**Research Article**

Mathematical Methods in the Applied Sciences

Received 4 August 2012 Published online 28 February 2013 in Wiley Online Library

(wileyonlinelibrary.com) DOI: 10.1002/mma.2752
MOS subject classification: 41A20; 30J99; 65E99

# On sparse representation of analytic signal in Hardy space

**Shuang Li*† and Tao Qian**

**Communicated by W. Sprößig**

**This paper is concerned with the sparse representation of analytic signal in Hardy space $H^2(\mathbb{D})$, where $\mathbb{D}$ is the open unit disk in the complex plane. In recent years, adaptive Fourier decomposition has attracted considerable attention in the area of signal analysis in $H^2(\mathbb{D})$. As a continuation of adaptive Fourier decomposition-related studies, this paper proves rapid decay properties of singular values of the dictionary. The rapid decay properties lay a foundation for applications of compressed sensing based on this dictionary. Through Hardy space decomposition, this program contributes to sparse representations of signals in the most commonly used function spaces, namely, the spaces of square integrable functions in various contexts. Numerical examples are given in which both compressed sensing and $\ell_1$-minimization are used.**


**Keywords:** Hardy space; singular value; reproducing kernels; $\ell_1$-minimization; compressed sensing; sparse representation

## 1. Introduction

Sparse representation of signals has long been of interest. Our study originates from a series of recent results on analytic signal decomposition and adaptive rational approximation by Qian *et al.* where the concept of adaptive Fourier decomposition (AFD) was introduced [1–4]. By maximal projection principle [1], AFD yields an approximation using only a few elements chosen adaptively from the set of shifted Cauchy kernels

$$\mathscr{D} = \left\{ e_a : e_a(z) = \frac{\sqrt{1-|a|^2}}{1-\bar{a}z}, a \in \mathbb{D} \right\}. \tag{1}$$

The parameters $\{a_n\}$ of $\{e_{a_n}\}$ do not necessarily satisfy the hyperbolic nonseparability condition

$$\sum_{k=1}^{\infty}(1-|a_k|) = \infty$$

which plays a fundamental role in the study of the Takenaka–Malmquist basis $\{B_n\}_{n=1}^{\infty}$ of $H^2(\mathbb{D})$,

$$B_n(z) = B_{\{a_1,\dots,a_n\}}(z) \triangleq \frac{1}{2\pi}\frac{\sqrt{1-|a_n|^2}}{1-\bar{a}_n z}\prod_{k=1}^{n-1}\frac{z-a_k}{1-\bar{a}_k z}.$$

The AFD is motivated by matching pursuit (MP), which is a greedy algorithm that selects the dictionary atoms sequentially. A typical MP is a substitution of the following representation problem

$$\min \|x\|_0 \quad \text{subject} \quad \text{to} \quad s = \mathcal{D}x. \tag{2}$$

The problem is NP-hard which means non-deterministic polynomial-time hard in general [5–7] because it requires combinatorial search through all the combinations of columns from the dictionary $\mathcal{D}$. Thus, it is necessary to rely on good but not optimal approximations with computational algorithms. Basis pursuit (BP) is another substitution to achieve this goal. Instead of (2), BP suggests solving an $\ell_1$-minimization problem

$$\min \|x\|_1 \quad \text{subject} \quad \text{to} \quad s = \mathcal{D}x. \tag{3}$$

*Department of Mathematics, University of Macau, Macau, China*
**Correspondence to: Shuang Li, Department of Mathematics, University of Macau, Macau, China.*
†*E-mail: ya97418@umac.mo*

This problem is convex and it can be recast as a linear program [8]. BP is a method for a more global optimization. Empirical evidence suggests that BP is more powerful than MP [8], and the stability of BP has been proved in the presence of noise for sparse enough representations [9]. In some cases, the solution of (3) coincides with that of (2) [10]. However, BP has a drawback that the optimization procedure (3) often suffers from heavy computational complexity. The reason for this may be the numerical features of $\mathcal{D}$ is not revealed thoroughly. In this work, we utilize SVD to analyze the intrinsic structure of the dictionary matrix $\mathcal{D}$ (defined in Section 2). The SVD is an important factorization scheme because of its unique ability to split up data space into orthogonal signal and noise subspaces [11]. Given a matrix $\mathcal{D}$, we can derive a decomposition as

$$\mathcal{D} = U\Sigma V^* \tag{4}$$

where $U$ and $V$ are unitary matrices and $\Sigma$ is a diagonal matrix $\text{diag}(\sigma_1, \sigma_2, \ldots, \sigma_r)$ with $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r \geq 0$, $r = \min\{\text{rank}(U), \text{rank}(V)\}$. Consequently, the linear system $s = \mathcal{D}x$ is rewritten as

$$U^*s = \Sigma V^* x. \tag{5}$$

The transformed signal $U^*s$ can be well approximated by discarding a large number of small entries that correspond to the small singular values in $\Sigma$ provided that the singular values of $\mathcal{D}$ decay fast. That is, $U^*s$ can be approximated by

$$U^*s \approx \begin{pmatrix} \lfloor U^*s \rfloor_K \\ 0 \end{pmatrix} \tag{6}$$

where $\lfloor U^*s \rfloor_K$ is a vector of the first $K$ entries of $U^*s$. And

$$\lfloor U^*s \rfloor_K = \text{diag}(\sigma_1, \ldots, \sigma_K)\lfloor V^* \rfloor_K x, \tag{7}$$

where $\lfloor V^* \rfloor_K$ stands for the first $K$ rows of $V^*$. Therefore, instead of (3), we solve

$$\min \|x\|_1 \quad \text{subject} \quad \text{to} \quad \lfloor U^*s \rfloor_K = \text{diag}(\sigma_1, \ldots, \sigma_K)\lfloor V^* \rfloor_K x = A_K x \tag{8}$$

to obtain a sparse representation of $s$ with a small error. The suitable value of $K$ depends on the decay rate of singular values. In this paper, we give a sharp estimation of singular values distribution. $K$ can be selected not quite large, and hence, the computational complexity is reduced to a large extent. To our knowledge, this is the first time such estimation is given.

Much of the recent interest on $\ell_1$-minimization has come in the emerging field of compressed sensing (CS) [12]. The idea of CS originated in Kashin's paper [13, 14] and was brought into the forefront by Candes, Romberg, and Tao [15–17] and Donoho[18]. This is a setting in which one wishes to recover a signal from a small number of compressive measurements. Suppose $\Phi$ is a suitable random matrix, then with very high probability, the sparse vector $x$ can be obtained from a small number of measurements by solving

$$\min \|x\|_1 \quad \text{subject} \quad \text{to} \quad y = \Phi x. \tag{9}$$

The CS is proved to be robust in the sense that it can also deal with approximately sparse signals [12]. Moreover, this technique can be extended to signals that not sparse in an orthonormal basis but rather in a redundant dictionary [19, 20]. Given a dictionary matrix $\mathcal{D}$ and a signal $s$, one can derive $x$ in solving

$$\min \|x\|_1 \quad \text{subject} \quad \text{to} \quad y = \Phi\mathcal{D}x \tag{10}$$

with high probability. The restricted isometry constants have been analyzed in [19]. We can enhance the recovery effect by increasing the number of rows of $\Phi$. In our work, both BP and CS are utilized to illustrate that the dictionary $\mathscr{D}$ (1) does give a sparse representation of analytic signals in $H^2(\mathbb{D})$.

The paper is organized as follows. Preliminaries and notations are given in Section 2. The main results are proved in Section 3, and numerical examples are presented in Section 4.

## 2. Preliminaries and notations

Let $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$, and let $\text{Hol}(\mathbb{D})$ be the space of analytic functions on $\mathbb{D}$.
For $1 \leq p < \infty$,

$$H^p(\mathbb{D}) = \left\{ f \in \text{Hol}(\mathbb{D}) : \|f\|_{H^p}^p = \sup_{0 \leq r < 1} \int_0^{2\pi} |f(re^{ix})|^p \mathrm{d}x/2\pi < \infty \right\}$$

or equivalently

$$H^p(\mathbb{D}) = \{f \in L^p(0, 2\pi) : \widehat{f}(k) = 0, k < 0\}$$

where $\widehat{f}(k)$ denotes the $k$-th Fourier coefficient of $f$. The definitions indicate that an analytic signal in $H^p(\mathbb{D})$ can be uniquely determined by its boundary value. Hence, when we refer to an analytic signal, we regard it as a one-dimensional signal defined on $[0, 2\pi]$, rather than a holomorphic function on the disk $\mathbb{D}$. $H^2(\mathbb{D})$ is a complete subspace of $L^2(0,2\pi)$, which is the closure of the set formed by finite linear combinations of $\{e^{int}\}_{n=0}^{\infty}$, and it inherits the inner product

$$\langle f,g\rangle = \frac{1}{2\pi}\int_0^{2\pi} f(e^{it})\bar{g}(e^{it})\mathrm{d}t, \qquad \forall f,g\in H^2(\mathbb{D}).$$

Moreover, $H^2(\mathbb{D})$ is equipped with reproducing kernels

$$\mathscr{K} = \left\{k_a : k_a(z) = \frac{1}{1-\bar{a}z}, a\in\mathbb{D}\right\}, \tag{11}$$

which gives

$$f(a) = \langle f, k_a\rangle, \qquad \forall f\in H^2(\mathbb{D}). \tag{12}$$

In fact, Equation (12) can be derived by the Cauchy integral formula, that is,

$$\begin{aligned}\langle f,k_a\rangle &= \frac{1}{2\pi}\int_0^{2\pi} f(e^{it})\overline{\frac{1}{1-\bar{a}e^{it}}}\mathrm{d}t\\ &= \frac{1}{2\pi i}\int_{\partial\mathbb{D}} f(\zeta)\frac{1}{\zeta-a}\mathrm{d}\zeta\\ &= f(a).\end{aligned}$$

Each $e_a\in\mathscr{D}$ (1) is the normalized reproducing kernel $k_a\in\mathscr{K}$. That means

$$e_a = \frac{k_a}{\|k_a\|} = \frac{k_a}{\sqrt{\langle k_a,k_a\rangle}} = k_a\sqrt{1-|a|^2}.$$

We next prove that $\mathscr{D}$ is a dictionary of Hardy space $H^2(\mathbb{D})$. A dictionary [21] is defined as a family of parameterized vectors $\mathscr{G}=\{g_\gamma\}_{\gamma\in\Gamma}$ in a Hilbert space $H$ such that $\|g_\gamma\|=1$ and $\overline{\mathrm{span}}\mathscr{G}=H$. Each $g_\gamma\in\mathscr{G}$ is usually called an atom.

*Lemma 2.1*
The set $\mathscr{D}$ (1) is a dictionary of $H^2(\mathbb{D})$.

*Proof*
It is obvious that with $e_a\in H^2(\mathbb{D})$, $\|e_a\|_2=1$, and $\overline{\mathrm{span}}\mathscr{D}\subseteq H^2(\mathbb{D})$, we need only to show $\overline{\mathrm{span}}\mathscr{D}=H^2(\mathbb{D})$. For any $f\in H^2(\mathbb{D})$, $\langle f,e_a\rangle=\sqrt{1-|a|^2}f(a)$. Therefore, $\langle f,e_a\rangle=0$ implies $f(a)=0$, which yields $\overline{\mathrm{span}}\mathscr{D}^{\perp}=\{0\}$. So, we obtain that $\overline{\mathrm{span}}\mathscr{D}=H^2(\mathbb{D})$. □

Here, we state the following three lemmas that will be used in Section 3 .

*Lemma 2.2*
For any fixed point $a\in\mathbb{D}$, $\langle e_{\gamma a},e_{\mu a}\rangle=\langle e_{\overline{\mu}\gamma a},e_a\rangle=\langle e_a,e_{\gamma\overline{\mu}a}\rangle$ where $|\mu|=|\gamma|=1$.

*Proof*

$$\langle e_{\gamma a},e_{\mu a}\rangle = \frac{1-|a|^2}{1-\overline{\gamma}\mu|a|^2} = \langle e_{\overline{\mu}\gamma a},e_a\rangle = \langle e_a,e_{\gamma\overline{\mu}a}\rangle.$$

□

*Lemma 2.3*
For any $n\in\mathbb{N}, n\geq 0, r<1$,

$$\int_0^{2\pi}\frac{(1-r^2)e^{-in\theta}}{1-r^2e^{i\theta}}\frac{\mathrm{d}\theta}{2\pi} = r^{2n}(1-r^2).$$

*Proof*
Because $r<1$,

$$\begin{aligned}\int_0^{2\pi}\frac{(1-r^2)e^{-in\theta}}{1-r^2e^{i\theta}}\frac{\mathrm{d}\theta}{2\pi} &= (1-r^2)\int_0^{2\pi}\sum_{k=0}^{\infty}(r^2e^{i\theta})^k e^{-in\theta}\frac{\mathrm{d}\theta}{2\pi}\\ &= (1-r^2)\sum_{k=0}^{\infty}\int_0^{2\pi}(r^2e^{i\theta})^k e^{-in\theta}\frac{\mathrm{d}\theta}{2\pi}\\ &= (1-r^2)\sum_{k=0}^{\infty}r^{2k}\delta_{k,n}\\ &= (1-r^2)r^{2n}.\end{aligned}$$

In general, we have

$$\int_0^{2\pi} \frac{\sqrt{1-r_1^2}\sqrt{1-r_2^2}}{1-r_1r_2e^{i\theta}} e^{-in\theta} \frac{\mathrm{d}\theta}{2\pi} = r_1^n r_2^n \sqrt{1-r_1^2}\sqrt{1-r_2^2}$$

provided that $r_1 < 1, r_2 < 1$. □

*Lemma 2.4* (Ky Fan's maximum principle [22])
Let $A$ be any Hermitian operator, then for $k = 1, 2, \ldots, n$, we have

$$\sum_{j=1}^{k} \lambda_j(A) = \max \sum_{j=1}^{k} \langle Ax_j, x_j \rangle$$

where eigenvalues $\lambda_1(A) \geq \lambda_2(A) \geq \cdots \geq \lambda_n(A)$, and the maximum is taken over all orthonormal $k$-tuples $\{x_1, \ldots, x_k\}$.

We introduce some notations. Given an analytic signal $s \in H^2(\mathbb{D})$ and the dictionary $\mathscr{D}$, the representation problem has the form

$$s = \sum_{a \in \mathbb{D}} x_a e_a.$$

Nevertheless, all the continuous-time signals $s$ and $e_a$'s should be discretized because computers can only process discrete values. Let $T = \{t_k : 0 = t_1 < t_2 <, \cdots, < t_M = 2\pi, k = 1, 2, \ldots, M, \Delta t = t_{k+1} - t_k = 1/(M-1)\}$. For any $a \in \mathbb{D}$, we sample $e_a$ on $T$ to obtain an $M$-dimensional column vector $v_a$, namely,

$$v_a = \begin{pmatrix} e_a(t_1) & e_a(t_2) & \cdots & e_a(t_M) \end{pmatrix}^T. \tag{13}$$

Denote $\underline{e_a}$ as the normalized vector of $v_a$, that is, $\underline{e_a} = v_a / \|v_a\|$. Sample $s$ on $T$, we have

$$\underline{s} = \begin{pmatrix} s(t_1) & s(t_2) & \cdots & s(t_M) \end{pmatrix}^T. \tag{14}$$

Let $\underline{\mathscr{D}} \in \mathbb{C}^{M \times N}$ be the dictionary matrix of $\mathscr{D}$, viz.

$$\underline{\mathscr{D}} = \begin{pmatrix} \underline{e_{a_0}} & \underline{e_{a_1}} & \cdots & \underline{e_{a_{N-1}}} \end{pmatrix}. \tag{15}$$

Then, the representation problem in discrete-time situation can be written as

$$\underline{s} = \underline{\mathscr{D}} x \tag{16}$$

where $x$ is the vector of coefficients and $M < N$. Throughout this paper, Equation (16) is our basic model, from which two facts can be derived. One is that the more columns $\underline{\mathscr{D}}$ are present, the sparser representation follows . The other is the solutions of (16) are strongly related with the positions of parameters $a_0, \ldots, a_{N-1}$. Intuitively, we should select $\{a_k\}_{k=0}^{N-1}$ in some manner equally spaced to reflect the information of the whole unit circle. Besides, the singular values distribution should be analyzed in the sense of $N$ tending to infinity. Denote

$$\underline{H} = \underline{\mathscr{D}}^* \underline{\mathscr{D}} = \begin{pmatrix} \overline{\underline{e_{a_0}}} \\ \overline{\underline{e_{a_1}}} \\ \vdots \\ \overline{\underline{e_{a_{N-1}}}} \end{pmatrix} \begin{pmatrix} \underline{e_{a_0}} & \underline{e_{a_1}} & \cdots & \underline{e_{a_{N-1}}} \end{pmatrix}. \tag{17}$$

In Section 3, we will study the eigenvalues of $\underline{H}$, which are squares of the singular values of $\underline{\mathscr{D}}$.

## 3. Main results

Let $H$ be a Hermitian matrix with entries $H_{ij} = \langle e_{a_{j-1}}, e_{a_{i-1}} \rangle$; it is easy to verify

$$\langle \underline{e_{a_{j-1}}}, \underline{e_{a_{i-1}}} \rangle \to \langle e_{a_{j-1}}, e_{a_{i-1}} \rangle, \qquad (M \to \infty).$$

We use the eigenvalues of $H$ to estimate the eigenvalues of $\underline{H}$ because the eigenvalues of a Hermitian matrix depend continuously on its entries [22].

*Theorem 1*

Suppose $N$ points $\{a_k\}_{k=0}^{N-1}$ are distributed equally spaced on the circle of radius $r$, that is, $\mathrm{Arg}(a_k) = \theta_k = 2k\pi/N$ and $|a_k| = r$. $H$ is a Hermitian matrix with entries $H_{ij} = \langle e_{a_{j-1}}, e_{a_{i-1}} \rangle$, $i, j \in \{1, 2, \cdots, N\}$. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ be the eigenvalues of $H$. Then, we have

$$\lim_{N\to\infty} \frac{\sum_{d=1}^{l} \lambda_d}{N} \geq 1 - r^{2l}. \tag{18}$$

*Proof*

By Lemma 2.2, the entries of $H$ satisfy $H_{ij} = \langle e_{a_{j-1}}, e_{a_{i-1}} \rangle = \langle e_{a_0}, e_{a_{j-i}} \rangle = b_{j-i}$, where

$$b_k = \langle e_{a_0}, e_{a_k} \rangle = \frac{1-r^2}{1-r^2 e^{i\theta_k}}, \qquad (k = 0, 1, \ldots, N-1),$$

and $b_{-k} = \overline{b}_k$. Let $B$ be an upper triangular matrix,

$$B = \begin{pmatrix} b_0 & b_1 & b_2 & b_3 & b_4 & \cdots & b_{N-1} \\ 0 & b_0 & b_1 & b_2 & b_3 & \cdots & b_{N-2} \\ 0 & 0 & b_0 & b_1 & b_2 & \cdots & b_{N-3} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & b_0 \end{pmatrix}.$$

Then, $H = B + B^* - I_N$, where $I_N$ is the $N \times N$ identity matrix. □

Consider the orthonormal functions $\{e^{-int}\}_{n=0}^{l-1}$ in $L^2(0, 2\pi)$, discretize them into $l$ vectors by equally spaced sampling, namely

$$x_n = \frac{1}{\sqrt{N}} \begin{pmatrix} e^{-2\pi \frac{0}{N} i} & e^{-2\pi \frac{n}{N} i} & e^{-2\pi \frac{2n}{N} i} & \cdots & e^{-2\pi \frac{(N-1)n}{N} i} \end{pmatrix}^T.$$

For $n_1, n_2 \in \{0, 1, \ldots, l-1\}$, Riemann summation shows

$$\langle x_{n_1}, x_{n_2} \rangle_{\mathbb{C}^N} \to \langle e^{-in_1 t}, e^{-in_2 t} \rangle_{L^2} = \delta_{n_1, n_2}, \qquad (N \to \infty), \tag{19}$$

where $\langle \cdot, \cdot \rangle_{\mathbb{C}^N}$ is the dot product of complex Euclidean space and $\langle \cdot, \cdot \rangle_{L^2}$ stands for $L^2(0, 2\pi)$ inner product. We denote

$$\omega \triangleq e^{-2\pi \frac{1}{N} i} \qquad \text{and} \qquad \Omega_n \triangleq \langle Bx_n, x_n \rangle = x_n^* B x_n,$$

then

$$\Omega_n = \frac{1}{N} \begin{pmatrix} 1 & \omega^{-n} & \cdots & \omega^{-(N-1)n} \end{pmatrix} B \begin{pmatrix} 1 \\ \omega^n \\ \vdots \\ \omega^{(N-1)n} \end{pmatrix}.$$

Setting $b_k^n = b_k \omega^{nk}$, calculation gives that $\Omega_n$ is the $N$-th *Cesàro* mean of the sequence $\{S_m^n\}_{m=0}^{N-1}$, namely

$$\Omega_n = \frac{1}{N} \left( N b_0^n + (N-1) b_1^n + \cdots + b_{N-1}^n \right) = \frac{1}{N} \sum_{m=0}^{N-1} S_m^n,$$

where $S_m^n = \sum_{k=0}^{m} b_k^n$. Again, Riemann summation gives

$$\begin{aligned} \frac{2\pi}{N} \sum_{k=0}^{N-1} b_k^n &= \frac{2\pi}{N} \sum_{k=0}^{N-1} \frac{1-r^2}{1-r^2 e^{i\theta_k}} e^{-in\theta_k} \\ &= \frac{2\pi}{N} S_{N-1}^n \to \int_0^{2\pi} \frac{(1-r^2) e^{-in\theta}}{1-r^2 e^{i\theta}} \mathrm{d}\theta, \qquad (N \to \infty). \end{aligned}$$

From Lemma 2.3, we conclude that

$$\int_0^{2\pi} \frac{(1-r^2) e^{-in\theta}}{1-r^2 e^{i\theta}} \frac{\mathrm{d}\theta}{2\pi} = (1-r^2) r^{2n}.$$

Hence,

$$\frac{1}{N} S_{N-1}^n \to (1-r^2) r^{2n}, \qquad (N \to \infty).$$

It is not hard to prove

$$\frac{\Omega_n}{N} = \frac{1}{N^2}\sum_{m=0}^{N-1} S_m^n \to \frac{1}{2}(1-r^2)r^{2n}, \qquad (N\to\infty). \tag{20}$$

Because

$$\frac{1}{N}\langle Hx_n, x_n\rangle = \frac{1}{N}\left(\langle Bx_n, x_n\rangle + \langle B^* x_n, x_n\rangle - \langle x_n, x_n\rangle\right) = \frac{1}{N}\left(\Omega_n + \overline{\Omega}_n - 1\right),$$

then

$$\frac{1}{N}\sum_{n=0}^{l-1}\langle Hx_n, x_n\rangle = \frac{1}{N}\sum_{n=0}^{l-1}\left(\Omega_n + \overline{\Omega}_n - 1\right) \to \sum_{n=0}^{l-1}(1-r^2)r^{2n}, \qquad (N\to\infty).$$

By Ky Fan's maximum principle and (19), we complete the proof in the sense of taking limit,

$$\lim_{N\to\infty}\frac{1}{N}\sum_{d=1}^{l}\lambda_d \geq \lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{l-1}\langle Hx_n, x_n\rangle = \sum_{n=0}^{l-1}(1-r^2)r^{2n} = 1-r^{2l}.$$

*Remark 1.1*
Notice that trace$(H) = N$ and $\sum_{n=0}^{\infty}(1-r^2)r^{2n} = 1$. Theorem 1 actually states

$$\frac{\lambda_i}{\mathrm{trace}(H)} = \frac{\lambda_i}{N} \approx (1-r^2)r^{2(i-1)}, \qquad (i = 1, 2, \ldots, N), \tag{21}$$

when $N$ is sufficiently large. We estimate that $\lambda_i \approx (1-r^2)r^{2(i-1)}N$. Given $M$, $N$, and $r$, the numerical eigenvalues of $\underline{H}$ can be derived via a common software (e.g., MATLAB). In Figures 1 and 2, they fit perfectly even though $M$ and $N$ are not very large.

*Remark 1.2*
The equality (20) is a direct corollary of the following proposition, which can be proved easily. If $a_N/N \to a$, then $\sum_{i=1}^{N} a_i/N^2 \to a/2$.

*Theorem 2*
Suppose $N$ points $\{a_k\}_{k=0}^{N-1}$ are selected equally spaced on the interval $[0,1)$. $H$ is a Hermitian matrix with entries $H_{ij} = \langle e_{a_{j-1}}, e_{a_{i-1}}\rangle$, $i,j \in \{1,2,\ldots,N\}$. Let $\lambda_1$ be the largest eigenvalue of $H$, then we have

$$\lim_{N\to\infty}\frac{\lambda_1}{N} \geq \int_0^1\int_0^1 \frac{\sqrt{1-s^2}\sqrt{1-r^2}}{1-sr}drds \approx 0.8158. \tag{22}$$

*Proof*
Divide the interval $[0, 1-\delta]$ into $N$ parts with $a_0 = 0, a_1 = \frac{1-\delta}{N}, \cdots, a_{N-1} = \frac{(1-\delta)(N-1)}{N}$, where $\delta$ is a little positive number. The entries of $H$ satisfy

$$H_{ij} = \frac{\sqrt{1-a_{i-1}^2}\sqrt{1-a_{j-1}^2}}{1-a_{i-1}a_{j-1}}.$$

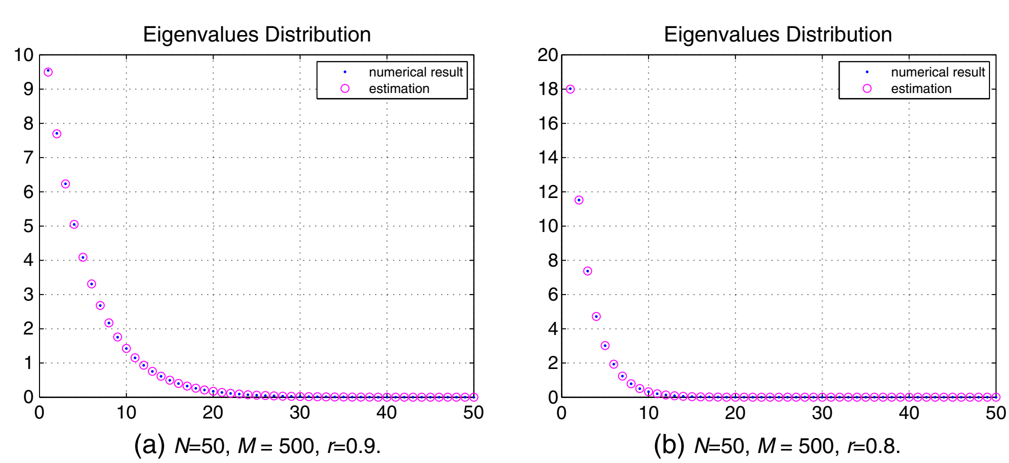



**Figure 1.** Eigenvalues of $\underline{H}$.

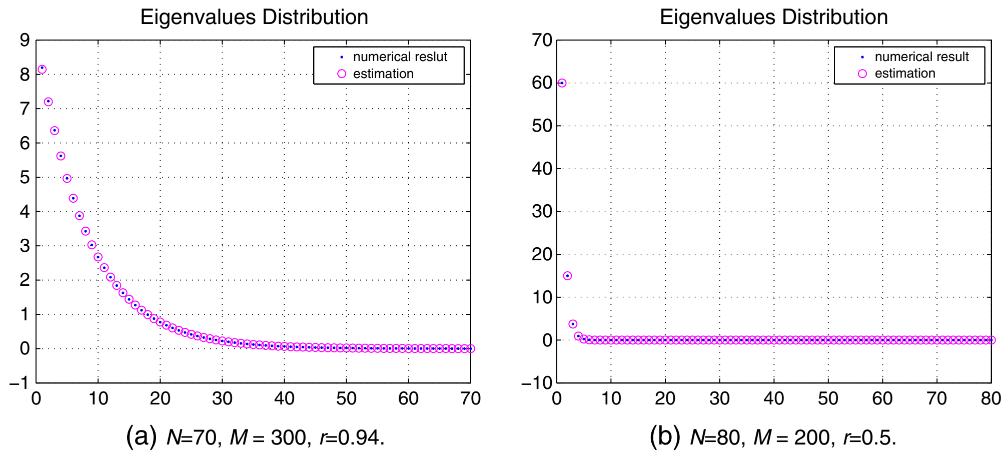



**Figure 2.** Eigenvalues of $\underline{H}$.

Let $x = \frac{1}{\sqrt{N}} \begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix}^T \in \mathbb{C}^N$, $\|x\|_2 = 1$, we have

$$\Omega = x^* H x = \frac{1}{N} \begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix} H \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix} = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} H_{ij}.$$

Riemann summation shows

$$\frac{(1-\delta)^2}{N^2} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\sqrt{1-a_i^2}\sqrt{1-a_j^2}}{1-a_i a_j} \to \int_0^{1-\delta} \int_0^{1-\delta} \frac{\sqrt{1-r^2}\sqrt{1-s^2}}{1-rs} \mathrm{d}r\mathrm{d}s \qquad (N \to \infty).$$

Consequently,

$$\frac{\Omega}{N} \to \frac{1}{(1-\delta)^2} \int_0^{1-\delta} \int_0^{1-\delta} \frac{\sqrt{1-r^2}\sqrt{1-s^2}}{1-rs} \mathrm{d}r\mathrm{d}s \qquad (N \to \infty).$$

Notice that the function

$$f(r,s) = \frac{\sqrt{1-r^2}\sqrt{1-s^2}}{1-rs}$$

is well defined in $[0,1) \times [0,1)$. In fact, $f$ is bounded on the region, which can be proved as follows. □

Taylor's series gives

$$\log(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - \ldots - \frac{x^n}{n} - \ldots \qquad (-1 < x \leq 1).$$

Rewrite $\sqrt{1-r^2}\sqrt{1-s^2}$ and $(1-rs)$ as

$$\begin{aligned} \sqrt{1-r^2}\sqrt{1-s^2} &= \exp\left(\log\sqrt{1-r^2} + \log\sqrt{1-s^2}\right), \\ &< \exp\left(\log\sqrt{1-r} + \log\sqrt{1-s} + \log 2\right), \\ &= 2\exp\left(\frac{1}{2}(\log(1-r) + \log(1-s))\right), \end{aligned}$$

and

$$1 - rs = \exp(\log(1-rs)).$$

Because

$$\log(1-r) + \log(1-s) = \sum_{k=1}^{\infty} \left(-\frac{r^k}{k} - \frac{s^k}{k}\right),$$

and

$$\frac{1}{2}(r^k + s^k) \geq \sqrt{r^k s^k} > r^k s^k, \qquad r, s \in (0, 1),$$

then we have

$$\frac{1}{2}(\log(1-r) + \log(1-s)) < -\sum_{k=1}^{\infty} \frac{r^k s^k}{k} = \log(1 - rs), \qquad r, s \in (0, 1).$$

Therefore,

$$\begin{aligned}\sqrt{1-r^2}\sqrt{1-s^2} &< 2\exp\left(\frac{1}{2}(\log(1-r) + \log(1-s))\right)\\ &< 2\exp(\log(1-rs)) = 2(1-rs), \qquad r, s \in (0, 1).\end{aligned}$$

That is,

$$0 \leq f(r, s) < 2, \qquad r, s \in [0, 1).$$

Obviously, $f(r, s) = 0$ if $r = 1, s \in [0, 1)$ or $s = 1, r \in [0, 1)$. Hence, $0 \leq f(r, s) < 2$ on $[0, 1] \times [0, 1] \backslash \{(1, 1)\}$, which means $f$ is Riemann integrable on $[0, 1] \times [0, 1]$.

Hence,

$$\lim_{\delta \to 0} \frac{1}{(1-\delta)^2} \int_0^{1-\delta} \int_0^{1-\delta} f(r, s) \mathrm{d}r\mathrm{d}s = \int_0^1 \int_0^1 f(r, s) \mathrm{d}r\mathrm{d}s.$$

Applying Ky Fan's maximum principle, we conclude

$$\lim_{N \to \infty} \frac{\lambda_1}{N} \geq \lim_{N \to \infty} \frac{\langle Hx, x\rangle}{N} = \lim_{N \to \infty} \frac{\Omega}{N} = \int_0^1 \int_0^1 f(r, s) \mathrm{d}r\mathrm{d}s \approx 0.8158.$$

*Remark 2.1*

In fact, an upper bound to $\lambda_1/N$ can be derived. It is known that

$$\lambda_1 = \sup_{\|x\|=1} \langle Hx, x\rangle,$$

where $x \in \mathbb{C}^N$. Then,

$$\lambda_1 = \sup_{\|x\|=1} \sum_{i=1}^{N} \sum_{j=1}^{N} \frac{\sqrt{1-a_{i-1}^2}\sqrt{1-a_{j-1}^2}}{1 - a_{i-1}a_{j-1}} \bar{x}_i x_j.$$

Let $g$ be a simple function corresponding to the vector $x$,

$$g(t) = \sqrt{N} \sum_{i=0}^{N-1} x_i I_{[t_i, t_{i+1})}(t),$$

where $t \in [0, 1)$ and $\Delta t = t_{i+1} - t_i = 1/(N-1)$. Hence,

$$\begin{aligned}\frac{\lambda_1}{N} &= \frac{1}{N} \sup_{\|x\|=1} \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \frac{\sqrt{1-a_i^2}\sqrt{1-a_j^2}}{1 - a_i a_j} \frac{\bar{g}_{t_i}}{\sqrt{N}} \frac{g_{t_j}}{\sqrt{N}}\\ &\to \sup_g \int_0^1 \int_0^1 f(r, s)\bar{g}(r)g(s) \mathrm{d}r\mathrm{d}s, \qquad (N \to \infty).\end{aligned}$$

Cauchy inequality implies that

$$\begin{aligned}\left(\int_0^1 \int_0^1 f(r, s)\overline{g}(r)g(s)\mathrm{d}r\mathrm{d}s\right)^2 &\leq \int_0^1 \int_0^1 f^2(r, s)\mathrm{d}r\mathrm{d}s \int_0^1 \int_0^1 |g(r)|^2 |g(s)|^2 \mathrm{d}r\mathrm{d}s\\ &= \int_0^1 \int_0^1 f^2(r, s)\mathrm{d}r\mathrm{d}s \left(\int_0^1 |g(r)|^2 \mathrm{d}r\right)^2\\ &= \int_0^1 \int_0^1 f^2(r, s)\mathrm{d}r\mathrm{d}s.\end{aligned}$$

So, we obtain

$$\lim_{N\to\infty} \frac{\lambda_1}{N} \le \left(\int_0^1 \int_0^1 f^2(r,s)drds\right)^{\frac{1}{2}} \approx 0.8427.$$

*Remark 2.2*

We proved that $0.8158 \le \lambda_1/N \le 0.8427$ when $N$ is reasonably large. Denote $\underline{\lambda_1}$ as the largest eigenvalue of $\underline{H}$, and it can be perfectly estimated by $\lambda_1$ (refer to Figures 3 and 4).

Lastly, we select the parameters $\{a_k\}_{k=0}^{N}$ equally spaced in $\mathbb{D}$ as follows. Let $\overrightarrow{r}$ and $\overrightarrow{\theta}$ be $N_1$-dimensional and $N_2$-dimensional vectors, respectively.

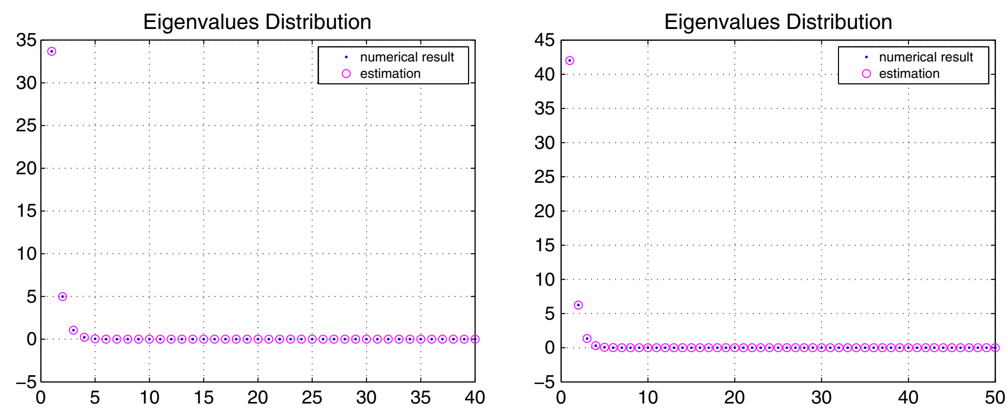
$$\overrightarrow{r} \triangleq \begin{pmatrix} 0 & \frac{1}{N_1} & \frac{2}{N_1} & \cdots & \frac{N_1-1}{N_1} \end{pmatrix},$$

$$\overrightarrow{\theta} \triangleq \begin{pmatrix} 1 & e^{\frac{2\pi i}{N_2}} & e^{\frac{2\pi 2i}{N_2}} & \cdots & e^{\frac{2\pi(N_2-1)i}{N_2}} \end{pmatrix}.$$

All the positions of $a$'s can be given by the tensor product of $\overrightarrow{r}$ and $\overrightarrow{\theta}$,

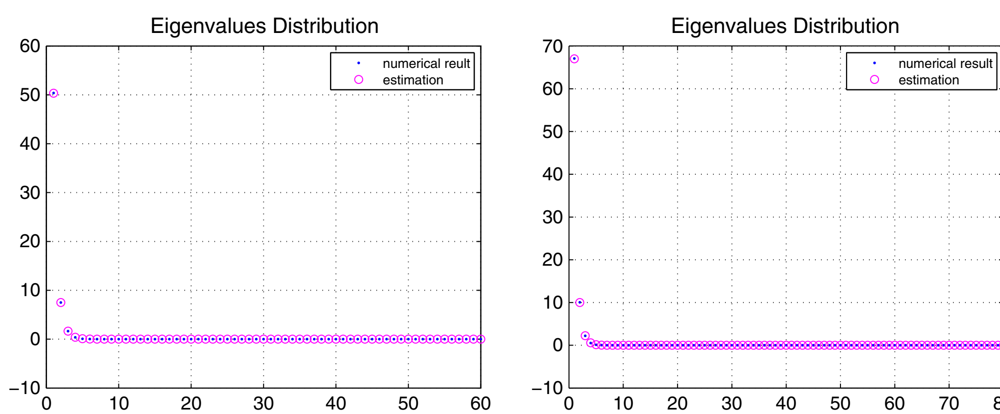
$$\overrightarrow{a} = \overrightarrow{r} \otimes \overrightarrow{\theta}.$$

Notice that the original point $a = (0,0) \in \overrightarrow{a}$ has a multiple of $N_2$, and we will count it $N_2$ times rather than one time for the sake of matrix computation. Thus, $\overrightarrow{a}$ is a row vector of $N = N_1 \times N_2$ entries, which are labeled from left to right by $a_0, a_1, \ldots, a_{N-1}$. It is clear that $\overrightarrow{a}$ has $N_1$ blocks, and all $a$'s in the same block are distributed equally spaced on the same circle.




(a) $N$=40, $M$ = 300. (b) $N$=50, $M$ = 200.

**Figure 3.** In (a), $\lambda_1 = 33.6819$, $\underline{\lambda_1} = 33.6865$, and $\lambda_1/N = 0.8420$. In (b), $\lambda_1 = 42.0148$, $\underline{\lambda_1} = 42.0288$, and $\lambda_1/N = 0.8403$.




(a) $N$=60, $M$ = 150. (b) $N$=80, $M$ = 120.

**Figure 4.** In (a), $\lambda_1 = 50.3464$, $\underline{\lambda_1} = 50.3790$, and $\lambda_1/N = 0.8391$. In (b), $\lambda_1 = 67.0077$, $\underline{\lambda_1} = 67.0829$, and $\lambda_1/N = 0.8385$.

*Theorem 3*
Suppose $N$ points $\{a_k\}_{k=0}^{N-1}$ are selected as previously discussed. Let $H$ be a Hermitian matrix with entries $H_{ij} = \langle e_{a_{j-1}}, e_{a_{i-1}}\rangle$, $i,j \in \{1,2,\ldots,N\}$. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ be eigenvalues of $H$, then we have

$$\lim_{\substack{N_1\to\infty \\ N_2\to\infty}} \frac{\sum_{k=1}^{l}\lambda_k}{N} \geq 1 - \frac{1}{2l+1}. \tag{23}$$

*Proof*
$H$ is a blocked matrix as follows:

$$H = \begin{pmatrix} B_1^*B_1 & B_1^*B_2 & B_1^*B_3 & \ldots & B_1^*B_{N_1} \\ B_2^*B_1 & B_2^*B_2 & B_2^*B_3 & \ldots & B_2^*B_{N_1} \\ B_3^*B_1 & B_3^*B_2 & B_3^*B_3 & \ldots & B_3^*B_{N_1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ B_{N_1}^*B_1 & B_{N_1}^*B_2 & B_{N_1}^*B_3 & \ldots & B_{N_1}^*B_{N_1} \end{pmatrix},$$

where each block $B_p^*B_q \in \mathbb{C}^{N_2\times N_2}$, $p,q \in \{1,2,\cdots,N_1\}$. Denote that

$$\vec{\theta}^{\,n} \triangleq \frac{1}{\sqrt{N_2}}\left(\begin{matrix} 1 & \left(e^{\frac{2\pi i}{N_2}}\right)^n & \left(e^{\frac{2\pi 2i}{N_2}}\right)^n & \cdots & \left(e^{\frac{2\pi(N_2-1)i}{N_2}}\right)^n \end{matrix}\right)^T, \qquad (n \geq 0), \tag{24}$$

and

$$r_p = \vec{r}(p) = \frac{p-1}{N_1},$$

then we have

$$\frac{1}{N_2}\left(\vec{\theta}^{\,n}\right)^*\left(B_p^*B_q\right)\vec{\theta}^{\,n} \to \int_0^{2\pi} \frac{\sqrt{1-r_p^2}\sqrt{1-r_q^2}e^{-in\theta}}{1-r_pr_qe^{i\theta}}\frac{d\theta}{2\pi}, \qquad (N_2\to\infty),$$

consequently,

$$\int_0^{2\pi} \frac{\sqrt{1-r_p^2}\sqrt{1-r_q^2}}{1-r_pr_qe^{i\theta}}e^{-in\theta}\frac{d\theta}{2\pi} = r_p^nr_q^n\sqrt{1-r_p^2}\sqrt{1-r_q^2}.$$

□

We define a family of normalized function in $L^2(0,2\pi)$ as

$$f_m(r) = \frac{r^m\sqrt{1-r^2}}{\|r^m\sqrt{1-r^2}\|}, \qquad (m \geq 0),$$

and a family of vectors as

$$\vec{f_m} = \frac{1}{\sqrt{N_1}}\left(\begin{matrix} f_m(r_1) & f_m(r_2) & f_m(r_3) & \cdots & f_m(r_{N_1}) \end{matrix}\right)^T. \tag{25}$$

Let

$$R = \begin{pmatrix} \vdots & \vdots & \vdots \\ \vdots & r_p^mr_q^m\sqrt{1-r_p^2}\sqrt{1-r_q^2} & \vdots \\ \vdots & \vdots & \vdots \end{pmatrix}_{N_1\times N_1},$$

then we can obtain

$$
\begin{aligned}
\frac{1}{N_1}\left(\vec{f_m}\right)^* R\left(\vec{f_m}\right) &= \frac{1}{N_1}\sum_{p=1}^{N_1}\sum_{q=1}^{N_1} r_p^m r_q^m \sqrt{1-r_p^2}\sqrt{1-r_q^2}\vec{f_m}(p)\vec{f_m}(q)\\
&\to \int_0^1\int_0^1 r^m s^m \sqrt{1-r^2}\sqrt{1-s^2} f_m(r) f_m(s) \mathrm{d}r\mathrm{d}s\\
&= \left(\int_0^1 r^m\sqrt{1-r^2} f_m(r)\mathrm{d}r\right)^2\\
&= \int_0^1 r^{2m}(1-r^2)\mathrm{d}r\\
&= \frac{1}{2m+1}-\frac{1}{2m+3}.
\end{aligned}
$$

Therefore,

$$
\frac{1}{N}\left(\vec{f_k}\otimes\vec{\theta}^{\,k}\right)^* H\left(\vec{f_k}\otimes\vec{\theta}^{\,k}\right) \to \frac{1}{2k+1}-\frac{1}{2k+3}, \tag{26}
$$

as $N_1 \to \infty$, $N_2 \to \infty$. Notice that

$$
\left\langle \vec{f_{k_1}}\otimes\vec{\theta}^{\,k_1}, \vec{f_{k_2}}\otimes\vec{\theta}^{\,k_2}\right\rangle \to \delta_{k_1,k_2}, \qquad (N_1\to\infty, N_2\to\infty), \tag{27}
$$

hence, by Ky-Fan's maximal principle, in the sense of taking limits, we have

$$
\lim_{\substack{N_1\to\infty,\\ N_2\to\infty}} \frac{\sum_{k=1}^{l}\lambda_k}{N} \ge \sum_{k=0}^{l-1}\left(\frac{1}{2k+1}-\frac{1}{2k+3}\right) = 1-\frac{1}{2l+1}. \tag{28}
$$

*Remark 3.1*
It is clear that $\operatorname{trace}(H) = N_1N_2 = N$ and

$$
\sum_{k=0}^{\infty}\left(\frac{1}{2k+1}-\frac{1}{2k+3}\right) = 1,
$$

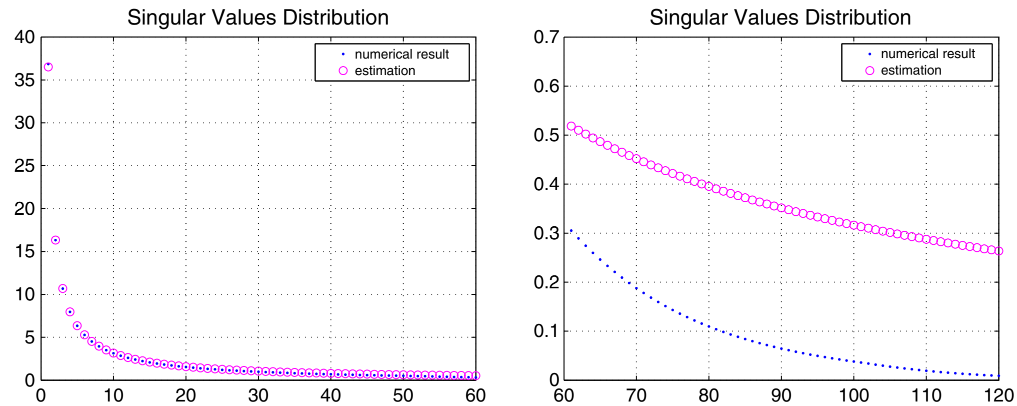
this theorem shows that

$$
\lambda_k \approx \left(\frac{1}{2k-1}-\frac{1}{2k+1}\right)\operatorname{trace}(H) = \frac{2N}{4k^2-1}, \qquad (k\ge 1), \tag{29}
$$

when $N$ is sufficiently large. Thus, we estimate the singular values as

$$
\sigma_k = \sqrt{\lambda_k} \approx \sqrt{\frac{2N}{4k^2-1}}, \qquad (k \ge 1). \tag{30}
$$

Given $N_1$, $N_2$, and $M$, we can calculate the singular values of $\mathscr{D}$, which are denoted by $\underline{\sigma_k}$. $\sigma_k$ in (30) is a quite sharp estimation of $\underline{\sigma_k}$ (Figure 5).




(a) $N_1$=40, $N_2 = 50$, $M = 400$. (b) $N_1$=40, $N_2 = 50$, $M = 400$.

**Figure 5.** $\underline{\sigma_k}$ and $\sigma_k$. (a) $1 \le k \le 60$ and (b) $61 \le k \le 120$.

## 4. Numerical examples

In this section, we give two numerical examples exhibiting sparse representations of analytic signals in $H^2(\mathbb{D})$. Because complex-valued signals are not numerically friendly in the sense of linear programming, we consider the complex signal as a real signal combining its real part and the imaginary part. That is, from (8), we set

$$s^r = \begin{pmatrix} \mathrm{Re}\lfloor U^* s \rfloor_K \\ \mathrm{Im}\lfloor U^* s \rfloor_K \end{pmatrix} \tag{31}$$

and

$$A_K^r = \begin{pmatrix} \mathrm{Re}A_K & -\mathrm{Im}A_k \\ \mathrm{Im}A_K & \mathrm{Re}A_k \end{pmatrix}. \tag{32}$$

Then, the equation

$$\lfloor U^* \underline{s} \rfloor_K = \mathrm{diag}(\sigma_1, \cdots, \sigma_K) \lfloor V^* \rfloor_K \tag{33}$$

is equivalent to

$$s^r = A_K^r \begin{pmatrix} \mathrm{Re}(x) \\ \mathrm{Im}(x) \end{pmatrix} = A_K^r y. \tag{34}$$

We solve

$$\min \|y\|_1 \quad \text{subject} \quad \text{to} \quad s^r = A_K^r y. \tag{35}$$

The dictionary matrix has been described in Theorem 3; we set $N_1 = 50$, $N_2 = 60$, and $M = 1000$ for $\mathscr{D}$. Furthermore, Theorem 3 states that $K = \mathcal{O}\left(\sqrt{N}\right) = \mathcal{O}\left(\sqrt{3000}\right) \approx 55c$. In the following examples, we deal with two cases with respect to $c = 1, 2$. Numerical result shows that sparse representations can be obtained by $\ell_1$-minimization even $K \ll M$.

The CS technique is also utilized in sparse recovery. Let $\Phi \in R^{n\times 2M}$ be a Gaussian random matrix satisfying $\Phi_{ij} \sim \mathcal{N}\left(0, \frac{1}{n}\right)$. We solve

$$\min \|y\|_1 \quad \text{subject} \quad \text{to} \quad \Phi \begin{pmatrix} \mathrm{Re}(\underline{s}) \\ \mathrm{Im}(\underline{s}) \end{pmatrix} = \Phi \begin{pmatrix} Re\mathscr{D} & -Im\mathscr{D} \\ Im\mathscr{D} & Re\mathscr{D} \end{pmatrix} y \tag{36}$$

to derive a sparse representation.

### 4.1. *Example 1*

$$s(z) = \frac{0.247z^4 + 0.0355z^3}{0.3329z^2 - 1.2727z + 1}$$

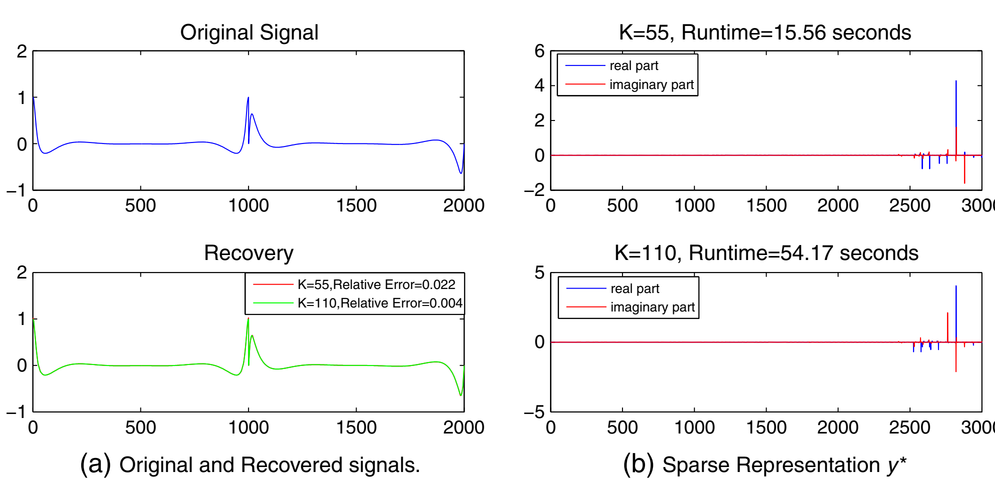
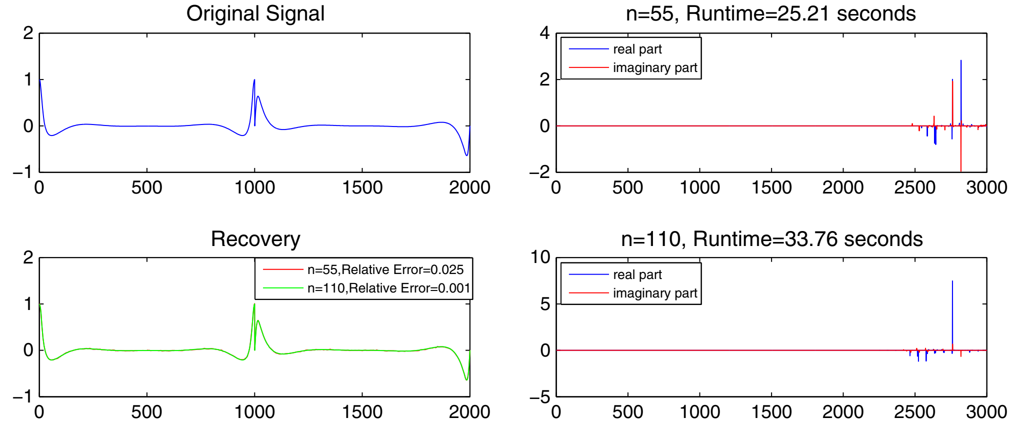
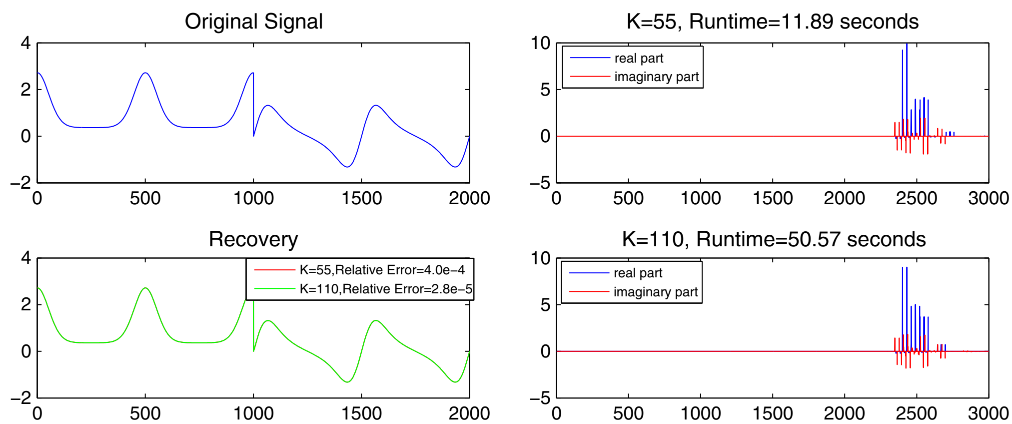



**Figure 6.** Basis pursuit recovery of Example 1.



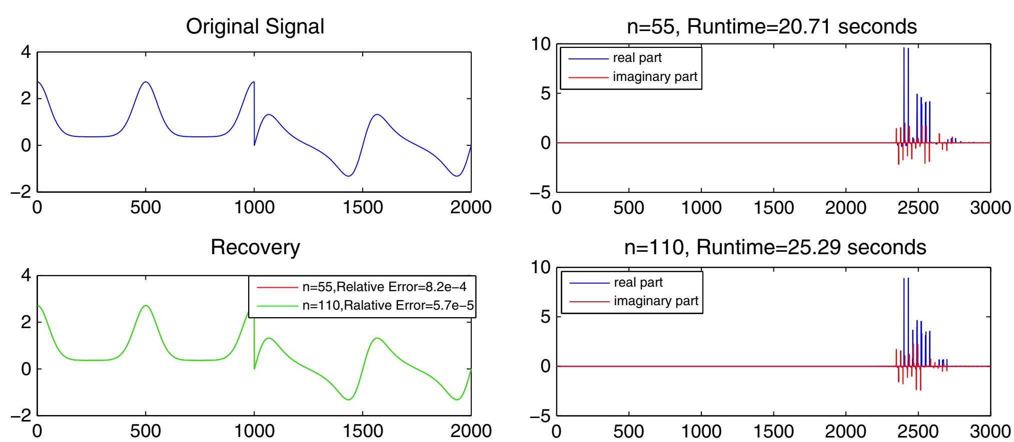

(a) Original and Recovered signals. (b) Sparse Representation $y^*$

**Figure 7.** Compressed sensing recovery of Example 1.




(a) Original and Recovered signals. (b) Sparse Representation $y^*$

**Figure 8.** Basis pursuit recovery of Example 2.




(a) Original and Recovered signals. (b) Sparse Representation $y^*$

**Figure 9.** Compressed sensing recovery of Example 2.

We sample $s$ to obtain a vector $\underline{s}$ of length $M = 1000$ as (14). Choose $K = \sqrt{3000} \approx 55$ and $K = 2\sqrt{3000} \approx 110$, respectively. The SVD of $\underline{\mathscr{D}}$ gives $U$ and $V$. Solving (35) and (36), we derive the optimal solution $y^*$, which is a sparse vector. The original signal $\underline{s}$ can be recovered by

$$\begin{pmatrix} \mathrm{Re}\underline{\mathscr{D}} & -\mathrm{Im}\underline{\mathscr{D}} \\ \mathrm{Im}\underline{\mathscr{D}} & \mathrm{Re}\underline{\mathscr{D}} \end{pmatrix} y^* \tag{37}$$

as shown in Figures 6 and 7.

*4.2. Example 2*

$$s(z) = e^{z^2}. \tag{38}$$

We do the same thing as in Example 1, as shown in Figures 8 and 9.

In conclusion, our dictionary does give sparse representations of analytic signals in $H^2(\mathbb{D})$, so the CS technique works almost as well as BP, and CS takes much less time.

## References

1. Qian T, Wang YB. Adaptive decomposition into basic signals of non-negative instantaneous frequencies—a variation and realization of greedy algorithm. *Advances in Computational Mathematics* 2011; **34**(3):279–293.
2. Dang P, Qian T. Analytic phase derivatives, all-pass filters and signals of minimum phase. *IEEE Transactions on Signal Processing* 2011; **59**(10):4708–4718.
3. Qian T, Zhang LM, Li ZX. Algorithm of adaptive Fourier decomposition. *IEEE Transactions on Signal Processing* 2011; **59**(12):5899–5906.
4. Mi W, Qian T. Frequency domain identification: an algorithm based on adaptive rational orthogonal system. *Automatica* 2012; **48**:1154–1162.
5. Davis G, Mallat S. Adaptive greedy approximations. In *Constructive Approximation*, Vol. 13. Springer-Verlag New York Inc., 1997; 57–98.
6. Davis G. Adaptive nonlinear approximations. *PhD thesis*, Department of Mathematics, Courant Institute of Mathematical Sciences, New York University, 1994.
7. Mallat S. *A Wavelets Tour of Signal Processing*. Academic Press: 525 B Street, Suite 1900, San Diego, California 92101-4495, USA, 2009.
8. Chen S, Donoho D, Saunders M. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing* 1999; **20**(1):33–61.
9. Donoho D, Elad M. On the stability of the basis pursuit in the presence of noise. *Signal Processing* 2006; **86**(3):511–532.
10. Donoho D, Huo X. Uncertainty principles and ideal atomic decomposition. *IEEE Transactions on Information Theory* 2001; **47**(7):2845–2852.
11. Knockaert L, Backer B, Zutter D. SVD compression, unitary transforms and computational complexity. *IEEE Transactions on Signal Processing* 1999; **47**(10):2724–2729.
12. Candes E, Wakin M, Boyd S. Enhancing sparsity by reweighted $\ell_1$ mimimization. *Journal of Fourier Analysis and Applications* 2008; **14**:877–905.
13. Kashin B. The widths of certain finite dimensional sets and classes of smooth functions. *Izvestia* 1977; **41**:334–351.
14. Cohen A, Dahmen W, Devore R. Compressed sensing and best *k*-term approximation. *Journal of the American Mathematical Society* 2009; **22**:211–231.
15. Candes E, Tao T. Near optimal signal recovery from random projections: universal encoding stategies? *IEEE Transactions on Information Theory* 2006; **52**(12):5406–5425.
16. Candes E, Romberg J, Tao T. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory* 2006; **52**(2):489–509.
17. Candes E, Tao T. Decoding by linear programming. *IEEE Transactions on Information Theory* 2005; **51**(12):4203–4215.
18. Donoho D. Compressed sensing. *IEEE Transactions on Information Theory* 2006; **52**(4):1289–1306.
19. Rauhut H, Schnass K, Vandergheynst P. Compressed sensing and redundant dictionaries. *IEEE Transactions on Information Theory* 2008; **54**(5): 2210–2219.
20. Candes E, Eldar Y, Needell D, Randall P. Compressed sensing with coherent and redundant dictionaries. *Applied and Computational Harmonic Analysis* 2011; **31**:59–73.
21. Mallat S, Zhang Z. Matching pursuit with time-frequency dictionaries. *IEEE Transactions on Signal Processing* 1993; **41**:3397–3415.
22. Bhatia R. *Matrix Analysis*. Spinger-Verlag: Springer-Verlag New York Inc., 1997.