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 ℓ_1-minimization are used.

Keywords: Hardy space; singular value; reproducing kernels; ℓ_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 ℓ_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

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

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

In general, we have

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

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 $k = 1, 2, \dots, n$, we have

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

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

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 $T = \{t_k : 0 = t_1 < t_2 <, \dots, < t_M = 2\pi, k = 1, 2, \dots, M, \Delta t = t_{k+1} - t_k = 1/(M-1)\}$. 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 $a_0, \dots, a_{N-1}$. 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

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

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].

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

I-x1n6734/F41.40155Tm0D(n)Tj/F6.8Tf8.9664008.9664331.68695.i08.9664218.755Tf6.9738006.97388..073688.87D-.1

$$\frac{1}{N}\sum_{n=0}^{l-1}\langle Hx_n, x_n\rangle = \frac{1}{N}\sum_{n=0}\Big($$

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

So we obtain

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

$$\vec{\theta}^{\,n} \triangleq \frac{1}{\sqrt{}}$$

$\big)$

then we can obtain

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

Therefore,

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

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

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

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

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

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 $\underline{\mathscr{D}}$. 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 ℓ_1-minimization even $K \ll M$.

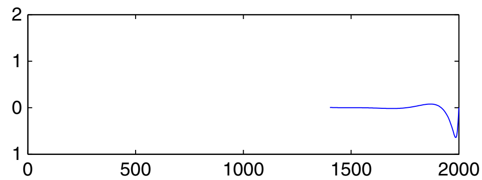
The CS technique is also utilized in sparse recovery. Let $\Phi \in R^{n \times 2M}$ be a Gaussian random matrix satisfying $\Phi_{ij} \sim \mathcal{N}\left(0, \frac{1}{n}\right)$. We solve

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

to derive a sparse representation.

4.1. Example 1

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


2
1
0
1
0
500
1000
1500
2000

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