \text{if } i \text{ is odd} \\ ca_i & \text{if } i \text{ is even} \end{cases}, \tag{20}$$

where $c = \pm 1$ or

$$|c|^p = \frac{\sum_{i=1}^{\lfloor \frac{k+1}{2} \rfloor} |a_{2i-1}|^p}{\sum_{i=1}^{\lfloor \frac{k}{2} \rfloor} |a_{2i}|^p}. \tag{21}$$

Then, we will show that, if c satisfies (21), but $c \neq \pm 1$, then with probability one

$$\partial_s^2 f_{\xi_L}[x](\Delta_{1,3}) \neq \partial_s^2 f_{\xi_L}[\widetilde{x}](\Delta_{1,3}).$$

Since $\widetilde{x}(t)$ (and therefore $\vec{\widetilde{a}}$) was chosen to depend on $\xi_1, \ldots, \xi_{L-1}$, but not ξ_L, these two facts together will imply that, with probability one, if $\widetilde{x}(t)$ is any signal such that $\mathcal{D}(\widetilde{x}) = \mathcal{D}(x)$ and $\partial_s^2 f_{\xi_\ell}[x](d) = \partial_s^2 f_{\xi_\ell}[\widetilde{x}](d)$ for all $d \in \mathcal{D}$ and all $1 \leq \ell \leq L$, then $\vec{\widetilde{a}} = \pm\vec{a}$ and therefore $\widetilde{x}(t)$ is equivalent to $\pm x(t)$ up to reflection and translation.

We first will show that (20) holds in the case where p is odd. Setting $\partial_s^2 f_{\xi_\ell}[x](\Delta_{i,i+1}) = \partial_s^2 f_{\xi_\ell}[\widetilde{x}](\Delta_{i,i+1})$ and using (12) implies that for all $1 \leq \ell \leq L-1$ and all $1 \leq i \leq k-1$ we have

$$\begin{aligned} &|a_i + a_{i+1}\mathbb{e}^{\mathbb{i}\xi_\ell\Delta_{i,i+1}}|^p - |a_{i+1}|^p - |a_i|^p \\ =&|\widetilde{a}_i + \widetilde{a}_{i+1}\mathbb{e}^{\mathbb{i}\xi_\ell\Delta_{i,i+1}}|^p - |\widetilde{a}_{i+1}|^p - |\widetilde{a}_i|^p. \end{aligned} \tag{22}$$

Therefore, $\xi_1\Delta_{i,i+1}, \ldots, \xi_{L-1}\Delta_{i,i+1}$ constitute $L-1$ solutions, which are distinct modulo 2π, to the equation

$$|a_i + a_{i+1}\mathbb{e}^{\mathbb{i}\theta}|^p - |\widetilde{a}_i + \widetilde{a}_{i+1}^{\mathbb{i}\theta}|^p = |\widetilde{a}_i|^p + |\widetilde{a}_{i+1}|^p - |a_i|^p - |a_{i+1}|^p.$$

Since $L-1 > 4p$, Lemma 3 implies that

$$a_i a_{i+1} = \widetilde{a}_i\widetilde{a}_{i+1} \quad \text{and} \quad \widetilde{a}_i^2 + \widetilde{a}_{i+1}^2 \tag{23}$$

for all $1 \leq i \leq k-1$. It follows from (23) that (20) holds with $c := a_1/\widetilde{a}_1$.

Now consider the case where $p = 2m$ is even. Similarly to (22), the assumption that $\partial_s^2 f_{\xi_\ell}[x](\Delta_{i,i+1}) = \partial_s^2 f_{\xi_\ell}[\widetilde{x}](\Delta_{i,i+1})$ implies that for all $1 \leq \ell \leq L$, $1 \leq i \leq k-1$,

$$\begin{aligned} &|a_i + a_{i+1}\mathbb{e}^{\mathbb{i}\xi_\ell\Delta_{i,i+1}}|^{2m} - |a_i|^{2m} - |a_{i+1}|^{2m} \\ =&|\widetilde{a}_i + \widetilde{a}_{i+1}\mathbb{e}^{\mathbb{i}\xi_\ell\Delta_{i,i+1}}|^{2m} - |\widetilde{a}_i|^{2m} - |\widetilde{a}_{i+1}|^{2m}. \end{aligned}$$

Therefore, for all $1 \leq i \leq k-1$, $\xi_1\Delta_{i,i+1}, \ldots, \xi_{L-1}\Delta_{i,i+1}$ are $L-1$ zeros of

$$\begin{aligned} h_i(\theta) := &|a_i + a_{i+1}\mathbb{e}^{\mathbb{i}\theta}|^{2m} - |\widetilde{a}_i + \widetilde{a}_{i+1}\mathbb{e}^{\mathbb{i}\theta}|^{2m} \\ &+ |\widetilde{a}_i|^{2m} + |\widetilde{a}_{i+1}|^{2m} - |a_i|^{2m} - |a_{i+1}|^{2m} \end{aligned}$$

which are distinct modulo 2π. Using the fact that

$$|a_i + a_{i+1}\mathbb{e}^{\mathbb{i}\theta}|^2 = a_i^2 + a_{i+1}^2 + 2a_i a_{i+1}\cos(\theta)$$

one may verify that $h_i(\theta)$ is a trigonometric polynomial of degree at most m given by

$$h_i(\theta) = (a_i^2 + a_{i+1}^2 + 2a_i a_{i+1}\cos(\theta))^m - (\widetilde{a}_i^2 + \widetilde{a}_{i+1}^2 + 2\widetilde{a}_i\widetilde{a}_{i+1}\cos(\theta))^m + \widetilde{a}_i^{2m} + \widetilde{a}_{i+1}^{2m} - a_i^{2m} - a_{i+1}^{2m}$$

Thus, since $L-1 > p = 2m$, this implies that $h(\theta)$ must be uniformly zero. In particular, setting the lead coefficient equal to zero implies

$$(a_i a_{i+1})^m = (\widetilde{a}_i\widetilde{a}_{i+1})^m$$

for all $1 \leq i \leq k-1$. Using the binomial theorem and setting the $\cos(\theta)$ coefficient equal to zero gives

$$(a_i^2 + a_{i+1}^2)^{m-1} a_i a_{i+1} = (\widetilde{a}_i^2 + \widetilde{a}_{i+1}^2)^{m-1}\widetilde{a}_i\widetilde{a}_{i+1}.$$

Together, the last two equations imply

$$a_i^2 + a_{i+1}^2 = \widetilde{a}_i^2 + \widetilde{a}_{i+1}^2 \quad \text{and} \quad a_i a_{i+1} = \widetilde{a}_i\widetilde{a}_{i+1}.$$

As in the case where p was odd, this implies that (20) must hold.

Combining (20) with the assumption that

$$\sum_{i=1}^{k} |a_i|^p = \sum_{i=1}^{k} |\widetilde{a}_i|^p$$

implies that either $c = \pm 1$ or that c satisfies (21). Thus the proof will be complete once we show that if c satisfies (21), but $c \neq \pm 1$,, then with probability one, $\partial_s^2 f_{\xi_L}[x](\Delta_{1,3}) \neq \partial_s^2 f_{\xi_L}[\widetilde{x}](\Delta_{1,3})$.

By (13), if $\partial_s^2 f_{\xi_L}[x](\Delta_{1,3}) = \partial_s^2 f_{\xi_L}[\widetilde{x}](\Delta_{1,3})$, then

$$\begin{aligned} &|a_1 + a_2 e^{i\xi_L\Delta_{1,2}} + a_3 e^{i\xi_L\Delta_{1,3}}|^p + |a_2|^p \\ &\quad - |a_2 e^{i\xi_L\Delta_{1,2}} + a_3 e^{i\xi_L\Delta_{1,3}}|^p - |a_1 + a_2 e^{i\xi_L\Delta_{1,2}}|^p \\ =&|\widetilde{a}_1 + \widetilde{a}_2 e^{i\xi_L\Delta_{1,2}} + \widetilde{a}_3 e^{i\xi_L\Delta_{1,3}}|^p + |\widetilde{a}_2|^p \\ &\quad - |\widetilde{a}_2 e^{i\xi_L\Delta_{1,2}} + \widetilde{a}_3 e^{i\xi_L\Delta_{1,3}}|^p - |\widetilde{a}_1 + \widetilde{a}_2 e^{i\xi_L\Delta_{1,2}}|^p. \end{aligned} \tag{24}$$

But (23) implies that for all i either $(a_i, a_{i+1}) = \pm(\widetilde{a}_i, \widetilde{a}_{i+1})$ or $(a_i, a_{i+1}) = \pm(\widetilde{a}_{i+1}, \widetilde{a}_i)$. In either case, we have that

$$|a_1 + a_2 e^{i\xi_L\Delta_{1,2}}| = |\widetilde{a}_1 + \widetilde{a}_2 e^{i\xi_L\Delta_{1,2}}|$$

and

$$|a_2 e^{i\xi_L\Delta_{1,2}} + a_3 e^{i\xi_L\Delta_{1,3}}| = |\widetilde{a}_2 e^{i\xi_L\Delta_{1,2}} + \widetilde{a}_3 e^{i\xi_L\Delta_{1,3}}|.$$

Combining this with (24) gives

$$\begin{aligned} &|a_1 + a_2 e^{i\xi_L\Delta_{1,2}} + a_3 e^{i\xi_L\Delta_{1,3}}|^p + |a_2|^p \\ &= |\widetilde{a}_1 + \widetilde{a}_2 e^{i\xi_L\Delta_{1,2}} + \widetilde{a}_3 e^{i\xi_L\Delta_{1,3}}|^p + |\widetilde{a}_2|^p. \end{aligned} \tag{25}$$

However, by Lemma 4 the set of $\xi_L \in \mathbb{R}$ such that (25) holds has measure zero, unless $c = \pm 1$. Since ξ_L is a normal random variable, this happens with probability zero. Thus, the proof is complete. □

V. Proofs of auxiliary lemmas

In this section, we will provide proofs of the lemmas stated in section III.

The Proof of Lemma 1. By linearity, it suffices to show that $\alpha_1, \ldots, \alpha_N$ are nonzero numbers, then $p(\theta) = \sum_{k=1}^N \alpha_k e^{i\gamma_k\theta}$ is not the zero function. We will restrict attention to the case where $|\gamma_N| > |\gamma_k|$ for all $1 \leq k \leq N-1$. The proofs of the other cases, where either $|\gamma_1| > |\gamma_k|$ for all $2 \leq k \leq N$ or where $|\gamma_1| = |\gamma_N| > |\gamma_k|$ for all $2 \leq k \leq N-1$ are similar.

For all $n \geq 1$, the n-th derivative of $p(\theta)$ is given by

$$p^{(n)}(\theta) = \sum_{k=1}^{N} \alpha_k \gamma_k^n e^{i\gamma_k\theta}.$$

Therefore, since $|\gamma_N| > \gamma_k$ for all $1 \leq k \leq N-1$, we have

$$\lim_{n\to\infty} \frac{p^{(n)}(0)}{\gamma_N^n} = \alpha_N.$$

In particular, there exists n such that $p^{(n)}(0) \neq 0$, and therefore $p(\theta)$ is not the zero function. □

The proof of Lemma 2. In the case where $p = 2m$ is even, then by (6) and (7), $|p_i(\theta)|^{2m} \in \mathcal{E}(md_i)$ for each i. Therefore, since $d_1 > d_2, d_3, d_4$, it follows from (7) that

$$|p_1(\theta)|^{2m} + |p_2(\theta)|^{2m} - |p_3(\theta)|^{2m} - |p_4(\theta)|^{2m}$$

is an element of $\mathcal{E}(md_1)$ and therefore vanishes on a set of measure zero.

Now consider the case where $p = 2m+1$ is odd. Squaring both sides of (8) implies

$$p_5(\theta) = 2(|p_1(\theta)p_2(\theta)|^{2m+1} - |p_3(\theta)p_4(\theta)|^{2m+1}),$$

where

$$p_5(\theta) := |p_1(\theta)|^{4m+2} + |p_2(\theta)|^{4m+2} - |p_3(\theta)|^{4m+2} - |p_4(\theta)|^{4m+2}.$$

Thus, squaring both sides again gives

$$p_6(\theta) = 8|p_1(\theta)p_2(\theta)p_3(\theta)p_4(\theta)|^{2m+1},$$

where

$$p_6(\theta) := p_5(\theta)^2 - 4|p_1(\theta)p_2(\theta)|^{4m+2} - 4|p_3(\theta)p_4(\theta)|^{4m+2}.$$

Therefore, squaring both sides one final time implies that

$$p_6(\theta)^2 - 64|p_1(\theta)p_2(\theta)p_3(\theta)p_4(\theta)|^{4m+2} = 0.$$

However, since $d_1 > d_2, d_3, d_4$, applying (5), (6), and (7), implies that $(p_6(\theta)^2 - 64|p_1(\theta)p_2(\theta)p_3(\theta)p_4(\theta)|^2) \in \mathcal{E}((8m+4)d_1)$ and therefore vanishes on a set of measure zero. □

The Proof of Lemma 3. If θ is a solution to

$$|p(\theta)|^p - |q(\theta)|^p = C,$$

then

$$|p(\theta)|^{2p} - |q(\theta)|^{2p} - C^2 = 2|q(\theta)|^p C.$$

Therefore,

$$f(\theta) := \left(|p(\theta)|^{2p} - |q(\theta)|^{2p} - C^2\right)^2 - 4|q(\theta)|^{2p}C^2 = 0.$$

Since

$$|p(\theta)|^2 = a^2+b^2+2ab\cos(\theta) \quad \text{and} \quad |q(\theta)|^2 = c^2+d^2+cd\cos(\theta),$$

$f(\theta)$ is a trigonometric polynomial of degree at most $2p$. Thus, if $f(\theta)$ has more than $4p$ zeros in $[0, 2\pi]$ it must be uniformly zero. Expanding out terms, we see

$$\begin{aligned} &f(\theta) \\ =&((a^2+b^2+2ab\cos(\theta))^p - (c^2+d^2+2cd\cos(\theta))^p - C^2)^2 \\ &-4C^2(c^2+d^2+2cd\cos(\theta))^p \end{aligned}$$

and so setting the $\cos^{2p}(\theta)$ coefficient equal to zero implies

$$0 = (2^p a^p b^p - 2^p c^p d^p)^2$$

which implies $ab = cd$ since p is odd. If $p \geq 3$, then we have $2(p-1) > p$. Therefore,

$$f_1(\theta) := ((a^2+b^2+2ab\cos(\theta))^p - (c^2+d^2+2cd\cos(\theta))^p - C^2)^2$$

has strictly greater degree than

$$f_2(\theta) := 4C^2(c^2+d^2+2cd\cos(\theta))^p$$

and so $f_1(\theta)$ must also be uniformly zero. Setting the $\cos^{p-1}(\theta)$ coefficient of $f_1(\theta)^{1/2}$ equal to zero yields

$$\left(p(a^2+b^2)(2ab)^{p-1} - p(c^2+d^2)(2cd)^{p-1}\right)^2 = 0,$$

but since $ab = cd$ this implies that $a^2+b^2 = c^2+d^2$. On the other hand, if $p = 1$, using the fact that $ab = cd$ we see that

$$0 = \left(a^2+b^2-(c^2+d^2)-C\right)^2 - 4C^2(c^2+d^2+2cd\cos(\theta)),$$

which can only happen for all θ if $C = 0$ and $a^2+b^2 = c^2+d^2$. □

The Proof of Lemma 4. Let

$$p(\theta) = a + b\mathbb{e}^{\mathbb{i}\theta} + c\mathbb{e}^{\mathbb{i}(\gamma+1)\theta}$$

and

$$q(\theta) = \kappa a + \frac{1}{\kappa} b\mathbb{e}^{\mathbb{i}\theta} + \kappa c\mathbb{e}^{\mathbb{i}(\gamma+1)\theta}.$$

Then squaring both sides of (9) yields,

$$|p(\theta)|^{2p} - |q(\theta)|^{2p} - C^2 = 2|q(\theta)|^p C,$$

and therefore if θ satisfies (9) it is a solution to $f(\theta) = 0$, where

$$f(\theta) := \left(|p(\theta)|^{2p} - |q(\theta)|^{2p} - C^2\right)^2 - 4|q(\theta)|^{2p}C^2.$$

$f(\theta)$ is an element of the class $\mathcal{E}$ of generalized exponential polynomials introduced earlier. Thus, it will follow that f vanishes on a set of measure zero as soon as we show that f is not uniformly zero. We will verify that that the lead cofficient of f is nonzero unless $\kappa = \pm 1$. Using the identity $\mathbb{e}^{\mathbb{i}x} + \mathbb{e}^{-\mathbb{i}x} = 2\cos(x)$ we see that

$$\begin{aligned} |p(\theta)|^2 = &a^2+b^2+c^2+2ab\cos(\theta) \\ &+2bc\cos(\gamma\theta)+2ac\cos((\gamma+1)\theta) \end{aligned}$$

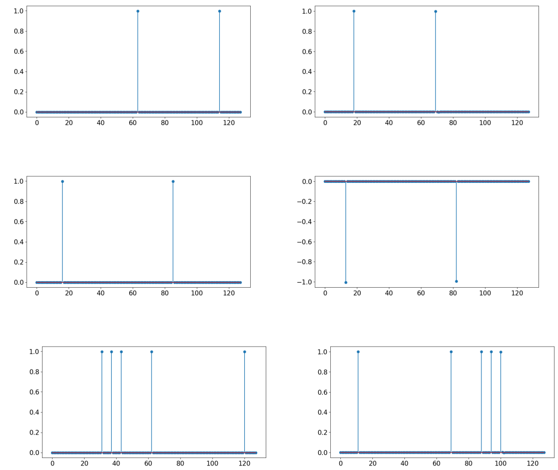


Fig. 4: Sparse signals reconstructed up to a global reflections, translations, and sign changes. Originals signals are on the left and reconstructed signals are on the right.

and likewise is given by

$$\begin{aligned} |q(\theta)|^2 = &\kappa^2 a^2 + \frac{1}{\kappa^2} b^2 + \kappa^2 c^2 \\ &+ 2ab\cos(\theta) + 2bc\cos(\gamma\theta) + 2\kappa^2 ac\cos((\gamma+1)\theta). \end{aligned}$$

Therefore, the lead coefficient of $f(\theta)$ vanishes if and only if $\kappa^2 = 1$.

□

VI. Signal Synthesis

In order to illustrate the ability of the measurements **??** to characterize sparse signals, we verify empirically that signals with the same measurement differ only by global reflections, translations and sign changes. Specifically, given a signal $x(t) = \sum_{j=1}^{k} a_j \delta_{v_j}(t)$ and a finite collection of measurements

$$\{\|g_{s_k,\xi_k} \star x\|_p\}_{k=1}^N$$

we use a greedy scheme to find another signal $\widetilde{x}(t) = \sum_{j=1}^{k} \widetilde{a}_j \delta_{v_j}(t)$ which minimizes

$$\sum_{k=1}^{N} \left| \|g_{s_k,\xi_k} \star x\|_p - \|g_{s_k,\xi_k} \star \widetilde{x}\|_p \right|^2.$$

Figure 4 shows the result of several signals $\tilde{x}$ which were obtained by solving this minimization problem.

In all of experiments, we set the signal lengther to be $N = 128$ and used two frequencies $\xi = (41/N)\pi, (23/N)\pi$. For the first signal we used scales $s = 1, 14, 27, 40, 53, 2, 65, 96, 106$. For the second singal we used $s = 1, 6, 11, 16, 21, 26, 31, 36, 41, 46, 51, 56, 61, 2, 4, 65, 96, 106$, and for the third we used $s = 1, 7, 13, 19, 25, 31, 37, 43, 49, 55, 61, 2, 4, 6, 65, 96, 106$.

REFERENCES

[1] Ahmad Bekir and Solomon W. Golomb. There are no further counterexamples to s. piccard's theorem. *IEEE Transactions on Information Theory*, 53(8):2864–2867, 2007.

[2] Stéphane Mallat. *A Wavelet Tour of Signal Processing, Third Edition: The Sparse Way*. Academic Press, 3rd edition, 2008.

[3] Stéphane Mallat. Group invariant scattering. *Communications on Pure and Applied Mathematics*, 65(10):1331–1398, October 2012.

[4] Stéphane Mallat and Iréne Waldspurger. Phase retrieval for the cauchy wavelet transform. *Journal of Fourier Analysis and Applications*, 21(6):1251–1309, 2015.

[5] Juri Ranieri, Amina Chebira, Yue M. Lu, and Martin Vetterli. Phase retrieval for sparse signals: Uniqueness conditions. *CoRR*, abs/1308.3058, 2013.