International Journal of Wavelets, Multiresolution
and Information Processing
Vol. 11, No. 3 (2013) 1350031 (14 pages)

DOI: 10.1142/S0219691313500318

SPARSE RECONSTRUCTION OF HARDY SIGNAL AND APPLICATIONS TO TIME-FREQUENCY DISTRIBUTION

TAO QIAN*, SHUANG LI† and WEIXIONG MAI‡

Department of Mathematics, University of Macau, Macau
*fsttq@umac.mo
†ya97418@umac.mo
‡maiweixiong@gmail.com

Received 29 January 2013
Revised 26 April 2013
Accepted 26 April 2013
Published

In this paper, we introduce a sparse recovery strategy for analytic signals in Hardy space $H^2(\mathbb{D})$, where $\mathbb{D}$ denotes the unit disk of the complex plane. The representation strategy is based on the optimization technique. We investigate the asymptotic singular values distribution of the dictionary matrix and give an estimation of the number of rows of the random matrix. To the best of our knowledge, this is the first time that such result is given. This result demonstrates that the dictionary of the normalized Szegö kernels (or reproducing kernels) is perfect for decompositions of analytic signals. A numerical example is presented exhibiting the theory. As applications, we still work on time-frequency analysis and propose a new type of non-negative time-frequency distribution associated with mono-components in the periodic case.

Keywords: Hardy space; singular value; analytic signal; sparse representation; time-frequency distribution.

AMS Subject Classification: 41A20, 30J99, 65E99, 42A50

1. Introduction

One of the problems of approximation theory is to approximate functions with elements from a large candidate set called a dictionary. Let H be a Hilbert space. Using terminology introduced by Mallat and Zhang,[15] a dictionary is defined as a family of parameterized vectors $\mathscr{D} = \{g_\gamma\}_{\gamma\in\Gamma}$ in H such that $\|g_\gamma\| = 1$ and $\overline{\text{a family}}$ $\overline{\mathrm{span}(g_\gamma)} = H$, the $g'_\gamma s$ are usually called atoms. For the discrete-time situation, the

*Corresponding author.

approximation problem can be written as

$$s = \mathscr{D}x \tag{1.1}$$

where s is the discrete signal, matrix $\mathscr{D}$ represents the dictionary with atoms as columns and x is the vector of coefficients. In general, $\mathscr{D}$ has more columns than rows because of its redundancy. A natural question is: can we find the best M-term approximation in a redundant dictionary for a given signal? That is an optimization problem

$$\min \|s - \mathscr{D}x\|_{l^2} \quad \text{subject to } \|x\|_{l^0} \leq M, \tag{1.2}$$

where $\|x\|_{l^0}$ is the number of nonzero coefficients of x. Unfortunately, finding an optimal M-term approximation in dictionaries is computationally intractable because it is NP-hard.[8,14] Until now, three main strategies have been mainly investigated, they are matching pursuit, basis pursuit and compressed sensing.

Matching pursuit (MP) introduced by Mallat and Zhang computes signal approximations from a redundant dictionary by iteratively selecting one vector at a time. If the dictionary is orthogonal, the method works perfectly. If the dictionary is not orthogonal, the situation is less clear.[5] The MP algorithm often yields locally optimal solutions depending on initial values. In contrast, basis pursuit (BP) perform a more global search. It finds signal representations by solving the following problem

$$\min \|x\|_{l^1} \quad \text{subject to } s = \mathscr{D}x. \tag{1.3}$$

Given s and $\mathscr{D}$, we find x with minimal l_1 norm. Basis pursuit is an optimization principle, not an algorithm. Empirical evidence suggests that BP is more powerful than MP.[5] And the stability of BP has been proven in the presence of noise for sufficiently sparse representation.[10] BP is closely connected with convex programming. The interior-point method and the homotopy method can be applied to BP in nearly linear time.[5] Compressed Sensing (CS) is a new concept in sparse representation. The ideas have their origins in certain abstract results by Kashin[6] but were brought into the forefront by the work of Candes, Romberg and Tao[3,2,4,16,12] and Donoho.[9,11] Basically, CS relies on random projection and BP. Suppose we have

$$y = \Phi x, \tag{1.4}$$

where x is a finite vector, Φ is observation matrix and y is the vector of available measurements. Then the BP solution x^* of

$$\min \|x\|_{l^1} \quad \text{subject to } y = \Phi x \tag{1.5}$$

recovers x exactly provided that x is sufficiently sparse and the matrix obeys the Restricted Isometry Property (RIP).[3,4,20,13] However the RIP of a fixed matrix is very hard to check, thus in practice we use random matrices instead. A Gaussian matrix $\Phi \in R^{m\times N}$ whose entries $\Phi_{i,j}$ are independent and follow a normal distribution with expectation 0 and variance $1/m$ is often adopted. To know more about the detailed technique, please see related paper in references.[3,2,4,9,11]

This paper originates from a series of recent results on analytic signal decomposition by Qian *et al.* where the concept of Adaptive Fourier Decomposition (AFD) was introduced.[19,7] AFD is a variation of MP, and it yields an approximation using a few elements chosen adaptively from the set of normalized reproducing kernels

$$\mathscr{D} = \left\{ d_a : d_a(z) = \frac{\sqrt{1-|a|^2}}{1-\overline{a}z}, a \in \mathbb{D} \right\}, \tag{1.6}$$

where $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$. One of the motivations of AFD research is finding time-frequency distributions of signals. In Refs. 17 and 18, Qian has given full study of decomposing Hardy signal into Mono-components and time-frequency analysis.

As continuation of AFD related studies, this paper analyzes the singular values of the dictionary matrix. We derive the asymptotic distribution of singular values in the sense that the number of atoms goes to infinity and study the CS based recovery. Then we use CS technique to derive a sparse representation of Hardy signal. As applications, we also discuss time-frequency distribution for analytic signal.

The paper is organized as follows. Preliminaries and notations are given in Sec. 2. The main results are proved in Sec. 3. A numerical example is presented in Sec. 4. Applications to time-frequency distribution are given in Sec. 5.

2. Preliminaries and Notations

Let $\mathbb{D} = \{z \in \mathbb{C} : |z| < 1\}$. Hardy space $H^2(\mathbb{D})$ is defined as

$$\left\{ f \in Hol(\mathbb{D}) : \sup_{0 \le r < 1} \int_0^{2\pi} \frac{|f(re^{it})|^2 dt}{2\pi} < \infty \right\}, \tag{2.1}$$

where $Hol(\mathbb{D})$ denotes the space of holomorphic (or analytic) functions on the unit disk $\mathbb{D}$. Or equivalently

$$\{f \in L^2(0, 2\pi) : \widehat{f}(n) = 0, n < 0\}, \tag{2.2}$$

where $\widehat{f}(n)$ denotes the n-th Fourier coefficient. $H^2(\mathbb{D})$ is a Hilbert space with inner product

$$\langle f, g \rangle = \frac{1}{2\pi} \int_0^{2\pi} f(e^{it}) \bar{g}(e^{it}) dt. \tag{2.3}$$

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

$$\mathscr{K} = \left\{ k_a : k_a(z) = \frac{1}{1-\overline{a}z}, a \in \mathbb{D} \right\}, \tag{2.4}$$

which gives $f(a) = \langle f, k_a \rangle$. Note that d_a in (1.6) satisfies

$$d_a = \frac{k_a}{\|k_a\|} = \frac{k_a}{\sqrt{\langle k_a, k_a \rangle}} = k_a \sqrt{1-|a|^2}. \tag{2.5}$$

In this letter, $\mathscr{D}$ plays a fundamental role because it forms a dictionary of $H^2(\mathbb{D})$.

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

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

In discrete-time system, we envision a decomposition of a signal $s \in H^2(\mathbb{D})$ as

$$\underline{s} = \mathcal{D}x, \tag{2.6}$$

where $\underline{s}$ is an M-dimensional discrete signal, $\mathcal{D}$ is an $M \times N$ dictionary matrix and x is a vector of coefficients. The signal $\underline{s}$ and columns $\{d_i\}_{i=1}^N$ are derived by sampling equally-spaced from s and respective d_a from $\mathscr{D}$. For simplicity, we normalize each d_i, i.e. $d_i = d_i/\|d_i\|_2$. The parameters $a_1, \ldots, a_N$ of N columns are selected as follows. Let $\overrightarrow{r}$ and $\overrightarrow{\theta}$ be N_1-dimensional and N_2-dimensional vectors respectively, i.e.

$$\overrightarrow{r} = \left(0\,\frac{1}{N_1}\,\frac{2}{N_1}\,\cdots\,\frac{N_1-1}{N_1}\right) \tag{2.7}$$

$$\overrightarrow{\theta} = (1\, e^{\frac{2\pi i}{N_2}}\, e^{\frac{2\pi 2i}{N_2}} \cdots e^{\frac{2\pi(N_2-1)i}{N_2}}). \tag{2.8}$$

Denote N-dimensional vector $\overrightarrow{a}$ as the tensor product of $\overrightarrow{r}$ and $\overrightarrow{\theta}$, i.e. $\overrightarrow{a} = \overrightarrow{r} \otimes \overrightarrow{\theta}$, $N = N_1 N_2$, let $a_i = \overrightarrow{a}(i+1)$, $i = 0, \ldots, N-1$.

Lemma 2.2. *SVD (Singular Value Decomposition). Let A denote an arbitrary matrix and $\{s_i\}$ be the singular values of A. Then A can be represented in the form*

$$A = UDV^H, \tag{2.9}$$

where U and V are unitary and matrix D has s_i in the (i,i) position.

Proof. For proof, please refer to Ref. 1. $\square$

A way to approximate the signal $\underline{s}$ is to project $\underline{s}$ onto the space spanned by the singular vectors corresponding to the largest several singular values of $\mathcal{D}$. Suppose that SVD of $\mathcal{D}$ gives $\mathcal{D} = U\Sigma V^H$, where U and V are unitary matrices, $\{\sigma_i\}_{i=1}^M$ is the set of singular values with $\sigma_1 \geq \cdots \geq \sigma_M \geq 0$, and $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_M)$. Then we have $U^H\underline{s} = \Sigma V^H x$. Thus the transformed signal $U^H\underline{s}$ is an approximately sparse signal provided that the singular values in Σ decay rapidly. In this case, compressed sensing theory states that $U^H\underline{s}$ can be recovered with overwhelming probability by solving

$$\min \|z\|_1 \quad y = \Phi(U^H\underline{s}) = \Phi z, \tag{2.10}$$

where Φ is an $\alpha \times M$ random matrix satisfying $\mathbb{E}\|\Phi x\|^2 = \|x\|^2$ and $\mathbb{P}(|\|\Phi x\|^2 - \|x\|^2| \geq \varepsilon\|x\|^2) \leq 2e^{-\alpha c(\varepsilon)}$,[2,4] and

$$\alpha \geq C \log\left(\frac{M}{\|U^H \underline{s}\|_0}\right) \|U^H \underline{s}\|_0. \tag{2.11}$$

Then using Basis Pursuit method,[5] we get a sparse representation x of (2.6) with overwhelming probability by solving

$$\min \|x\|_1 \quad AU^H \underline{s} = A\Sigma V^H x. \tag{2.12}$$

Notice that the random matrix AU^H also satisfies $\mathbb{E}\|AU^H x\|^2 = \|x\|^2$ and $\mathbb{P}(|\|AU^H x\|^2 - \|x\|^2| \geq \varepsilon\|x\|^2) \leq 2e^{-\alpha c(\varepsilon)}$ since U is unitary. Consequently, we can solve

$$\min \|x\|_1 \quad A\underline{s} = AU\Sigma V^H x = A\mathcal{D}x \tag{2.13}$$

instead of (2.12) to derive a sparse representation x if the singular values decay fast. Hence, we have to analyze the singular values of $\mathcal{D}$, which are the square roots of the eigenvalues of $\mathcal{D}^H\mathcal{D}$. Two facts should be mentioned here. One is that the more columns $\mathcal{D}$ has, the sparser representation follows. The other is the solutions of (2.6) 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 equally-spaced manner 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.

Riemann sum of the integral shows that the entries of $\mathcal{D}^H\mathcal{D}$ satisfy

$$\lim_{M\to\infty} (\mathcal{D}^H\mathcal{D})_{ij} = \lim_{M\to\infty} \langle d_j, d_i\rangle_E = \langle d_{a_j}, d_{a_i}\rangle_{H^2}, \tag{2.14}$$

where $\langle\cdot,\cdot\rangle_E$ stands for Euclidean inner product and $\langle\cdot,\cdot\rangle_{H^2}$ means Hardy space inner product. Indeed, we have

$$\lim_{M\to\infty} \frac{2\pi}{M-1} \sum_{k=1}^{M} d_{a_j}(k)\overline{d_{a_i}}(k) = \int_0^{2\pi} d_{a_j}(e^{it})\overline{d_{a_i}}(e^{it})dt. \tag{2.15}$$

So, $\langle d_j, d_i\rangle_E/(\|d_j\| \cdot \|d_i\|) = \langle d_j, d_i\rangle_E \to \langle d_{a_j}, d_{a_i}\rangle_{H^2}$ as M goes to infinity.

Let H be a Hermitian matrix with entries $H_{ij} = \langle d_{a_j}, d_{a_i}\rangle_{H^2}$. The eigenvalues of H can be used to estimate the eigenvalues of $\mathcal{D}^H\mathcal{D}$. This can be proven as follow. Denoting $\Delta = \mathcal{D}^H\mathcal{D} - H$, by (2.14) we assert $\Delta_{ij} \to 0$ as $M \to \infty$. Thus for any $\epsilon > 0$, we have $|\Delta_{ij}| < \epsilon/N$ given that M is reasonably large. Let $D_i = \{z : |z - \Delta_{ii}| \leq \Sigma_{j\neq i}|\Delta_{ij}|\}, 1 \leq i \leq N$. Then Gersgorin Disk theorem[1] shows, $\lambda_k(\Delta) \in \bigcup_{i=1}^N D_i$ where $\lambda_k(\Delta)$ is the eigenvalues of Δ, $1 \leq k \leq N$. It is clear that $D_i \subseteq E_i = \{z : |z| \leq \Sigma_j|\Delta_{ij}|\}$. However E_i represents a disk with center 0 and radius $r_i = \Sigma_j|\Delta_{ij}| < \Sigma_j\epsilon/N = \epsilon$. Therefore, $\lambda_k(\Delta) < \epsilon$, $1 \leq k \leq N$. On the other hand, Weyl's inequalities[1] say that $\lambda_k(\mathcal{D}^H\mathcal{D}) \leq \lambda_k(H) + \lambda_1(\Delta)$ and $\lambda_k(\mathcal{D}^H\mathcal{D}) \geq \lambda_k(H) + \lambda_N(\Delta)$, where $\lambda_N \leq \lambda_{N-1} \leq \cdots \leq \lambda_1$. Thus we finally derive that $\lambda_k(H) - \epsilon \leq \lambda_k(\mathcal{D}^H\mathcal{D}) \leq \lambda_k(H) + \epsilon$ if the number of sampling points M is reasonably large. In summary, the eigenvalues of $\mathcal{D}^H\mathcal{D}$ can be estimated by the eigenvalues of H. It is well-known that the spectrum of a Hermitian matrix is real and the eigenvectors associated with distinct eigenvalues are orthogonal. Moreover, there exists an orthonormal basis consisting entirely of normalized eigenvectors. In the next section, we analyze the asymptotic eigenvalues of H by constructing the associated asymptotic eigenvectors. The following simple lemma is needed.

Lemma 2.3. *For any fixed* $a \in \mathbb{D}$, $\langle d_{\gamma a}, d_{\mu a}\rangle = \langle d_{\overline{\mu}\gamma a}, d_a\rangle = \langle d_a, d_{\gamma\overline{\mu}a}\rangle$ *where* $|\mu| = |\gamma| = 1$.

Proof.

$$\langle d_{\gamma a}, d_{\mu a}\rangle = \frac{1-|a|^2}{1-\overline{\gamma}\mu|a|^2} = \langle d_{\overline{\mu}\gamma a}, d_a\rangle = \langle d_a, d_{\gamma\overline{\mu}a}\rangle. \qquad \square$$

3. Asymptotic Eigenvalues

Theorem 3.1. *Denote N-dimensional column vectors as*

$$v_n = \frac{1}{\sqrt{N}}(1\ e^{-2\pi in/N}\ \cdots\ e^{-2\pi i(N-1)n/N})^T, \quad n \geq 0. \tag{3.1}$$

Let H be an $N \times N$ Hermitian matrix with entries

$$H_{ml} = \frac{\alpha}{1-\beta e^{i(l-m)2\pi/N}}, \tag{3.2}$$

where α and β are real constants and $\beta < 1$. Then we have

$$\lim_{N\to\infty}\frac{1}{N}\langle Hv_n, v_n\rangle_E = \alpha\beta^n. \tag{3.3}$$

Proof. Let B be an upper triangular matrix with $B_{ml} = H_{ml}$ for $l \geq m$. Then $H = B + B^H - (\alpha/(1-\beta))I_N$ where I_N is the $N \times N$ identity matrix. We set $\delta = 2\pi/N$,

$$\delta_k = k\delta, \quad b_k = \frac{\alpha}{1-\beta e^{i\delta_k}}, \quad k = 0, 1, \ldots, N-1. \tag{3.4}$$

Denote $\tau = e^{-i\delta}$. One derives that

$$\begin{aligned}\langle Bv_n, v_n\rangle_E &= v_n^H Bv_n \\ &= \frac{1}{N}(Nb_0 + (N-1)b_1\tau^n + (N-1)b_2\tau^{2n} + \cdots + b_{N-1}\tau^{(N-1)n}) \\ &= \frac{1}{N}\sum_{m=0}^{N-1}\sum_{k=0}^{m} b_k\tau^{nk}. \end{aligned} \tag{3.5}$$

Riemann sum of the integral gives

$$\begin{aligned}\lim_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N-1} b_k\tau^{nk} &= \lim_{N\to\infty}\frac{1}{N}\sum_{k=0}^{N-1}\frac{\alpha}{1-\beta e^{i\delta_k}}e^{-in\delta_k} \\ &= \frac{1}{2\pi}\int_0^{2\pi}\frac{\alpha}{1-\beta e^{it}}e^{-int}dt\end{aligned}$$

$$
\begin{aligned}
&= \frac{\alpha}{2\pi}\int_0^{2\pi}\sum_{k=0}^{\infty}\beta^k e^{i(k-n)t}dt \\
&= \alpha\beta^n. \qquad (3.6)
\end{aligned}
$$

Denoting $S_m = \sum_{k=0}^{m} b_k \tau^{nk}$, one can derive from (3.6) that

$$\lim_{N\to\infty}\frac{1}{N}S_{N-1} = \alpha\beta^n. \qquad (3.7)$$

On the other hand, by Stolzs theorem, (3.5) gives

$$
\begin{aligned}
\frac{1}{N}\langle Bv_n, v_n\rangle_E &= \frac{1}{N^2}(S_0 + S_1 + \cdots + S_{N-1}) \\
&= \frac{1}{N^2}\sum_{m=0}^{N-1} S_m \to \frac{1}{2}\alpha\beta^n \qquad (3.8)
\end{aligned}
$$

as $N \to \infty$. Then we conclude that

$$
\begin{aligned}
\frac{1}{N}\langle Hv_n, v_n\rangle_E &= \frac{1}{N}\left(v_n^H Bv_n + v_n^H B^H v_n - \frac{\beta}{1-\gamma}\|x_n\|^2\right) \\
&= \frac{1}{N}\left(v_n^H Bv_n + \overline{v_n^H Bv_n} - \frac{\beta}{1-\gamma}\right) \\
&\to \frac{1}{2}\alpha\beta^n + \frac{1}{2}\alpha\beta^n = \alpha\beta^n \qquad (3.9)
\end{aligned}
$$

as $N \to \infty$, which completes the proof. □

Remark 3.1. Notice that

$$\langle d_{a_1}, d_{a_2}\rangle_{H^2} = \frac{\sqrt{1-|a_1|^2}\sqrt{1-|a_2|^2}}{1-\bar{a}_1 a_2}.$$

If parameters $\{a_k\}_{k=0}^{N-1}$ are distributed on the circle of radius r as follows

$$a_k = r\exp\left(\frac{2\pi ik}{N}\right), \qquad (3.10)$$

where $r < 1$, then we can obtain a Hermitian matrix H with entries

$$H_{ml} = \langle d_{a_{l-1}}, d_{a_{m-1}}\rangle_{H^2}. \qquad (3.11)$$

By Lemma (2.3),

$$H_{ml} = \langle d_{a_{l-1}}, d_{a_{m-1}}\rangle_{H^2} = \langle d_{a_0}, d_{a_{l-m}}\rangle_{H^2} = \frac{1-r^2}{1-r^2e^{i(l-m)\delta}}, \qquad (3.12)$$

where $\delta = 2\pi/N$. It is clear that Trace$(H) = N$. Theorem 1 actually states that

$$\langle Hv_n, v_n\rangle \approx \text{Trace}(H)\times(1-r^2)r^n, \quad n \geq 0, \qquad (3.13)$$

AQ: Do you mean Theorem 3.1?

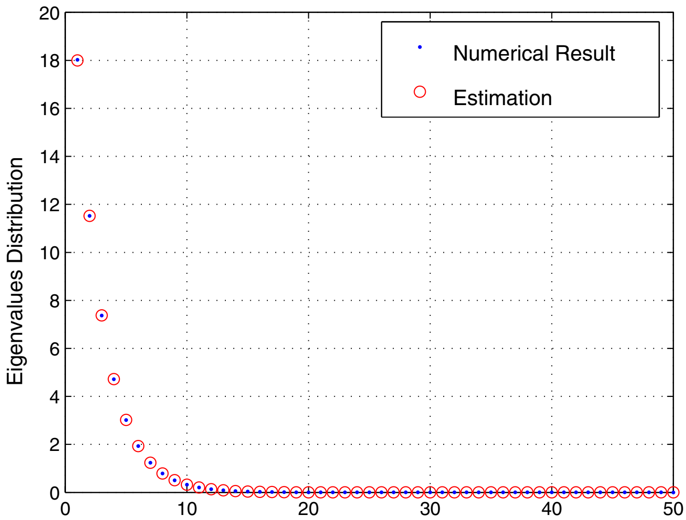


Fig. 1. $N = 50$, $r = 0.8$. Numerical eigenvalues of (3.11) are perfectly estimated by (3.13) even though N is not very large.

when N is sufficiently large. On the other hand, it is easy to verify

$$\sum_{n=0}^{N-1}(1-r^2)r^n = 1 - r^N \approx 1. \tag{3.14}$$

Hence, $\{v_n\}_{n\geq 0}$ are the asymptotic eigenvectors associated with the asymptotic eigenvalues $N(1-r^2)r^n$ as N tends to infinity. The eigenvalues of H in (3.11) decay as a geometry series. See Fig. 1.

Theorem 3.2. *Let H be a Hermitian matrix with entries $H_{ij} = \langle d_{a_{j-1}}, d_{a_{i-1}}\rangle_{H^2}$, $i, j \in \{1, 2, \ldots, N\}$. Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ be eigenvalues of H. Then we have*

$$\lim_{\substack{N_1\to\infty \\ N_2\to\infty}} \frac{\lambda_k}{N_1 N_2} = \frac{1}{2k-1} - \frac{1}{2k+1}. \tag{3.15}$$

Proof. H is a blocked matrix:

$$H = \begin{pmatrix} B_1^H B_1 & B_1^H B_2 & \ldots & B_1^H B_{N_1} \\ B_2^H B_1 & B_2^H B_2 & \ldots & B_2^H B_{N_1} \\ \vdots & \vdots & \vdots & \vdots \\ B_{N_1}^H B_1 & B_{N_1}^H B_2 & \ldots & B_{N_1}^H B_{N_1} \end{pmatrix} \tag{3.16}$$

where each block $B_p^H B_q$ is an $N_2 \times N_2$ matrix, $p, q \in \{1, 2, \ldots, N_1\}$. The entries of $B_p^H B_q$ satisfy

$$(B_p^H B_q)_{ml} = \frac{\sqrt{1-r_p^2}\sqrt{1-r_q^2}}{1 - r_p r_q e^{i(l-m)\delta}}, \tag{3.17}$$

where $\delta = 2\pi/N_2$, $r_p = \overrightarrow{r}(p)$ and $m, l \in \{1, 2, \ldots, N_2\}$. Denote that

$$\overrightarrow{\theta}^n = \frac{1}{\sqrt{N_2}}(1\, e^{-i\delta n} \cdots e^{-i\delta(N_2-1)n})^T, \quad n \geq 0.$$

AQ: Do you mean Theorem 3.1?

By Theorem 1, we derive:

$$\frac{1}{N_2}(\overrightarrow{\theta}^n)^H (B_p^H B_q)\overrightarrow{\theta}^n = \frac{1}{N_2}\langle (B_p^H B_q)\overrightarrow{\theta}^n, \overrightarrow{\theta}^n\rangle_E \to \alpha\beta^n \tag{3.18}$$

as $N_2 \to \infty$, where

$$\alpha = \sqrt{1-r_p^2}\sqrt{1-r_q^2}, \quad \beta = r_p r_q. \tag{3.19}$$

Define a family of normalized functions as following

$$f_n(r) = \frac{r^n\sqrt{1-r^2}}{\|r^n\sqrt{1-r^2}\|_{L^2(0,1)}}, \quad (n \geq 0) \tag{3.20}$$

and a family of column vectors

$$\overrightarrow{f_n} = \frac{1}{\sqrt{N_1}}(f_n(r_1)\ f_n(r_2)\cdots f_n(r_{N_1}))^T. \tag{3.21}$$

Let R be an $N_1 \times N_1$ matrix with entries

$$R_{pq} = r_p^n r_q^n \sqrt{1-r_p^2}\sqrt{1-r_q^2}, \tag{3.22}$$

then we obtain that

$$\begin{aligned}
\frac{(\overrightarrow{f_n})^H R(\overrightarrow{f_n})}{N_1} &= \frac{\langle R(\overrightarrow{f_n}), \overrightarrow{f_n}\rangle}{N_1} \\
&= \frac{1}{N_1}\sum_{p=1}^{N_1}\sum_{q=1}^{N_1} r_p^n r_q^n \sqrt{1-r_p^2}\sqrt{1-r_q^2}\overrightarrow{f_n}(p)\overrightarrow{f_n}(q) \qquad (3.23)\\
&\to \int_0^1\int_0^1 r^n s^n \sqrt{1-r^2}\sqrt{1-s^2} f_n(r) f_n(s)\,dr\,ds \\
&= \left(\int_0^1 r^n\sqrt{1-r^2} f_n(r)dr\right)^2 = \int_0^1 r^{2n}(1-r^2)dr \\
&= \frac{1}{2n+1} - \frac{1}{2n+3} \qquad (3.24)
\end{aligned}$$

as $N_1 \to \infty$. Therefore, tensor product gives

$$\frac{1}{N_1 N_2}(\overrightarrow{f_n} \otimes \overrightarrow{\theta}^n)^H H(\overrightarrow{f_n} \otimes \overrightarrow{\theta}^n) \to \frac{1}{2n+1} - \frac{1}{2n+3} \tag{3.25}$$

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

$$\langle \overrightarrow{f_{n_1}} \otimes \overrightarrow{\theta}^{n_1}, \overrightarrow{f_{n_2}} \otimes \overrightarrow{\theta}^{n_2} \rangle \to \delta_{n_1,n_2}, (N_1 \to \infty, N_2 \to \infty), \tag{3.26}$$

i.e. $\overrightarrow{f_n} \otimes \overrightarrow{\theta}^n$ $(n \geq 0)$ are asymptotic orthogonal vectors. In other words, the vectors $\overrightarrow{f_n} \otimes \overrightarrow{\theta}^n / \sqrt{N_1 N_2}$ are asymptotic eigenvectors associated with asymptotic eigenvalues $N_1 N_2(1/(2n+1) - 1/(2n+3))$ in the sense of N tending to infinity, $n \geq 0$. Hence, from (3.25) we have

$$\lim_{\substack{N_1 \to \infty, \\ N_2 \to \infty}} \frac{\lambda_k}{N_1 N_2} = \frac{1}{2(k-1)+1} - \frac{1}{2(k-1)+3}. \tag{3.27}$$

□

Remark 3.2. Since $\sum_{k=1}^{\infty}(\frac{1}{2k-1} - \frac{1}{2k+1}) = 1$ and $\mathrm{Trace}(H) = N_1 N_2 = N$, Theorem 2 actually shows that

$$\lambda_k \approx \left(\frac{1}{2k-1} - \frac{1}{2k+1}\right) \mathrm{Trace}(H) = \frac{2N}{4k^2 - 1}, \tag{3.28}$$

when N is sufficiently large, $k \geq 1$. Thus we estimate the singular values of H as

$$\sigma_k = \sqrt{\lambda_k} \approx \sqrt{\frac{2N}{4k^2 - 1}}. \tag{3.29}$$

The singular values of $\mathcal{D}$ are also estimated by (3.29) since $\lambda_k(H)$ and $\lambda_k(\mathcal{D}^H \mathcal{D})$ can be arbitrarily near, provided that M is sufficiently large. See Fig. 2. They fit very well even M, N_1, N_2 are not large.

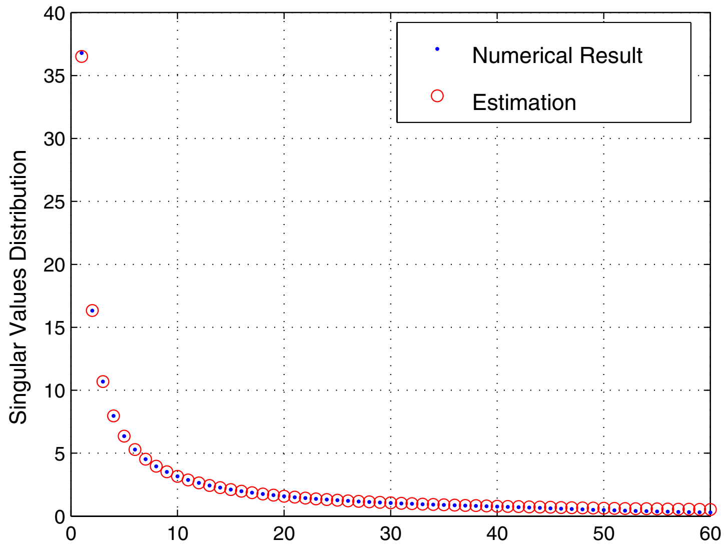


Fig. 2. $N_1 = 50$, $N_2 = 40$, $M = 300$. Blue points stand for the largest 60 singular values of $\mathcal{D}$ and circles stand for our estimation from (3.29). We have made a sharp estimation even M is not very large.

4. Numerical Example

In Sec. 3, we have derived a sharp estimation of asymptotic singular values of $\mathcal{D}$, i.e. $\sigma_k \approx \sqrt{\frac{2N}{4k^2-1}} \approx \frac{1}{k}\sqrt{\frac{N}{2}}$. Hence the support of the transformed vector $U^H \underline{s}$ in Sec. 2 is a set of cardinality $\mathcal{O}(\sqrt{N})$. Singular values tend to zero rapidly with order $\mathcal{O}(\frac{1}{k})$. We conclude from the theory of compressed sensing that

$$\alpha \geq C\sqrt{N} \log\left(\frac{M}{\sqrt{N}}\right), \tag{4.1}$$

where α is the number of rows of the Gaussian matrix and C is a universal constant. We give an example to illustrate our recovery algorithm. Let $s \in H^2(\mathbb{D})$ be an analytic signal such that

$$s = \frac{0.0247z^4 + 0.0355z^3}{(1-0.9048z)(1-0.3679z)}. \tag{4.2}$$

For the dictionary matrix $\mathcal{D}$, we set $M = 1000$, $N_1 = 50$ and $N_2 = 60$. Then $N = 3000$. Let $\alpha = 110 \approx 2\sqrt{3000}$, we get an $\alpha \times M$ Gaussian matrix A with entries $A_{ij} \sim N(0, 1/110)$. See Figs. 3 and 4. The relative error is 0.0013 and runtime is 16.59s.

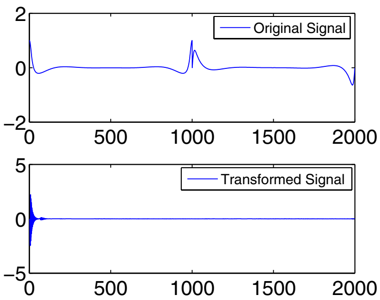


Fig. 3. The original signal $\underline{s}$ and transformed signal $U^H\underline{s}$ are plotted. Both of them are 1000-dimensional complex vectors or equivalently 2000-dimensional real vectors. It can be observed that $U^H\underline{s}$ is a really sparse signal due to the rapid decay of singular values of $\mathcal{D}$.

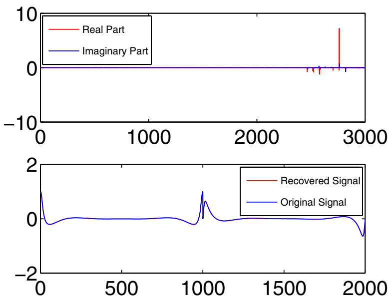


Fig. 4. The optimal solution of (2.13) and the recovered signal are illustrated. The dictionary (1.6) does give sparse representations of analytic signals.

5. Time-Frequency Distribution

In this section, we deal with time-frequency distribution of analytic signal using our method. A signal $s(t) = \rho(t)e^{i\theta(t)}$ is a complex mono-component signal, if $\theta^{'}(t) > 0$ and

$$H(\rho(t)e^{i\theta(t)}) = -i\rho(t)e^{i\theta(t)}, \tag{5.1}$$

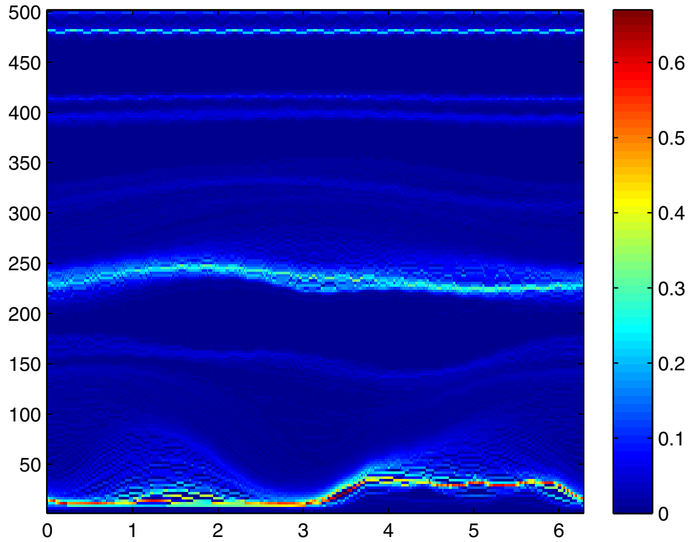
where $\rho(t) > 0$ and $H(\cdot)(x) = \int_0^{2\pi} cot(\frac{x-t}{2})(\cdot)dt$ denotes the Hilbert transform on the circle. The Transient Time Frequency Distribution (TTFD) of a mono-component signal $s(t) = \rho(t)e^{i\theta(t)}$ is defined by

$$P(t,\xi) = \rho^2(t)\delta_M(\xi - \theta'(t)), \quad (t,\xi) \in \mathbb{R} \times \left[-\frac{1}{2M}, +\infty\right),$$

where

$$\delta_M(\xi - \theta'(t)) = \begin{cases} M, & \text{if } \xi \in \left[\theta'(t) - \frac{1}{2M}, \theta'(t) + \frac{1}{2M}\right], \\ 0, & \text{if } \xi \notin \left[\theta'(t) - \frac{1}{2M}, \theta'(t) + \frac{1}{2M}\right], \end{cases}$$

where M is a large enough positive number to be determined in practice. When M goes to infinity, δ_M becomes the distributional Dirac function. It is a convenience and practical application that we make M to be a finite number. In fact, finite Blaschke products which are orthogonal functions under Gram-Schimitt procedure of the dictionary (1.6) give common examples of mono-components.



AQ: Please provide Fig. 5 citation in text

Fig. 5. Time-Frequency Distribution.

If $s(t)$ is a multi-component signal, then through our method $s(t)$ can be decomposed into a sum of mono-component signals

$$s(t) = \sum_{k=1}^{\infty} s_k(t) = \sum_{k=1}^{\infty} c_k B_k(e^{it}) = \sum_{k=1}^{\infty} \rho_k(t) e^{i\theta_k(t)} \tag{5.2}$$

then the corresponding TTFD is defined as

$$P(t,\xi) = \sum_{k=1}^{\infty} P_k(t,\xi) = \sum_{k=1}^{\infty} \rho_k^2(t)\delta_M(\xi - \theta_k'(t)), \quad (t,\xi) \in \mathbb{R} \times \left[-\frac{1}{2M}, +\infty\right) \tag{5.3}$$

where $P_k(t,\xi)$ is the TTFD of $s_k(t)$.

Time-frequency example:

$$s(t) = \begin{cases} \cos(10t), & \text{if } t < \pi, \\ \cos(30t), & \text{if } t \geq \pi. \end{cases}$$

Acknowledgments

The authors would like to show their deepest gratitude to Prof. Song Li and Prof. Luoqing Li for providing their papers and manuscripts. The ideas in these documents are very helpful.

References

1. R. Bhatia, *Matrix Analysis* (Springer-Verlag, 1997).
2. E. Candes, J. Romberg and T. Tao, Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information, *IEEE. Trans. Information Theory* **52**(2) (2006) 489–509.
3. E. Candes, J. Romberg and T. Tao, Stable signal recovery from incomplete and inaccurate measurments, *Comm. Pure Appl. Math.* **59**(8) (2006) 1207–1223.
4. E. Candes and T. Tao, Decoding by linear programming, *IEEE. Trans. Inf. Theory* **51**(12) (2005) 4203–4215.
5. S. Chen, D. Donoho and M. Saunders, Atomic decomposition by basis pursuit, *SIAM J. Sci. Comput.* **43**(1) (2001) 129–159.
6. A. Cohen, W. Dahmen and R. DeVore, Compressed sensing and best k-term approximation, *J. Amer. Math. Soc.* **22** (2009) 211–231.
7. P. Dang and T. Qian, Analytic phase derivatives, all-pass filters and signals of minimum phase, *IEEE Trans. Signal Processing* **59**(10) (2011) 4708–4718.
8. G. Davis, S. Mallat and M. Avellaneda, Adaptive greedy approximations, *Constr. Approx.* **13**(1) (1997) 57–98.
9. D. Donoho, Compressed sensing, *IEEE. Trans. Inf. Theory* **52**(4) (2006) 1289–1306.
10. D. Donoho and M. Elad, On the stability of the basis pursuit in the presence of noise, *Signal Processing* **86**(3) (2006) 511–532.
11. D. Donoho and Y. Tsaig, Extensions of compressed sensing, *Signal Processing* **86**(3) (2006) 549–571.
12. J. Lin and S. Li, Nonuniform support recovery from noisy random measurements by orthogonal matching pursuit, *J. Approx. Theory* **165**(1) (2013) 20–40.

13. Y. Liu, T. Mi and S. Li, Compressed sensing with general frames via optimal-dual-based ℓ_1-analysis, *IEEE Trans. Inf. Theory* **58**(7) (2012) 4201–4214.
14. S. Mallat, *A Wavelets Tour of Signal Processing* (Academic Press, 2009).
15. S. Mallat and Z. Zhang, Matching pursuit with time-frequency dictionaries, *IEEE Trans. Signal Process.* **41**(12) (1993) 3397–3415.
16. Q. Mo and S. Li, New bounds on the restricted isometry constant δ_{2k}, *Appl. Comput. Harmon. Anal.* **31**(3) (2011) 460–468.
17. T. Qian, Characterization of boundary values of functions in Hardy spaces with application in signal analysis, *J. Integral Equations and Applications* **17**(2) (2005) 159–198.
18. T. Qian, Mono-components for decomposition of signals, *Mathematical Methods in the Applied Sciences* **29**(10) (2006) 1187–1198.
19. T. Qian, L. M. Zhang and Z. X. Li, Algorithm of adaptive fourier decomposition, *IEEE Trans. on Signal Processing* **59**(2) (2011) 5899–5906.
20. Y. Shen and S. Li, Restricted p-isometry property and its application for nonconvex compressive sensing, *Adv. Comput. Math.* **37**(3) (2012) 441–452.