Proceedings of the Project Review, Geo-Mathematical Imaging Group (Purdue University, West Lafayette IN), Vol. 1 (2009) pp. 1-26.

SPARSE APPROXIMATION OF FUNCTIONS USING SUMS OF EXPONENTIALS

MARCUS CARLSSON*

Abstract. We consider the problem of approximating a function $A \in C_0(\mathbb{R}^+)$ by a sum of few exponentials functions $\{e^{i\zeta_k x}\}_k$, where $\mathrm{Im}\zeta_k > 0$. The difficulty lies in finding an efficient algorithm to calculate suitable location of the "nodes" $\{\zeta_k\}$. We improve the algorithm invented by Beylkin and Monzón [3], develop a theory that explains their numerical observations, as well as introduce another algorithm in the same spirit that is more suitable for approximation with respect to $L^2(\mathbb{R}^+)$.

1. Introduction. Let $R \in \mathbb{R}^+ \cup \{\infty\}$ be fixed and let $x : \mathbb{R}^+ \to \mathbb{R}^+$ be the identity function $x(t) = t$. This paper is concerned with the following problem. Given a continuous function A on the interval $[0, R)$ and a certain accuracy $\epsilon > 0$, we wish to find a small number $n \in \mathbb{N}$, "nodes" $\zeta_k \in \mathbb{C}$ and "weights" c_k, $(1 \leqslant k \leqslant n)$, such that

$$\|A(x) - \sum_{k=1}^{n} c_k e^{i\zeta_k x}\| \leqslant \epsilon. \tag{1.1}$$

Clearly, the problem also depends heavily on which norm is used in (1.1). We will consider different norms as we go along. In particular L^2, BMO, and certain operator norms related to L^∞ will be discussed.[1]

The theoretical part of this paper is motivated by applications, and therefore we do not strive to find the smallest number n for which (1.1) holds, (given a fixed A and ϵ). Instead, we want to provide the theoretical framework to build stable approximation-algorithms which yield a number n which is small compared to other compression algorithms.

This paper builds on results by G. Beylkin and L. Monzón in [3]. It improves their results in the sense that it provides the theoretical framework to understand the phenomena observed numerically by Beylkin and Monzón, and extends their results in the sense that we invent new algorithms more suitable for approximation in L^2. We present two algorithms. One that is similar to the one by Beylkin and Monzón but has the advantage that the location of the "nodes" are known to lie in the unit disc $\mathbb{D}$, which makes it faster to execute. The second algorithm yields estimates for approximation with respect to L^2.

The paper is outlined as follows. In Section 2 we present the algorithm by Beylkin and Monzón and in 3 we present our first algorithm. Sections 4, 5 and 6 provide the theoretical framework that allows us to understand the algorithm in 3. In Section 6 we also motivate that Beylkin and Monzón's numerical observations likely are consequences of Sections 4 and 5 as well. Finally, in Section 7 we introduce the other algorithm which yields estimates for approximation with respect to L^2.

2. Review of the algorithm by Beylkin and Monzón. The algorithm by G. Beylkin and L. Monzón in [3] deals with (1.1) in the case of the finite interval $[0, 1]$. They point out that their numerical results are far better than the theoretical results indicate. To clarify the contribution of the present paper, we therefore have to explicitly point out some weaknesses in [3], which should not be taken as criticism.

For $p \geqslant 1$ and any $N \in \mathbb{N}$ let the sampling operator $\mathcal{S}_{p,N} : C([0, 1]) \to \mathbb{C}^{N+1}$ be defined by

$$\mathcal{S}_{p,N} A = \left(\frac{1}{(N)^{1/p}} A\left(\frac{k}{N}\right) \right)_{k=0}^{N}. \tag{2.1}$$

*Department of Mathematics, Purdue University, West Lafayette, IN 47907

[1]For any $p \geqslant 1$, L^p will always be defined with respect to standard Lebesgue measure on some subinterval of $\mathbb{R}^+$ that is determined by the context.

The factor $1/(N)^{1/p}$ is natural for considering approximations in L^p, because clearly

$$\|A\|_{L^p} = \lim_{N\to\infty} \|S_{p,N}\|_{l^p}. \tag{2.2}$$

Whenever p, N are clear from the context, we will simply write $\mathcal{S}$.

In [3] it is not entirely clear which norm is intended in (1.1). On one hand, the problem is stated for $L^\infty([0,1])$ and this norm is also used in the examples, whereas the L^2-norm (as well as a certain operator norm to be specified below) is considered in the proofs. Clearly any solution for (1.1) with respect to L^∞ is also a solution with respect to L^2, so we will here focus on the latter norm. Their method goes as follows. Let $C([0,1])$ denote the space of continuous functions on $[0,1]$. Given an $A \in C([0,1])$, pick an $N \in \mathbb{N}$ such that $\mathcal{S}(A) = \mathcal{S}_{\infty,2N}(A)$ is "sufficiently oversampled". The next step is to form the finite Hankel matrix

$$\gamma_{\mathcal{S}A} = \begin{pmatrix} \mathcal{S}A(0) & \mathcal{S}A(1) & \cdots & \mathcal{S}A(N) \\ \mathcal{S}A(1) & \mathcal{S}A(2) & \cdots & \mathcal{S}A(N+1) \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{S}A(N) & \mathcal{S}A(N+1) & \cdots & \mathcal{S}A(2N) \end{pmatrix}$$

and find its so called "con-eigenvectors" and "con-eigenvalues" in the Takagi factorization of $\gamma_{\mathcal{S}A}$, that is, we find $u_0, \ldots, u_N \in \mathbb{C}^{N+1}$ and $\sigma_0 \geqslant \sigma_1 \geqslant \ldots \geqslant \sigma_N \geqslant 0$ such that

$$\gamma_{\mathcal{S}A} u_k = \sigma_k \overline{u_k}$$

for all $0 \leqslant k \leqslant N$, where the bar denotes complex conjugation. The con-eigenvectors are nothing but the singular vectors (from the standard singular value decomposition of $\gamma_{\mathcal{S}A}$) rotated in appropriate position, and the con-eigenvalues are the singular values of $\gamma_{\mathcal{S}A}$. We will in the remainder refer to these as the singular vectors/values. One then picks an n such that $\sigma_n \leqslant \epsilon$, where ϵ is the desired approximation accuracy in (1.1). u_n gives rise to the polynomial

$$\breve{u_n}(z) = \sum_{k=0}^{N} u_n(k) z^k \tag{2.3}$$

Let $z_1, \ldots, z_N$ denote the roots of $\breve{u_n}$, and let $\boxed{z_k} \in \mathbb{C}^{2N+1}$ denote the vector

$$\boxed{z_k} = (1, z_k, z_k^2, \ldots, z_k^{2N}) \tag{2.4}$$

Beylkin and Monzón obtain the following result

THEOREM 2.1. *Assume that the zeroes to $\breve{u_n}$ are distinct. Then there are coefficients $c_1, \ldots, c_N$ such that, setting*

$$(\mathcal{S}A)_{ap} = \sum_{k=1}^{N} c_k \boxed{z_k}, \tag{2.5}$$

we have

$$\|\gamma_{\mathcal{S}A} - \gamma_{(\mathcal{S}A)_{ap}}\| = \sigma_n, \tag{2.6}$$

where the norm refers to the operator norm for matrixes, and [2]

$$\|\mathcal{S}A - (\mathcal{S}A)_{ap}\|_{l^2} \leqslant \sigma_n. \tag{2.7}$$

[2] We write $\|\cdot\|_{l^2}$ even though only finite sequences are involved.

The coefficients c_k such that (2.6) holds can be easily found by solving a certain Vandermonde system, we refer to [3] for the details. (2.7) holds for these coefficients as well, but if we are interested in the best l^2 approximation, the optimal coefficients c_k are different, and can be found with the least squares method.

To get back to the original function A, one observes that with $\zeta_k = -\mathrm{i}2N\log z_k$, (where log denotes the branch of the complex logarithm defined on $\mathbb{C}\backslash(-\infty,0]$), we have

$$\boxed{z_k} = \mathcal{S}\mathrm{e}^{\mathrm{i}\zeta_k x}. \tag{2.8}$$

It is thus natural to expect that A is close to $\sum_{k=1}^{N} c_k \mathrm{e}^{\mathrm{i}\zeta_k x}$ in some sense. In particular, note that

$$\|\mathcal{S}(A - \sum_{k=1}^{N} c_k \mathrm{e}^{\mathrm{i}\zeta_k \frac{k}{N}})\|_{l^2} = \|\mathcal{S}A - (\mathcal{S}A)_{ap}\|_{l^2} \leqslant \sigma_n$$

where the left hand side is close to $\sqrt{2N}\|A(x) - \sum_{k=1}^{N} c_k \mathrm{e}^{\mathrm{i}\zeta_k x}\|_{L^2([0,1])}$ if both functions are smooth enough in comparison with the sampling interval $1/2N$, (cf. (2.2)). Thus one might hope to be able to prove that

$$\|A(x) - \sum_{k=1}^{N} c_k \mathrm{e}^{\mathrm{i}\zeta_k x}\|_{L^2([0,1])} \lesssim \frac{\sigma_n}{\sqrt{2N}} \tag{2.9}$$

(although this is not stated explicitly in [3]). The obstacle here is that both ζ_k and σ_n depend on N, and therefore any attempt to take the limit as $N \to \infty$ runs into difficulties. We will come back to this issue in (2.11) and (2.12), but first we shall discuss some other aspects of the above theorem and the approximation algorithm.

Taken by itself, Theorem 2.1 is of limited value. This is due to the fact that, in order for $\mathcal{S}A$ to properly reflect A, one needs to oversample, so that N becomes a rather large number, and hence we fail to obtain a *sparse* approximation of A. Secondly, if n is chosen to be large, (so that the accuracy σ_n of the approximation improves), one can show that the algorithm is close to what is known as Prony's method, which is from the 19th century and is known to be unstable. One complication with Prony's method is that the Vandermonde-system above (cf. (2.5)) can become very ill-conditioned. We refer to [3], Section 2.3, for further details. The value of Theorem 2.1 lies in combining it with the following numerical observations made by Beylkin and Monzón:

1. By choosing such that $\sigma_n \approx \epsilon$, the numerical instabilities associated with Prony's method disappear.
2. For the functions considered in [3], the σ_k's decay very fast, and hence one may obtain $\sigma_n \approx \epsilon$ for quite small values of n.
3. Roughly $N - n$ of the terms $c_k \boxed{z_k}$ in (2.5) become so small that they can be omitted without significantly changing the sum.

Hence, upon renumbering the z_k's so that $|c_k|$ becomes decreasing with k, we obtain by Theorem 2.1 and (2.8) that

$$\|\mathcal{S}(A - \sum_{k=1}^{n} c_k \mathrm{e}^{\mathrm{i}\zeta_k x})\|_{l^2} \approx \sigma_n \tag{2.10}$$

and perhaps a corresponding approximation for A;

$$\|A(x) - \sum_{k=1}^{n} c_k \mathrm{e}^{\mathrm{i}\zeta_k x}\|_{L^2([0,1])} \approx \frac{\sigma_n}{\sqrt{2N}} \ ? \tag{2.11}$$

However, we shall show that

$$\lim_{N\to\infty} \frac{\sigma_n(N)}{\sqrt{2N}} = \infty \tag{2.12}$$

so it turns out that the value of $\sigma_n(N)$, or even $\sigma_n(N)/\sqrt{2N}$, actually gives a very bad estimate of the actual error in (2.9), (2.10) and (2.11). This fact also shows that the choice of n such that $\sigma_n \approx \epsilon$ in point 1 above is not suitable. One of the main issues solved in this paper is how to appropriately relate the σ_n's and the error ϵ in various norms.

As an example of the magnitudes involved in a specific case, we take A to be the Bessel function $J_0(100\pi x)$ on $[0,1]$ with 50 oscillations that is studied in the introduction of [3]. They obtain

$$\|A(x) - \sum_{k=1}^{n} c_k e^{i\zeta_k x}\|_{L^\infty([0,1])} \approx \epsilon$$

with $\epsilon \approx 10^{-11}$, $\sigma_{28} \approx 10^{-10}$, $n = 28$ and $N = 214$. However, by (2.12) the coincidence of the magnitudes of σ_{28} and ϵ is clearly a coincidence.

We will in this paper not discuss observation 2, but investigate number 3 quite extensively. At first sight, number 3 is a mystery, but it becomes less surprising when compared with the celebrated result by V. M. Adamjan, D. Z. Arov and M. G. Krein [1] concerning infinite dimensional Hankel matrixes, (Theorem 2.2 below). Given any $a \in l^2(\mathbb{Z}^+)$ define

$$\Gamma_a = \begin{pmatrix} a(0) & a(1) & \cdots \\ a(1) & a(2) & \cdots \\ \vdots & \vdots & \ddots \end{pmatrix} \tag{2.13}$$

For the moment we assume that a is such that Γ_a defines a compact operator on l^2. Let σ_n, $n = 0, 1, \dots$ be the singular values and u_n the corresponding singular vectors, that is, $\{u_n\}$ is an orthonormal basis of l^2 such that

$$\Gamma_a^* \Gamma_a u_k = \sigma_k^2 u_k$$

and $\sigma_k \geqslant \sigma_{k+1}$ for all $k \in \mathbb{Z}^+$. The function $\breve{u}_n = \sum_{k=0}^{\infty} u_n(k) z^k$ now has to be interpreted as a function in the Hardy space $H^2(\mathbb{D})$ of the unit disc $\mathbb{D}$, (cf (2.3)), and may therefore have infinitely many zeroes in $\mathbb{D}$. If $z \in \mathbb{D}$ is one such zero we define $\boxed{z} \in l^2$ in analogy with (2.4). ([7] and [10] are both good expositions on $H^2(\mathbb{D})$. For proofs of Theorem 2.2, see [13] a more accessible version than [1], or [4] for a different proof).

THEOREM 2.2. *(AAK) Let $a \in l^2(\mathbb{Z}^+)$ be given and let Γ_a, u_n etc. be as above. Let $\sigma_n > 0$ be a fixed distinct singular value. Then $\breve{u}_n$ has exactly n zeroes $z_1, \dots, z_n$ in $\mathbb{D}$, repeated according to multiplicity. If the zeroes are distinct then there are coefficients $c_1, \dots, c_n \in \mathbb{C}$ such that*

$$\|\Gamma_a - \Gamma_{\sum_{j=1}^{n} c_j \boxed{z_j}}\| = \sigma_n \tag{2.14}$$

This should be compared with equation (2.6) in Theorem 2.1. The counterpart (2.7) clearly also follows from Theorem 2.2, as the following calculation shows;

$$\|a - \sum c_k \boxed{z_k}\,\|_{l^2(\mathbb{Z}^+)} = \|(\Gamma_a - \Gamma_{\sum c_k \boxed{z_k}})(e_0)\|_{l^2(\mathbb{Z}^+)} \leqslant \|\Gamma_a - \Gamma_{\sum c_k \boxed{z_k}}\| \|e_0\|_{l^2(\mathbb{Z}^+)} = \sigma_n,$$

where e_0 denotes the first vector in the standard basis for l^2.

The astounding (and hard) part of Theorem 2.2 is the fact that $\breve{u}_n$ has exactly n zeroes. Once this is established, the estimate (2.14) follows rather easily, and Theorem 2.1 can be proved along the same lines, see [2]. The novelty of Beylkin and Monzón's work lies in the idea of using tightly sampled values of some function A to form Hankel matrices, rather than its Fourier coefficients as is customary in complex analysis. We will in this paper explain the numerical observations 1 and 3.

Now, let the dependence of σ_n, c_k and ζ_k on N be explicit, i.e. $\sigma_n = \sigma_n(N)$, $c_k = c_k(N)$ and $\zeta_k = \zeta_k(N)$. Clearly, one would like to know what happens when $N \to \infty$, and in particular wether some estimate similar to (2.11) is asymptotically stable as $N \to \infty$. For our slightly modified version of the algorithm, we will show that $\zeta_k(N)$ and $c_k(N)$, if ordered properly, has limits ζ_k^* and c_k^* as $N \to \infty$. Unfortunately, as already has been mentioned, we also show that $\sigma_n(N)/\sqrt{2N} \to \infty$ as $N \to \infty$ so the estimate (2.11) is unreliable. This issue is not merely a theoretical flaw, but important in applications. Say, for example, that we have obtained sparse approximations of two functions $A_1 \approx \sum_{k=1}^n c_k^1 e^{i\zeta_k^1 x}$ and $A_2 \approx \sum_{k=1}^n c_k^2 e^{i\zeta_k^2 x}$ and say that we wish to calculate their scalar product. In [3] it is pointed out that it is much faster to do this algebraically using the approximations, i.e. the formula

$$\langle A_1, A_2 \rangle \approx \sum_{k_1,k_2} c_{k_1}^1 \overline{c_{k_2}^2} \frac{e^{i\zeta_{k_1}^1 + \overline{i\zeta_{k_1}^1}} - 1}{i\zeta_{k_1}^1 + \overline{i\zeta_{k_1}^1}}.$$

However, if we do not have good estimates on

$$\|A_j - \sum_{k=1}^n c_k^j e^{i\zeta_k^j x}\|_{L^2(\mathbb{R}^+)}, \quad j = 1, 2$$

then clearly the numbers produced as above are unreliable and might lead to errors. Our modified algorithm has certain advantages which we now list. (All objects below, like e.g. u_k, refer to the corresponding object for the modified algorithm.)

1. We can prove that each $\widetilde{u_n}$ generically has exactly n zeroes in the unit disc, and that these are the significant ones for the approximation. This is not merely of theoretical interest; it is computationally time consuming to find all the N roots of $\widetilde{u_n}$ and then determine which are the significant ones needed for the approximation in question by solving a large Vandermonde system.
2. We show that if we sample A according to $\mathcal{S}_{1,N} f = \left(\frac{1}{2N} A(\frac{k}{2N})\right)_{k=0}^{2N}$ and fix n, then both $\zeta_k(N)$, (ordered appropriately), and $\sigma_n(N)$ converge when $N \to \infty$. Moreover, we provide norms in which (2.11) does hold, i.e. if we denote the respective limits above by ζ_k^* and σ_n^*, then

$$\|A(x) - \sum_{k=1}^n c_k^* e^{i\zeta_k^* x}\| \leqslant \sigma_n^*. \tag{2.15}$$

   For example, with some modifications, one may take the Sobolev norm $W^{-1,2}$ or the BMO norm of the Fourier transform, (see Section 5.4 for more details.)

A drawback of the result in 2 is that these norms are quite weak. We offer a more intricate algorithm which is proven to be asymptotically stable and yield an estimate similar to (2.15) with respect to the L^2-norm in Section 7.

3. Algorithm 1. This section introduces our modification of Beylkin and Monzón's algorithm. We consider a function $A \in L^1(\mathbb{R}^+) \cap C_0(\mathbb{R}^+)$ which we aim to approximate by a sum of few exponentials, ($C_0(\mathbb{R}^+)$ denotes the set of continuous functions with compact support in $\mathbb{R}^+ = [0, \infty)$). We redefine the operator $\mathcal{S}_{p,N}$ (cf. eq. (2.1)) to an operator from the set of piesewise continuous functions on $\mathbb{R}^+$ by setting

$$\mathcal{S}_{p,N} A = \left(\frac{1}{(N)^{1/p}} A\left(\frac{k}{N}\right) \right)_{k=0}^{\infty}. \tag{3.1}$$

We then sample A using the operator $\mathcal{S} = \mathcal{S}_{1,N}$ where $N \in \mathbb{N}$ is some "sufficiently high" number. We will mainly be concerned with showing that the algorithm is stable as $N \to \infty$, and therefore we omit a further discussion of what "sufficiently high" means. The main changes compared to Section 2 is thus that the function A is assumed to be zero at the right endpoint and the factor $1/N$ in the sampling operator, (recall that we used $\mathcal{S}_{\infty,2N}$ in Section 2). Note that σ_n changes accordingly with $1/N$.

The next step is to define the infinite Hankel matrix $\Gamma_{\mathcal{S}A} = (\mathcal{S}A(i+j))_{i,j}$. As A has compact support, this matrix only contain zeroes for all rows and columns with index greater than some number I. We can therefore treat it as a finite matrix and find singular values etc. using computers. Let I be the smallest number such that $\mathcal{S}A(i)=0$ for all $i>I$ and set $\gamma_{\mathcal{S}A}=(\mathcal{S}A(i+j))_{0\leqslant i,j\leqslant I}$. For example, if the support of A is precisely $[0,1]$, then $\gamma_{\mathcal{S}A}$ is the $N\times N$-matrix defined by

$$\gamma_{\mathcal{S}A} = \begin{pmatrix} \mathcal{S}A(0) & \mathcal{S}A(1) & \cdots & \mathcal{S}A(N-2) & \mathcal{S}A(N-1) \\ \mathcal{S}A(1) & \mathcal{S}A(2) & \cdots & \mathcal{S}A(N-1) & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathcal{S}A(N-2) & \mathcal{S}A(N-1) & \cdots & 0 & 0 \\ \mathcal{S}A(N-1) & 0 & \cdots & 0 & 0 \end{pmatrix}.$$

Let $\sigma_0 \geqslant \sigma_1 \geqslant \ldots \geqslant \sigma_I$ be the singular values of $\gamma_{\mathcal{S}A}$ and let $u_0,\ldots,u_I \in \mathbb{C}^{I+1}$ be the corresponding singular vectors. Theorem 2.2 then immediately implies

COROLLARY 3.1. *Assume that all non-zero σ_k's are different and let $n\in\mathbb{N}$ be fixed such that $\sigma_n>0$. Then $\breve{u_n}$ has exactly n zeroes $z_k\in\mathbb{D}$, $k=1,\ldots,n$ (counted with multiplicity). Assume that the zeroes are distinct and do not lie on the negative axis. Set $\zeta_k=-\mathrm{i}N\log z_k$. Then there are coefficients $c_1,\ldots,c_n\in\mathbb{C}$ such that, with*

$$A_{ap}(x)=\sum_{k=1}^{n} c_k \mathrm{e}^{\mathrm{i}\zeta_k x},$$

we have

$$\|\Gamma_{\mathcal{S}(A-A_{ap})}\| = \|\Gamma_{\mathcal{S}A} - \Gamma_{\frac{c_k}{N}\boxed{z_k}}\| = \sigma_n \tag{3.2}$$

Remark 1: Recall that $\boxed{z_k}$ was first defined as a finite sequence in (2.4) and later used as an element of l^2 by extending it in the obvious fashion. To simplify notation, we will continue to let the actual meaning of $\boxed{z_k}$ be determined by the environment, as there is no risk for confusion. Obviously the z_k's and c_k's also depend heavily on n, but as n for the majority of the paper will be considered to be a fixed number chosen at the beginning of the algorithm, we will omit this dependence in the notation.

Remark 2: To compute the c_k's, one proceeds as follows. Set $v=\overline{\breve{u_n}}/\breve{u_n}$, let $\hat{v}$ denote its Fourier coefficients and set $d(k)=\hat{v}(-k)$ for $k=0,\ldots,N$. It will follow from the material in Section 5 that the system

$$\mathcal{S}A-d=\sum_{k=1}^{n} c_k\boxed{z_k}$$

has a unique solution, and that these are the coefficients to use for Corollary 3.1.

We will in the next two sections study the asymptotic behavior of $\sigma_n=\sigma_n(N)$, $c_k=c_k(N)$, $\zeta_k=\zeta_k(N)$ and $A_{ap}(N)$ as $N\to\infty$. We shall prove that they all converge (if ordered appropriately) to some limits which we shall call σ_n^*, c_k^*, ζ_k^* and A_{ap}^*.

Let us assume these facts for the moment and note that as a consequence of Corollary 3.1 we also get the brutal estimate

$$\frac{1}{N}\sqrt{\sum_{j=0}^{\infty}\left|(A-A_{ap}(N))\left(\frac{j}{N}\right)\right|^2} = \|\mathcal{S}(A-A_{ap}(N))\|_{l^2} = \|\Gamma_{\mathcal{S}(A-A_{ap}(N))}(e_0)\|_{l^2} \leqslant \sigma_n(N),$$

which should be compared with (2.7), keeping in mind that $\sigma_n(N)$ in Section 2 is essentially the same object as $N\sigma_n(N)$ in this section, due to the use of different sampling operators. The above estimate is clearly of limited value as

$$\|A - A^*_{ap}\|_{L^2} = \lim_{N\to\infty} \frac{1}{\sqrt{N}} \sqrt{\sum_{j=0}^{\infty} \left|(A - A_{ap}(N))\left(\frac{j}{N}\right)\right|^2},$$

which yields

$$\|A - A_{ap}(N)\|_{L^2(\mathbb{R}^+)} \lesssim \sqrt{N}\sigma_n(N).$$

The outline for the coming sections is as follows. In Section 4 we show that $\sigma_n(N)$ converges, in section 5.1, 5.2 and 5.3 we review the necessary Hankel operator theory and in Section 5.4 we provide estimates of $\|A - A^*_{ap}\|$ in norms that are more tractable than the one obtained by taking the limit of (3.2) as $N \to \infty$. Finally, in Section 6 we prove that the other sequences $\zeta_k(N)$ and $c_k(N)$ converge.

4. Asymptotic behavior of $\Gamma_{\mathcal{S}_{1,N}A}$. Given $A \in L^1(\mathbb{R}^+) \cap C^1(\mathbb{R}^+)$ and $F \in L^2(\mathbb{R}^+) \cap C(\mathbb{R}^+)$, it is easy to see that, for every $k \in \mathbb{N}$, we have

$$\lim_{N\to\infty} \left(\Gamma_{\mathcal{S}_{1,N}A}\mathcal{S}_{\infty,N}F\right)(kN) = \int_0^{\infty} A(k+y)F(y)dy.$$

We therefore define the "real line Hankel operators" $\Gamma_A : L^2(\mathbb{R}^+) \to L^2(\mathbb{R}^+)$ by

$$\Gamma_A F(x) = \int_0^{\infty} A(x+y)F(y)dy.$$

Remark: This notation collides with the notation for a "classical" Hankel operator (2.13). We will let the type of the symbol A determine wether Γ_A is an operator on $l^2(\mathbb{Z}^+)$ or $L^2(\mathbb{R}^+)$. Elements of $l^2(\mathbb{Z}^+)$ will generally be denoted by lower case letters and elements of $L^2(\mathbb{R}^+)$ by upper case letters.

As Γ_A is in some sense a limit of $\Gamma_{\mathcal{S}_{1,N}A}$, it is natural to expect that Γ_A inherits many of the properties of Hankel operators. This is indeed true, and will be treated in Section 5. In this section we shall make precise the statement that Γ_A is a limit of $\Gamma_{\mathcal{S}_{1,N}A}$

Let $\chi(S,x)$ denote the characteristic function of a set S. For each $N \in \mathbb{N}$, set

$$b_k^N(x) = \sqrt{N}\chi\left(\left[\frac{k}{N}, \frac{k+1}{N}\right), x\right)$$

and let $P_N : L^2(\mathbb{R}^+) \to L^2(\mathbb{R}^+)$ be the orthogonal projection on the subspace spanned by $\{b_k^N\}_{k=0}^{\infty}$. Note that $\{b_k^N\}_{k=0}^{\infty}$ is an orthonormal set in $L^2(\mathbb{R}^+)$. Recall the definition of $\mathcal{S}_{2,N}$ in (3.1) and note that $\mathsf{Ran}\ P_N = \mathsf{Ran}\ \mathcal{S}_{2,N}$. We also define $\mathcal{S}_{2,N}^{inv} : l^2(\mathbb{Z}^+) \to L^2(\mathbb{R}^+)$ by

$$\mathcal{S}_{2,N}^{inv}(a) = \sum a_k b_k^N, \tag{4.1}$$

where the funny notation is motivated by the formula

$$\mathcal{S}_{2,N}^{inv}\mathcal{S}_{2,N}P_N = P_N,$$

which shows that $\mathcal{S}_{2,N}^{inv}$ acts as an inverse to $\mathcal{S}_{2,N}$ restricted to the subspace $\mathsf{Ran}\ \mathcal{S}_{2,N}$. We shall show that the operator $\mathcal{S}_{2,N}^{inv}\Gamma_{\mathcal{S}_{1,N}A}\mathcal{S}_{2,N}$ converges to Γ_A as $N \to \infty$. In order to simplify the notation we set

$$\Gamma_A^N = \mathcal{S}_{2,N}^{inv}\Gamma_{\mathcal{S}_{1,N}A}\mathcal{S}_{2,N},$$

but one should keep in mind that Γ_A^N is not a real line Hankel operator. Let $\rho_N : C^1(\mathbb{R}^+) \to C(\mathbb{R}^+)$ be defined by

$$\rho_N(A)(x) = \sup\{|A'(y)| : |x-y| \leqslant N^{-1}\}$$

PROPOSITION 4.1. *Let $A \in L^1(\mathbb{R}^+) \cap C^1(\mathbb{R}^+)$ be given. Then*

$$\|\Gamma_A - \Gamma_A^N\| \leqslant \frac{4}{N}\sqrt{\int_0^\infty x\big|\rho_{2N}(A)(x)\big|^2 dx}$$

Proof. We shall first give an estimate of $\Gamma_A - \Gamma_A^N$ restricted to $\mathsf{Ran}\ P_N$. For each fixed $x \in \mathbb{R}^+$ we have

$$\Gamma_A b_k^N(x) = \sqrt{N}\int_{x+k/N}^{x+(k+1)/N} A(y)dy = N^{-1/2}A(x+y_x), \tag{4.2}$$

for some $y_x \in \mathbb{R}^+$ with $k/N \leqslant y_x \leqslant (k+1)/N$ by the mean value theorem. On the other hand, note that $\mathcal{S}_{2,N}b_k^N = e_k$ so

$$\Gamma_A^N b_k^N(x) = (\mathcal{S}_{2,N}^{inv}\Gamma_{\mathcal{S}_{1,N}A}e_k)(x) = \sum N^{-1}A\left(\frac{l+k}{N}\right)b_l^N(x) = N^{-1/2}A\left(\frac{l_x+k}{N}\right)$$

where $l_x \in \mathbb{Z}^+$ is such that $l_x/N \leqslant x < (l_x+1)/N$. Another application of the mean value theorem yields

$$\left|\Gamma_A b_k^N(x) - \Gamma_A^N b_k^N(x)\right| = N^{-1/2}\left|A(x+y_x) - A\left(\frac{l_x+k}{N}\right)\right| \leqslant 2N^{-3/2}\rho_N(A)(x+k/N). \tag{4.3}$$

Now, let $a \in l^2$ be arbitrary but satisfy $\|a\| = 1$. Then

$$\begin{aligned}\left|\left(\left(\Gamma_A - \Gamma_A^N\right)\left(\sum a_k b_k^N\right)\right)(x)\right| &= \left|\sum a_k\left(\left(\Gamma_A - \Gamma_A^N\right)b_k^N\right)(x)\right| \leqslant \\ &\leqslant 2N^{-3/2}\left|\sum a_k\rho_N(A)(x+k/N)\right| = 2N^{-1}\sqrt{N^{-1}\sum\left(\rho_N(A)(x+k/N)\right)^2} \leqslant \\ &\leqslant 2N^{-1}\sqrt{\sum\int_{\frac{k}{N}}^{\frac{k+1}{N}}\left(\rho_{2N}(A)(x+y)\right)^2 dy} = 2N^{-1}\sqrt{\int_0^\infty\left(\rho_{2N}(A)(x+y)\right)^2 dy}\end{aligned}$$

Finally, we obtain

$$\begin{aligned}\|(\Gamma_A - \Gamma_A^N)\left(\sum a_k b_k^N\right)\|_{L^2} &\leqslant 2N^{-1}\sqrt{\int_0^\infty\int_0^\infty\left(\rho_{2N}(A)(x+y)\right)^2 dydx} = \\ &= 2N^{-1}\sqrt{\int_0^\infty\int_{-u}^{u}\left(\rho_{2N}(A)(u)\right)^2\frac{1}{2}dvdu} = 2N^{-1}\sqrt{\int_0^\infty u\left(\rho_{2N}(A)(u)\right)^2 du}\end{aligned}$$

which, upon noting that $\|\sum a_k b_k\|_{L_2} = 1$, yields

$$\|(\Gamma_A - \Gamma_A^N)P_N\| \leqslant 2N^{-1}\sqrt{\int_0^\infty x\left(\rho_{2N}(A)(x)\right)^2 dx}. \tag{4.4}$$

We turn to the estimate for $(\Gamma_A - \Gamma_A^N)(I-P_N) = \Gamma_A(I-P_N)$. Define subsets $S_{k,i,j}^N \subset [\frac{k}{N}, \frac{k+1}{N}]$ for $j \geqslant 1$ and $0 \leqslant i \leqslant 2^j - 1$ by

$$S_{k,i,j}^N = \left[\frac{k}{N} + \frac{i}{2^jN}, \frac{k}{N} + \frac{i+1}{2^jN}\right],$$

and let $d^N_{k,i,j}$ be functions defined by

$$d^N_{k,i,j}(x) = \sqrt{2^{j-1}N}\left(\chi(S^N_{k,2i,j}, x) - \chi(S^N_{k,2i+1,j}, x)\right)$$

for $j \geqslant 1$ and $0 \leqslant i < 2^{j-1}$. It is easy to see that, (for N fixed),

$$\mathsf{Span}\ (\{b^N_k\} \cup \{d^N_{k,i,j}\}) = \mathsf{Span}\ (\{\chi_{S^N_{k,i,j}}\}),$$

and it follows from basic integration theory that the right hand side is dense in $L^2(\mathbb{R}^+)$. Moreover, $\{b^N_k\} \cup \{d^N_{k,i,j}\}$ is clearly an orthonormal set, and hence it is a basis for $L^2(\mathbb{R}^+)$. Thus $\mathsf{Ran}\ (I - P_N) = \mathsf{Span}\ \{d^N_{k,i,j}\}$. In a similar fashion as in (4.2) and (4.3) we obtain

$$|\Gamma_A d^N_{k,i,j}(x)| = \frac{\sqrt{2^{j-1}N}}{2^j N}\left|A\left(x + \frac{k}{N} + \frac{2i}{2^j N} + \delta_1\right) - A\left(x + \frac{k}{N} + \frac{2i+1}{2^j N} + \delta_2\right)\right| \leqslant$$
$$\leqslant \frac{1}{2\sqrt{2^{j-1}N}}\left|\frac{1}{2^{j-1}N}\rho_N(A)\left(x + \frac{k}{N}\right)\right| = \frac{\sqrt{2}}{(2^j N)^{3/2}}\left|\rho_N(A)\left(x + \frac{k}{N}\right)\right|,$$

where $0 \leqslant \delta_1, \delta_2 \leqslant (2^j N)^{-1}$. Thus

$$\|\Gamma_A d^N_{k,i,j}(x)\| \leqslant \frac{\sqrt{2}}{(2^j N)^{3/2}}\|\rho_N(A)\left(\cdot + \frac{k}{N}\right)\|_{L^2(\mathbb{R}^+)}.$$

Now let $a_{k,i,j} \in \mathbb{C}$ be any numbers (indexed by the same index set as $\{d^N_{k,i,j}\}$) such that $\sum |a_{k,i,j}|^2 = 1$. By repeated use of Cauchy-Schwartz inequality we get

$$\|\Gamma_A\left(\sum a_{k,i,j} d^N_{k,i,j}\right)\| \leqslant \sum_{k=0}^{\infty}\sum_{j=1}^{\infty}\sum_{i=0}^{2^{j-1}-1} |a_{k,i,j}|\frac{\sqrt{2}}{(2^j N)^{3/2}}\|\rho_N(A)\left(\cdot + \frac{k}{N}\right)\| \leqslant$$
$$\leqslant \sum_{k=0}^{\infty}\sum_{j=1}^{\infty}\sqrt{\sum_i |a_{k,i,j}|^2 2^{(j-1)/2}}\frac{\sqrt{2}}{(2^j N)^{3/2}}\|\rho_N(A)\left(\cdot + \frac{k}{N}\right)\| \leqslant$$
$$\leqslant \sum_{k=0}^{\infty}\sqrt{\sum_{j,i}|a_{k,i,j}|^2}\sqrt{\sum_{j\geqslant 1} 2^{-2j}}\frac{1}{N^{3/2}}\|\rho_N(A)\left(\cdot + \frac{k}{N}\right)\| \leqslant$$
$$\leqslant \sqrt{\frac{1}{1 - 1/4}}\frac{1}{N^{3/2}}\sqrt{\sum_{k,j,i}|a_{k,i,j}|^2}\sqrt{\sum_k \int_0^{\infty}\left|\rho_N(A)\left(\cdot + \frac{k}{N}\right)\right|^2} \leqslant$$
$$\leqslant \frac{2}{\sqrt{3}N}\sqrt{\int_0^{\infty}\int_0^{\infty}|\rho_{2N}(A)(y + x)|^2\,dydx} \leqslant \frac{2}{\sqrt{3}N}\sqrt{\int_0^{\infty} x|\rho_{2N}(A)(x)|^2 dx}.$$

It follows that

$$\|(\Gamma_A - \Gamma^N_A)(I - P_N)\| \leqslant \frac{2}{\sqrt{3}N}\sqrt{\int_0^{\infty} x\left|\rho_{2N}(A)(x)\right|^2 dx}$$

which combined with (4.4) yields the desired result. □

We are now in a position to prove the first claim at the end of Section 3, namely that $\sigma_n(N)$ has some limit value σ^*_n as $N \to \infty$. By standard theorems concerning Hankel operators, (see Section 5), it follows that Γ_A and Γ_a are compact operators whenever $A \in L^1$ or $a \in l^1$. We need the following fact about compact operators. We have not found a reference, so we outline a proof. For a any operator T on some Hilbert space $\mathcal{X}$, recall that its singular values are defined by

$$\sigma_n = \inf\left\{\|T|_{\mathcal{N}}\| : \mathcal{N} \subset \mathcal{X} \text{ is a linear subspace with } \mathsf{codim}\ \mathcal{N} = n\right\}.$$

If T is compact, it is easy to see from the polar decomposition that σ_n is the n:th eigenvalue of $\sqrt{T^*T}$, counted with multiplicity in decreasing order.

LEMMA 4.2. *Let T and T_N, $N \in \mathbb{N}$, be compact operators and assume that* $\lim_N \|T - T_N\|$. *Let $\sigma_0(N) \geqslant \sigma_1(N) \geqslant \dots$ be the singular values of T_N and $\sigma_0 \geqslant \sigma_1 \geqslant \dots$ be the singular values of T. Then*

$$\lim_N \sigma_n(N) = \sigma_n$$

for all $n \in \mathbb{Z}^+$. Moreover, let $\mathcal{E}_N$ and $\mathcal{E}$ be the corresponding projection-valued spectral measures associated with $\sqrt{T_N^* T_N}$ and $\sqrt{T^*T}$. Let $\Sigma \subset \mathbb{R}^+$ be any open interval such that

$$cl(\Sigma) \cap \{\sigma_k : k \in \mathbb{Z}_+\} = \Sigma \cap \{\sigma_k : k \in \mathbb{Z}_+\} = \{\sigma_n\}$$

for some $n \in \mathbb{Z}_+$, where $cl(\Sigma)$ denotes the closure of Σ. Then

$$\lim_N \|\mathcal{E}_N(\Sigma) - \mathcal{E}(\Sigma)\| = 0.$$

Proof. Note that the first part of the lemma follows by the second. Put a small circle $\Delta \subset \mathbb{C}$ that touches both endpoints of Σ. The resolvent of $\sqrt{T^*T}$ will then be defined on Δ and

$$\mathcal{E}(\Sigma) = \int_\Delta \left(\lambda - \sqrt{T^*T}\right)^{-1} \frac{d\lambda}{2\pi \mathrm{i}}$$

by the Riesz functional calculus, see e.g. [6]. The lemma follows as $\left(\lambda - \sqrt{T_N^* T_N}\right)$ is invertible on Δ for large N and

$$\lim_{N\to\infty} \sup_{\lambda\in\Delta}\{\| \left(\lambda - \sqrt{T_N^* T_N}\right)^{-1} - \left(\lambda - \sqrt{T^*T}\right)^{-1} \|\},$$

by standard resolvent estimates. □

Given a real line Hankel operator Γ_A and corresponding Hankel operators $\Gamma_{\mathcal{S}_{1,N}A}$, we will always denote the singular values of Γ_A by σ_n and those of $\Gamma_{\mathcal{S}_{1,N}A}$ by $\sigma_n(N)$. Due to the following corollary, we will in the remainder omit the star in the notation σ_n^* for the limit of the sequence $\sigma_n(N)$, and the same goes for ζ_k^*, c_k^* and A_{ap}^*.

COROLLARY 4.3. *Let $A \in L^1(\mathbb{R}^+) \cap C^1(\mathbb{R}^+)$ be such that*

$$\sqrt{\int_0^\infty x \left|\rho_N(A)(x)\right|^2} < \infty$$

and $\mathcal{S}_{1,N}A \in l^1$ for all N. Then

$$\lim_N \sigma_n(N) = \sigma_n.$$

Proof. $\Gamma_A^N(I - P_N) = 0$ and $\mathsf{Ran}\ P_N$ is clearly a reducing subspace for Γ_A^N, so the singular values of Γ_A^N are completely determined by $\Gamma_A^N|_{\mathsf{Ran}\ P_N}$ as an operator from $\mathsf{Ran}\ P_N$ to $\mathsf{Ran}\ P_N$. From the definition of Γ_A^N it follows immediately that the matrix representing Γ_A^N in the orthonormal basis $\{b_k^N\}_{k=0}^\infty$ is precisely $\Gamma_{\mathcal{S}_{1,N}A}$, and hence $\{\sigma_n(N)\}$ are the non-zero singular values of Γ_A^N. With this in mind, the Corollary follows immediately from Proposition 4.1 and Lemma 4.2. □

Our next goal will be to show that (using the notation from Section 3) the sequences of complex exponents $\zeta_k(N)$ are convergent. In order to do so, we will need to go through a fair amount of Hankel operator theory, which is the content of Section 5. In particular, we will explicitly write down a version of the AAK-theorem for real line Hankel operators. The fact that such a theorem holds is probably known to specialists although we have not been able to find it in print. The result concerning the convergence of the $\zeta_k(N)$'s will be proven first in Section 6.

5. Hankel operators on $H^2(\mathbb{D})$ and $H^2(\mathbb{C}^+)$. For the readers not familiar with Hardy space theory and the AAK-theorem in particular (Theorem 2.2), sections 5.1 and 5.2 serves as a very brief introduction. Other readers may just glance through it to get familiar with the notation. Good expositions on Hardy space theory are the books by P. Duren [7] and P. Koosis [10], but these books do not include the AAK-theorem. Section 4 in V. Peller's book on Hankel operators [13] is devoted to a complete proof of the AAK-theorem. The proof in the case when the Hankel operator is compact is simpler and can also be found in [14]. The main steps will be outlined below.

5.1. A review of $H^2(\mathbb{D})$. We will use the letter m to denote the normalized arc-length measure on $\mathbb{T}$, $L^2(\mathbb{T})$ will denote the corresponding function space on $\mathbb{T}$, and $\mathcal{F}: L^2(\mathbb{T}) \to l^2(\mathbb{Z})$ will denote the unitary Fourier transform, i.e. $\mathcal{F}(\phi)(k) = \int \phi(z) z^{-k} dm$. As usual, we will also denote $\mathcal{F}(\phi)$ by $\hat{\phi}$ and similarly $\check{\phi}$ denotes the inverse transform, so that $\check{\hat{\phi}} = \phi$. The space $H^2(\mathbb{D})$ is customarily defined as a Hilbert space of analytic functions on $\mathbb{D}$ satisfying a certain mean growth restriction near the boundary $\mathbb{T}$. More precisely, an analytic function ϕ on $\mathbb{D}$ is in $H^2(\mathbb{D})$ if

$$\|\phi\|_{H^2(\mathbb{D})} = \frac{1}{2\pi} \sup_{r \in (0,1)} \int_0^{2\pi} |\phi(re^{it})|^2 dt < \infty. \tag{5.1}$$

Alternatively, one may define $H^2(\mathbb{D})$ as the subset of $L^2(\mathbb{T})$ given by $\mathcal{F}^{-1}(l^2(\mathbb{Z}^+))$, i.e. the set of functions whose Fourier coefficients with negative index are zero. Clearly, each element $\phi = \sum_{k=0}^{\infty} c_k z^k$ of the latter definition defines an analytic function in $\mathbb{D}$ by considering z as an independent variable in $\mathbb{D}$. It is a standard fact that the set of functions obtained in this way coincides with the first definition and the two norms are the same. Conversely, given an analytic function ϕ in $H^2(\mathbb{D})$ using the first definition, the corresponding function on $\mathbb{T}$ is obtained by taking "non-tangential limits" of ϕ $m - a.e.$ Therefore, when dealing with $H^2(\mathbb{D})$ we will make no distinction between the analytic function in $\mathbb{D}$ and its "boundary function" on $\mathbb{T}$.

Analogous facts hold for all $H^p(\mathbb{D})-$spaces, $1 \leqslant p \leqslant \infty$, where $H^p(\mathbb{D})$ is defined in analogy with (5.1) except for $p = \infty$; $H^\infty(\mathbb{D})$ is then simply defined as the space of all bounded analytic functions on $\mathbb{D}$.

We now recall some elementary results concerning $H^2(\mathbb{D})$. Given $z_0 \in \mathbb{D}$, set

$$k_{z_0}(\zeta) = \frac{1}{1 - \overline{z_0}\zeta}$$

where the bar denotes complex conjugation. k_{z_0} is called the reproducing kernel for z_0 due to the formula

$$\phi(z_0) = \langle \phi, k_{z_0} \rangle, \quad \forall \phi \in H^2(\mathbb{D})$$

which follows immediately from Parseval's identity and

$$\widehat{k_{z_0}} = \boxed{z_0} = (1, z_0, z_0^2, \ldots).$$

Let $z: \mathbb{D} \to \mathbb{D}$ denote the function $z(\zeta) = \zeta$. Note that the operator $M_z: H^2(\mathbb{D}) \to H^2(\mathbb{D})$ given by $M_z(\phi) = z\phi$ is a bounded operator with $\|M_z\| = 1$. It is usually referred to as the unilateral shift operator, due to its effect on the Fourier coefficients of ϕ. The proposition below is a special case of Beurling's theorem which characterizes all M_z invariant subspaces in $H^2(\mathbb{D})$. Given $\phi \in \mathcal{H}^2(\mathbb{D})$ let $inv_{M_z}(\phi)$ denote the invariant subspace generated by ϕ and M_z, i.e.

$$inv_{M_z}(\phi) = cl\{z^k \phi : k \in \mathbb{Z}^+\},$$

where cl denotes the closure. Given a sequence $(z_1, \ldots, z_K)$ in $\mathbb{D}$ with distinct numbers, let $\mathcal{M}\left((z_k)_{k=1}^K\right) \subset H^2(D)$ be defined by

$$\mathcal{M}\left((z_k)_{k=1}^K\right) = \{\phi \in H^2(\mathbb{D}) : \phi(z_j) = 0, \ j = 1, \ldots, K\}.$$

If repetitions are present in $(z_1, \dots, z_K)$ and z_j appear n_j times, say, then we simply modify the above definition to demand that ϕ has a zero of multiplicity n_j at z_j. Note that in the case when all z_k's are distinct, we have

$$\mathcal{M}\left((z_k)_{k=1}^K\right) = \{k_{z_j},\ j = 1, \dots, K\}^\perp = H^2(\mathbb{D}) \ominus \{k_{z_j},\ j = 1, \dots, K\}.$$

PROPOSITION 5.1. *Let $\phi \in H^2(\mathbb{D})$ be a polynomial and let $z_1, \dots, z_K \in \mathbb{D}$ be its zeroes in $\mathbb{D}$, repeated according to multiplicity. Then*

$$inv_{M_z}(\phi) = \mathcal{M}\left((z_k)_{k=1}^K\right).$$

Moreover any M_z-invariant subspace $\mathcal{M} \subset \mathcal{H}$ with finite codimension K is of the above form.

5.2. Hankel operators and the AAK-theorem. We will in this section give a very brief summary of Hankel operators and outline some of the ideas in the proof of the AAK-theorem (Theorem 2.2). Set $H_-^2(\mathbb{D}) = L^2(\mathbb{T}) \ominus H^2(\mathbb{D})$ and let $\mathbb{Z}^- = \mathbb{Z} \setminus \mathbb{Z}^+ = \{-k\}_{k=1}^\infty$. We define P_+ and P_- to be the orthogonal projections on $l^2(\mathbb{Z})$ with ranges $l^2(\mathbb{Z}^+)$ and $l^2(\mathbb{Z}^-)$. By abuse of notation we use the same notation for the corresponding operators on $L^2(\mathbb{T})$ onto $H^2(\mathbb{D})$ and $H_-^2(\mathbb{D})$ respectively. Given $b \in L^2(\mathbb{T})$, we define the Hankel operator $H_b : H^2(\mathbb{D}) \to H_-^2(\mathbb{D})$ via

$$H_b(\phi) = P_-(b\phi), \tag{5.2}$$

whenever b is such that this yields a bounded operator. It is a simple exercise to check that if we set $a(k) = \hat{b}(-1-k)$ and take $(z^k)_{k=0}^\infty$ as a basis in $H^2(\mathbb{D})$ and $(z^{-k})_{k=1}^\infty$ as a basis in $H_-^2(\mathbb{D})$, then the matrix corresponding to H_b is precisely Γ_a as defined in (2.13). Thus H_b and Γ_a are the same from an operator theoretic viewpoint. We will always use the notation H_b for Hankel operators defined via (5.2) and Γ_a for Hankel operators defined via (2.13).

Now, given a Hankel operator Γ_a, any function b such that H_b is equivalent with Γ_a as above will be called a symbol of Γ_a. Note that $P_+ b$ does not affect the operator, so there are many symbols for a given Hankel matrix Γ_a. In fact, a is completely determined by $P_- b$. A natural question is clearly for which b's the operator H_b is bounded and how to compute the norm. If $b \in L^\infty$ we clearly have $\|\Gamma_a\| \leqslant \|b\|_{L^\infty}$. These issues were completely resolved by Z. Nehari in 1957, [11].

THEOREM 5.2. (NEHARI) *A Hankel matrix Γ_a, ($a \in l^2(\mathbb{Z}^+)$), is bounded if and only if there exists a symbol b in L^∞. In this case*

$$\|\Gamma_a\| = \inf\{\|b\|_{L^\infty}\}$$

where the infimum is taken over all possible symbols.

Combined with the celebrated characterization of BMO by C. Fefferman, this implies that H_b is bounded if and only if $P_- b \in BMO$. Moreover H_b is compact if and only if $P_- b \in VMO$, which follows from the next theorem which describes all compact Hankel operators, [8].

THEOREM 5.3. (HARTMAN) *A Hankel matrix Γ_a, ($a \in l^2(\mathbb{Z}^+)$), is compact if and only if there exists a continuous symbol.*

We will now collect a few observations to be used in the "proof" of the AAK-theorem. Note that, given any $\phi \in H^2(\mathbb{D})$ and we have

$$\widehat{(H_b\phi)}(-k-1) = \int z^{k+1} b\phi \, dm = \langle z^k \phi, P_+(\overline{zb})\rangle.$$

From this we see that $\mathsf{Ker}\ H_b$ is M_z-invariant and that $H_b\phi = 0$ if and only if

$$P_+(\overline{zb}) \perp inv_{M_z}(\phi). \tag{5.3}$$

For the applications we have in mind, the Hankel operators will always be compact and their non-zero singular values will generically be distinct, so we will in the remainder assume that this is the case. This assumption is only made in order to simplify the exposition, and moreover it is not a major restriction because one can prove that by randomly generating finite Hankel matrices, the probability to get one with multiple singular values is zero. We omit a proof but note that this is proven for Hankel operators in 2 dimensions in [2].

From now on let H_b be such an operator and denote its non-zero singular values by $\sigma_0 > \sigma_1 > \ldots$. In order to use the same notation as in the introduction, we define the con-eigenvectors u_n as elements of $l^2(\mathbb{Z}^+)$ that satisfy $\sigma_n^2 \breve{u_n} = H_b^* H_b \breve{u_n}$. A short argument using the polar decomposition of H_b and the obvious identity $H_b^* = \overline{H_b}$ shows that the u_n's can be chosen to satisfy

$$\sigma_n \overline{z \breve{u_n}} = H_b \breve{u_n}. \tag{5.4}$$

(Beylkin and Monzón calls these vectors con-eigenvectors, following [9]). Now consider n to be fixed and set $v = \sigma_n \frac{\overline{z u_n}}{u_n}$. Note that

$$\|v\|_{L^\infty} = \sigma_n \tag{5.5}$$

and $\sigma_n \overline{z \breve{u_n}} = H_v \breve{u_n}$. In particular, $0 = H_b \breve{u_n} - H_v \breve{u_n} = H_{b-v} \breve{u_n}$ so

$$P_+ \left(\overline{z(b-v)} \right) \perp inv_{M_z}(\breve{u_n}) \tag{5.6}$$

by (5.3). We are now almost in a position to give an briefly outline the proof of the AAK-theorem. The missing, and hard, piece lies in the following curious lemma.

Lemma 5.4. *$inv_{M_z}(\breve{u_n})$ has codimension n.*

See [4] and [13] for two completely different proofs. To see that Theorem 2.2 follows, let $a \in l^2(\mathbb{Z}^+)$ be as in Theorem 2.2 and let $b \in L^\infty(\mathbb{T})$ be such that

$$P_-(b) = \sum_{j=0}^{\infty} a(j) z^{-j-1}$$

in analogy with the remarks following (5.2). Let n be fixed. By Lemma 5.4 $\breve{u_n}$ has precisely n zeroes, which is the first part of Theorem 2.2. To see the second part, let us denote the zeroes of $\breve{u_n}$ by $z_1, \ldots, z_n$, and recall that these are assumed to be distinct. By equation (5.6) and Proposition 5.1 we conclude that

$$P_+(\overline{z(b-v)}) \in \mathsf{Span}\ \{k_{z_1}, \ldots, k_{z_n}\}.$$

For any $f \in L^2(\mathbb{T})$, we have $P_- f = \overline{z P_+ \overline{z f}}$, and hence we may choose $c_1, \ldots, c_n \in \mathbb{C}$ such that

$$P_-(b-v) = \sum_{j=1}^{n} c_j \overline{z k_{z_j}}. \tag{5.7}$$

Note that $H_{\sum_{j=1}^n c_j \overline{z k_{z_j}}}$ corresponds to the Hankel matrix $\Gamma_{\sum_{j=1}^n c_j \boxed{z_j}}$, (cf (2.14)), and hence Theorem 2.2 follows if we show that

$$\|H_b - H_{\sum_{j=1}^n c_j \overline{z k_{z_j}}}\| = \sigma_n. \tag{5.8}$$

But the analytic part of a function does not affect the Hankel operator, and hence by (5.5) and (5.7) we have

$$\|H_b - H_{\sum_{j=1}^n c_j \overline{z k_{z_j}}}\| = \|H_{P_- b} - H_{\sum_{j=1}^n c_j \overline{z k_{z_j}}}\| = \|H_{P_- v}\| = \|H_v\| \leqslant \|v\|_{L^\infty} = \sigma_n.$$

Conversely

$$\|H_b - H_{\sum_{j=1}^n c_j \overline{z k_{z_j}}}\| \geqslant \|H_v \breve{u_n}\| = \|\sigma_n \overline{z \breve{u_n}}\| = \sigma_n.$$

The "proof" of the AAK-theorem is complete.

For future reference, we state the Adamjan, Arov, Krein theorem for compact Hankel operators in the $H^2(\mathbb{D})$ environment, which is the usual form.

THEOREM 5.5. *Let $b \in C(\mathbb{T})$ be given and let H_b, σ_n, u_n etc. be as above. Let $\sigma_n > 0$ be distinct and assume that all zeros to the $\breve{u_n}$ are simple. Then u_n has exactly n zeroes $z_1, \ldots, z_n \in \mathbb{D}$ and there is a rational function r with simple poles at $\overline{z_1}^{-1}, \ldots, \overline{z_n}^{-1}$ such that*

$$\|H_b - H_{\overline{r}}\| = \sigma_n$$

Moreover, H_r has rank n.

Proof. The main part is clear by noting that

$$\overline{zk_{z_j}} = P_{-}\overline{\overline{z_j}^{-1}k_{z_j}}$$

which combined with (5.8) implies that $r = \sum_{j=1}^{n} \frac{c_j}{\overline{z_j}^{-1}} k_{z_j}$ is the desired function. The rank of H_r is clearly $\leqslant n$ as

$$H_r = \sum_{j=1}^{n} c_j H_{\overline{zk_{z_j}}}$$

and $H_{\overline{zk_{z_j}}}$ has rank 1, which is easily seen by considering the corresponding Hankel matrix $\Gamma_{\boxed{z_j}}$. To see that the rank can not be less than n, note that

$$\sigma_n = \inf\{\|H_b - K\| : K, \mathsf{Rank}(K) \leqslant n\}$$

so if $\mathsf{Rank} H_{\overline{r}} < n$, then $\sigma_{n-1} \leqslant \|H_b - H_{\overline{r}}\| = \sigma_n$, which contradicts the assumption σ_n is distinct. □

Remark: The assumption $b \in C(\mathbb{T})$ is not necessary but simplifies the statement.

5.3. Hankel operators on $\mathbb{R}$. We will first briefly recapitulate some of the corresponding theory for $H^2(\mathbb{C}^+)$ and its Hankel operators, and towards the end we will derive some results that we have not been able to find in the literature. $\mathbb{C}^+$ stands for the open upper half plane in $\mathbb{C}$, i.e. $\{\zeta : \mathsf{Im}\zeta > 0\} = \mathbb{C}^+$. Much of the theory is similar to that of $H^2(\mathbb{D})$ and we will therefore often use the same symbols for corresponding objects, occasionally separating them by the supindex $\mathbb{T}$ and $\mathbb{R}$ respectively. We will reserve the letters z, ζ, x for the independent variable in $\mathbb{D}$, $\mathbb{C}^+$ and $\mathbb{R}^+$ respectively. In expressions like $\mathrm{e}^{\mathrm{i}x}$ we will without comment consider x to be the identity function on $\mathbb{R}$, (so $\mathrm{e}^{\mathrm{i}x}$ is a function, not a number.) Also, ϕ will typically denote functions in $H^2(\mathbb{D})$, whilst Φ and F belong to $H^2(\mathbb{C}^+)$ and $L^2(\mathbb{R}^+)$ respectively. We will treat $L^2(\mathbb{R}^+)$ as a space of functions identically 0 on $\mathbb{R}^-$ and we will in the remainder use the notation $\chi_{\mathbb{R}^+}$ instead of $\chi(\mathbb{R}^+, x)$ to denote the characteristic function of the interval $\mathbb{R}^+$.

Let $\mathcal{F}$ denote the unitary Fourier transform defined by

$$\mathcal{F}(\Phi)(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \Phi(\zeta)\mathrm{e}^{-\mathrm{i}x\zeta} d\zeta.$$

One defines $H^2(\mathbb{C}^+)$ as a space of analytic functions on $\mathbb{C}^+$ in analogy with $H^2(\mathbb{D})$, where the integrals in (5.1) are replaced by integrals on lines parallel with $\mathbb{R}$. Equivalently, one may consider $H^2(\mathbb{C}^+)$ as $\mathcal{F}^{-1}(L^2(\mathbb{R}^+))$. Given $\Phi \in H^2(\mathbb{C}^+)$ the corresponding function on $\mathbb{R}$, (also denoted by Φ), is given *a.e.* by non-tangential limits and for $\zeta \in \mathbb{C}^+$ and we have

$$\Phi(\zeta) = \mathcal{F}^{-1}(\hat{\Phi}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \hat{\Phi}(x)\mathrm{e}^{\mathrm{i}\zeta x} dx,$$

where as usual $\hat{\Phi} = \mathcal{F}(\Phi)$. Note that $\hat{\Phi} \in L^2(\mathbb{R}^+)$ by the Paley-Wiener theorem. Given any $\zeta_0 \in \mathbb{C}^+$, the above formula implies that

$$\Phi(\zeta_0) = \left\langle \Phi, \mathcal{F}^{-1}\left(\overline{\chi_{\mathbb{R}^+}\mathrm{e}^{\mathrm{i}\zeta_0 x}}\right) \right\rangle = \left\langle \Phi, (\sqrt{2\pi})^{-1} \frac{-1}{\mathrm{i}(\zeta - \overline{\zeta_0})} \right\rangle \tag{5.9}$$

so we see that

$$k_{\zeta_0}(\zeta) = \frac{\mathrm{i}}{\sqrt{2\pi}(\zeta - \overline{\zeta_0})} \tag{5.10}$$

is the reproducing kernel for ζ_0.

Let $\alpha : \mathbb{D} \to \mathbb{C}^+$ be the analytic bijection given by

$$\alpha(z) = \mathrm{i}\frac{1+z}{1-z},$$

and let $\beta : \mathbb{C}^+ \to \mathbb{D}$ denote its inverse $\beta(\zeta) = \frac{\zeta - \mathrm{i}}{\zeta + \mathrm{i}}$. Moreover let $\mathcal{U} : L^2(\mathbb{T}) \to L^2(\mathbb{R})$ be the unitary map given by

$$\mathcal{U}(\phi)(\zeta) = \frac{\phi(\beta(\zeta))}{\sqrt{\pi}(\zeta + \mathrm{i})}.$$

The inverse is

$$\mathcal{U}^{-1}(\Phi)(z) = \frac{2\sqrt{\pi}\mathrm{i}\Phi(\alpha(z))}{1-z}.$$

It can be shown that $\mathcal{U}$ maps $H^2(\mathbb{D})$ onto $H^2(\mathbb{C}^+)$, (see e.g. [10]), and hence $\mathsf{Ran}\ \mathcal{U}|_{H^2_-(\mathbb{D})} = L^2(\mathbb{R}) \ominus H^2(\mathbb{C}^+)$. Explicitly this can be seen by noting that if $\phi \in H^2(\mathbb{D})$ then

$$\mathcal{U}(\overline{z\phi}) = \overline{\mathcal{U}(\phi)}. \tag{5.11}$$

Set $L^2(\mathbb{R}) \ominus H^2(\mathbb{C}^+) = H^2_-(\mathbb{C}^+)$. In analogy with section (5.1), we let P_+ denote the projection from $L^2(\mathbb{R})$ onto either $H^2(\mathbb{C}^+)$ or $L^2(\mathbb{R}^+)$, where the context determines which one is intended. If there is risk for confusion we will denote them $P_{H^2(\mathbb{C}^+)}$ and $P_{L^2(\mathbb{R}^+)}$. P_- is defined similarly. Note that $\mathcal{F}P_{H^2(\mathbb{C}^+)} = P_{L^2(\mathbb{R}^+)}\mathcal{F}$ whereas $\mathcal{F}^{-1}P_{H^2(\mathbb{C}^+)} = P_{L^2(\mathbb{R}^-)}\mathcal{F}^{-1}$. Moreover note that $\mathcal{U}P_{H^2(\mathbb{D})} = P_{H^2(\mathbb{C}^+)}\mathcal{U}$, that

$$\mathcal{U}|_{H^2(\mathbb{D})}(\{\phi : \phi(z_0) = 0\}) = \{\Phi : \Phi(\alpha(z_0)) = 0\} \tag{5.12}$$

and in particular that

$$\mathcal{U}|_{H^2(\mathbb{D})}(k_{z_0}) = \frac{-\sqrt{2}\mathrm{i}}{(1-\overline{z_0})}\ k_{\alpha(z_0)} \tag{5.13}$$

We now turn to the Hankel operators on $H^2(\mathbb{C}^+)$. In the beginning of Section 4 we have already argued that the counterpart of Hankel matrices $\Gamma_a : l^2(\mathbb{Z}^+) \to l^2(\mathbb{Z}^+)$ on $L^2(\mathbb{R}^+)$ should be operators of the form

$$\Gamma_A(F) = \int_0^\infty A(x+y)F(y)\ dy$$

where A is a function in $L^1(R_+)$. We will now conclude the same thing from a different angle, namely by showing that each real line Hankel operator Γ_A is actually unitarily equivalent to a traditional Hankel operator Γ_a. The converse is also true, but one then has to extend the definition of Γ_A to include symbols A that are certain distributions. We will have no need for this so we omit it, the details can be found in [13], Chapter 1.8. For any $B \in L^\infty(\mathbb{R})$ we define $H_B : H^2(\mathbb{C}^+) \to H^2_-(\mathbb{C}^+)$ via

$$H_B(\Phi) = P_{H^2_-(\mathbb{C}^+)}(B\Phi),$$

The following proposition shows that the analogues of the results from Section 5.2, like Nehari's and Hartman's theorems, transfer immediately to H_B and Γ_A.

PROPOSITION 5.6. *Given $A \in L^1(\mathbb{R}^+)$, set $B = \sqrt{2\pi}\hat{A}$ and $b = B \circ \alpha$. Then $\Gamma_A = \mathcal{F}^{-1} H_B \mathcal{F}^{-1}$ and $\mathcal{U}^{-1} H_B \mathcal{U} = H_b$. In particular, if $\breve{u_n}$ is a singular vector to H_b with zero set $\{z_k\}$, then $U_n = \mathcal{F}\mathcal{U}(\breve{u_n})$ is a singular vector to Γ_A and $\zeta_k = \alpha(\{z_k\})$ are the zeroes of $\breve{U_n}$.*

Proof. For any $F \in L^2(\mathbb{R}^+)$ we have

$$\mathcal{F}^{-1} H_B \mathcal{F}^{-1}(F) = P_{L^2(\mathbb{R}^+)} \mathcal{F}^{-1}(B\check{F}) = (2\pi)^{-1/2} P_+(\check{B} * \check{\check{F}}) =$$
$$= (2\pi)^{-1/2} P_+ \int_{\mathbb{R}} \check{B}(x-y) F(-y)dy = (2\pi)^{-1/2} P_+ \int_0^\infty \check{B}(x+y)F(y)dy.$$

which establishes the first claim. For the second, we have

$$\mathcal{U}^{-1} H_B \mathcal{U}\phi = \mathcal{U}^{-1} P_-(B\, \mathcal{U}\phi) = P_- \mathcal{U}^{-1}(B\, \mathcal{U}\phi) = P_-(B \circ \alpha\, \mathcal{U}^{-1}\mathcal{U}\phi) = H_{B\circ\alpha}(\phi).$$

The statement about the singular vectors is an obvious consequence of these unitary equivalences and (5.12). □

We can now easily show the following theorem, which is a combination of Hartman's theorem and Lemma 5.4:

THEOREM 5.7. *Given any $A \in L^1(\mathbb{R}^+)$, Γ_A is a compact operator. Then there is a basis of singular vectors $U_1, U_2, \ldots$ such that $\sigma_n \overline{U_n} = \Gamma_A U_n$ for all n. If the non-zero singular values are distinct, then each $\breve{U_n}$ has exactly n zeroes $\zeta_1, \ldots, \zeta_n$ in $\mathbb{C}^+$, (repeated according to multiplicity).*

Proof. Define B and b as in Proposition 5.6. As $A \in L^1(\mathbb{R})$ it follows that B is continuous and approaches zero at infinity. Hence $b = B \circ \alpha$ is continuous on $\mathbb{T}$, (even at $z = 1$), so by Theorem 5.3 we get that H_b is compact. Thus Γ_A is compact as well by Proposition 5.6. Let σ_n be the singular values of H_b. By (5.4) we can choose the singular vectors $\breve{u_n}$ such that $\sigma_n \overline{z\breve{u_n}} = H_b \breve{u_n}$. Set $U_n = \mathcal{F}\mathcal{U}(u_n)$. Using (5.11) and Proposition 5.6 we get

$$\Gamma_A U_n = \Gamma_A \mathcal{F}\mathcal{U}(\breve{u_n}) = \mathcal{F}^{-1} H_B \mathcal{U}(\breve{u_n}) = \mathcal{F}^{-1}\mathcal{U} H_b(\breve{u_n}) =$$
$$\mathcal{F}^{-1}\mathcal{U}(\sigma_n \overline{z\breve{u_n}}) = \sigma_n \mathcal{F}^{-1}\overline{\mathcal{U}(\breve{u_n})} = \sigma_n \overline{\mathcal{F}\mathcal{U}(\breve{u_n})} = \sigma_n \overline{U_n}$$

which establishes the second statement. The third follows immediately from Theorem 5.5 and Proposition 5.6. □

By combining Proposition 5.6 and Nehari's theorem (Theorem 5.2) we immediately get

THEOREM 5.8. *Given any $A \in L^1(\mathbb{R}^+)$ set $B = \sqrt{2\pi}\hat{A}$. Then $\|\Gamma_A\| = \|B\|_{L^\infty/H^\infty(\mathbb{C}^+)}$.*

The equality $\|H_b - H_{\sum_{j=1}^n c_j \overline{z k_{z_j}}}\| = \sigma_n$, (i.e. (5.8)), and Proposition 5.6 can however not be combined directly to show a corresponding equality for H_B, because $(\overline{z k_{z_j}}) \circ \beta$ does not equal $\overline{k_{\alpha(z_j)}}$. Nevertheless, the result is true.

THEOREM 5.9. *Given $A \in L^1(\mathbb{R}^+)$, let U_n be the singular vectors to Γ_A. Then there are coefficients $c_1, \ldots, c_n \in \mathbb{C}$ such that*

$$\|\Gamma_A - \sum_{j=1}^n c_j \Gamma_{e^{i\zeta_j x}}\| = \sigma_n.$$

Proof. By (5.9) and (5.10) we get

$$\mathcal{F}(\chi_{\mathbb{R}^+} e^{i\zeta_j x}) = \overline{k_{\zeta_j}}. \tag{5.14}$$

Thus, with $B = \sqrt{2\pi}\hat{A}$, Proposition 5.6 implies that the theorem is equivalent to

$$\|H_B - \sqrt{2\pi} \sum_{j=1}^n c_j H_{\overline{k_{\zeta_j}}}\| = \sigma_n. \tag{5.15}$$

To see that (5.15) indeed holds, apply Theorem 5.5 to $b = B \circ \alpha$ and use Proposition 5.6 and to conclude that the rational function $R = r \circ \beta$ has simple poles at the points $\alpha(\overline{z_1^{-1}}), \dots, \alpha(\overline{z_n^{-1}})$, no other poles and satisfies

$$\|H_B - H_{\overline{R}}\| = \sigma_n \tag{5.16}$$

Note that

$$\alpha\left(\overline{z^{-1}}\right) = \mathrm{i}\frac{1 + \overline{z^{-1}}}{1 - \overline{z^{-1}}} = -\mathrm{i}\frac{1 + \overline{z}}{1 - \overline{z}} = \overline{\alpha(z)}$$

and $\alpha(z_j) = \zeta_j$ by Proposition 5.6. Now, $R(\infty) = r(1)$ so we can find a polynomial p with $\deg p < n$ such that

$$R(\zeta) - r(1) = \frac{p(\zeta)}{\prod(\zeta - \overline{\zeta_j})}.$$

If we consider the functions k_{ζ_j} as rational functions in the obvious way (cf. eq. (5.10)), it follows by elementary calculus that there are coefficients $c_1, \dots, c_n$ such that

$$R = r(1) + \sqrt{2\pi}\sum_{j=1}^{n} \overline{c_j} k_{\zeta_j} \tag{5.17}$$

which combined with (5.16) and the fact that $H_{r(1)} = 0$ gives (5.15). □

5.4. What does it mean that $\|\Gamma_A - \sum_{j=1}^n c_j \Gamma_{e^{i\zeta_j x}}\|$ is small?. Given any $A \in L^1(\mathbb{R}^+)$, we will in the next section show that the algorithm from Section 3 produces approximations $A_{ap}(N)$ that satisfy

$$\|\Gamma_{A - A_{ap}}\| =\approx \sigma_n,$$

where $\{\sigma_k\}$ are the singular values of Γ_A. It is not so clear what this actually means in terms of standard norms, so to shed some light here we offer a few estimates, neither of which is perfect.

Let m denote the Lebesgue measure on R. Given an interval $I \subset \mathbb{R}$ and a locally integrable function F, set $F_I = m(I)^{-1}\int_I |F| dm$. Recall that the BMO-space on $\mathbb{R}$ is defined as

$$\{F \in L^1_{loc} : sup_I\left\{m(I)^{-1}\int_I |F - F_I| dm\right\},$$

where the supremum is taken over all finite intervals $I \subset \mathbb{R}$. Note that $\|1\|_{BMO} = 0$, so it is only a semi-norm.

PROPOSITION 5.10. *There exists a constant C such that*

$$C\|\Gamma_A\| \leqslant \|\hat{A}\|_{BMO} \leqslant 2.6\|\Gamma_A\|$$

for all $A \in L^1(\mathbb{R}^+)$.

Proof. Set $B = \sqrt{2\pi}\hat{A}$. By Theorem 5.8 we have $\|\Gamma_A\| = \|B\|_{L^\infty/H^\infty}$ and it follows from the celebrated result of C. Fefferman that the right hand side is bounded above and below by $\|B\|_{BMO}$. As we have not been able to find the values of the constants involved in print, we outline the argument. We refer to [10] for proofs of the formulas to be used below. Write $B = B_r - \mathrm{i}\tilde{B}_r$, where $B_r = \mathrm{Re}B$ and $\tilde{B}_r$ denotes the Hilbert transform of B_r. This can be done as B is anti-analytic and bounded in $\mathbb{C}^+$. Recall the formulas $\int \tilde{\Phi}\tilde{\Psi} dm = \int \Phi\Psi dm$ and $\tilde{\tilde{\Phi}} = -\Phi$. Note that $L^\infty(\mathbb{R})/H^\infty(\mathbb{C}^+) = H^1(\mathbb{C}^+)^*$, so we can take $D_k \in L^\infty$ and $E_k \in L^\infty$, $k = 1, 2, \dots$, to be real valued functions such that $\int \Phi(D_k + iE_k) dm = \int \Phi B dm$ for all $\Phi \in H^1(\mathbb{C}^+)$ and $\lim_k \|D_k + iE_k\|_{L^\infty} = \|B\|_{L^\infty/H^\infty}$. Now

$$\int \Phi_r(D_k + \widetilde{E_k}) dm = \int \Phi_r D_k - \widetilde{\Phi_r} E_k dm = \mathrm{Re}\int \Phi(D_k + iE_k) dm =$$

$$= \mathrm{Re}\int \Phi B dm = \int \Phi_r B_r + \widetilde{\Phi_r}\widetilde{B_r} = 2\int \Phi_r B_r dm$$

for all $\Phi \in H^1(\mathbb{C}^+)$ so $2B_r = (D_k + \widetilde{E_k})$. By calculations very similar to those in Section X, [10], we get $\|D_k\|_{BMO} \leqslant 2\|D_k\|_{L^\infty}$ and $\|\widetilde{E_k}\| \leqslant (1+2\sqrt{3})\|E_k\|_{L^\infty}$. Thus

$$2\|B_r\|_{BMO} \leqslant \lim_k (2\|D_k\|_{L^\infty} + (1+2\sqrt{3})\|E_k\|_{L^\infty}) \leqslant (2+(1+2\sqrt{3}))\|B\|_{L^\infty/H^\infty}.$$

By analogous calculations for the imaginary part we obtain $2\|\widetilde{B_r}\|_{BMO} \leqslant (3+2\sqrt{3})\|B\|_{L^\infty/H^\infty}$ and hence

$$\|B\|_{BMO} = \|B_r - \mathrm{i}\widetilde{B_r}\|_{BMO} \leqslant \|B_r\|_{BMO} + \|\widetilde{B_r}\|_{BMO} \leqslant (3+2\sqrt{3})\|B\|_{L^\infty/H^\infty}$$

so that

$$\|\sqrt{2\pi}\hat{A}\|_{BMO} \leqslant (3+2\sqrt{3})\|B\|_{L^\infty/H^\infty} = (3+2\sqrt{3})\|\Gamma_A\|.$$

Finally, $(3+2\sqrt{3})/\sqrt{2\pi} \approx 2.5788$. □

COROLLARY 5.11. *Let A etc. be as in Theorem 5.9. Then*

$$C\sigma_n \leqslant \|\mathcal{F}(A - \chi_{\mathbb{R}^+} \sum_{j=1}^n c_j e^{\mathrm{i}\zeta_j x})\|_{BMO} \leqslant 2.6\sigma_n$$

We include a few other ways of estimating $\|\Gamma_A - \sum_{j=1}^n c_j \Gamma_{e^{\mathrm{i}\zeta_j x}}\|$. First a lemma.

LEMMA 5.12. *Let B, R, U_n etc. be as in Theorem 5.9. Put $v = \sigma_n z\overline{\breve{u}_n}/\breve{u}_n$ and $V = \sigma_n \overline{\breve{u}_n}/\breve{u}_n$. Then*

$$V \circ \alpha = v$$

and

$$b - \overline{r} - (b(0) - \overline{r(0)}) = P_- v.$$

Proof. For $\zeta \in \mathbb{R}$ we have

$$V = \sigma_n \frac{\overline{\breve{u}_n}}{\breve{u}_n} = \sigma_n \frac{\zeta + \mathrm{i}}{\zeta - \mathrm{i}} \frac{\overline{\breve{u}_n \circ \beta}}{\breve{u}_n \circ \beta} = \sigma_n \beta \frac{\overline{\breve{u}_n \circ \beta}}{\breve{u}_n \circ \beta} = v \circ \beta.$$

From (5.7) and the proof of Theorem 5.5 we conclude that $P_-(b - \overline{r}) = P_- v$, which, as $\overline{b} \in H^\infty(\mathbb{D})$, gives the second identity. □

Let δ denotes the Dirac distribution at 0 and let $W^{-1,2}$ denotes the Sobolev space on $\mathbb{R}$.

PROPOSITION 5.13. *Let A etc. be as in Theorem 5.9. Then there exists a $c_0 \in \mathbb{C}$ such that*

$$\|A - \sum_{j=1}^n c_j e^{\mathrm{i}\zeta_j x} + c_0\delta\|_{W^{-1,2}} \leqslant \sqrt{\frac{\pi}{2}}\sigma_n$$

Proof. By Proposition 5.12 we have

$$b - \overline{r} - (b(0) - \overline{r(0)}) = P_- v \tag{5.18}$$

which, as $P_- = P_{H^2_-(\mathbb{D})}$ is a contraction, implies that

$$\|b - \overline{r} - (b(0) - \overline{r(0)})\|_{L^2} \leqslant \|v\|_{L^2} = \sigma_n.$$

This yields

$$\int \left|B-\overline{R}-(b(0)-\overline{r(0)})\right|^2 \frac{1}{\pi^2(1+\zeta^2)} dm = \|\frac{B-\overline{R}-(b(0)-\overline{r(0)})}{\pi(\mathrm{i}+\zeta)}\|_{L^2}^2 =$$
$$= \|\mathcal{U}(b-\overline{r}-(b(0)-\overline{r(0)}))\|_{L^2}^2 = \|b-\overline{r}-(b(0)-\overline{r(0)})\|_{L^2}^2 \leqslant \sigma_n^2.$$

which in turn, combined with (5.17), gives

$$\int \left|B-\sqrt{2\pi}\sum_{j=1}^{n} c_j\overline{k_{\zeta_j}} - (b(0)+\overline{r(1)-r(0)})\right|^2 \frac{1}{(1+\zeta^2)} dm \leqslant (\pi\sigma_n)^2.$$

Now, $b(1)=B(\infty)=0$, so evaluation of (5.18) at 1 gives $b(0)+\overline{r(1)-r(0)} = -(P_- v)(1)$ and moreover $\sqrt{2\pi}\mathcal{F}(A-\chi_{\mathbb{R}^+}\sum_{j=1}^n c_j \mathrm{e}^{\mathrm{i}\zeta_j x}) = B-\sqrt{2\pi}\sum_{j=1}^n c_j\overline{k_{\zeta_j}}$ by Proposition 5.6 and (5.14). Summing up, we have shown that

$$\|A-\chi_{\mathbb{R}^+}\sum_{j=1}^{n} c_j \mathrm{e}^{\mathrm{i}\zeta_j x} + (P_- v)(1)\delta\|_{W^{-1,2}} \leqslant \sqrt{\frac{\pi}{2}}\sigma_n.$$

□

Our final estimate is a little trickier and goes as follows: Let A etc. be as in Theorem 5.9. Note that $\overline{(B-\sqrt{2\pi}\sum_{j=1}^n c_j\overline{k_{\zeta_j}})} \in \mathcal{H}^2(\mathbb{C}^+)$. Lets temporarily set $\Phi = \overline{(B-\sqrt{2\pi}\sum_{j=1}^n c_j\overline{k_{\zeta_j}})}$. We then have

$$\sigma_n = \|\Gamma_A - \sum_{j=1}^{n} c_j\Gamma_{\mathrm{e}^{\mathrm{i}\zeta_j x}}\| = \|H_{\overline{\Phi}}\| \geqslant \|H_{\overline{\Phi}}\Phi\|_{L^2}\Big/\|\Phi\|_{L^2} = \|P_-|\Phi|^2\|_{L^2}\Big/\|\Phi\|_{L^2}$$
$$= \|P_-|\Phi|^2\|_{L^2}\Big/\|\Phi\|_{L^2} = \|\Phi^2\|_{L^2}\Big/2\|\Phi\|_{L^2} = \|(A-\sum_{j=1}^{n} c_j \mathrm{e}^{\mathrm{i}\zeta_j x})^2\|_{L^2(\mathbb{R}^+)}\Big/2\|A-\sum_{j=1}^{n} c_j \mathrm{e}^{\mathrm{i}\zeta_j x}\|_{L^2(\mathbb{R}^+)},$$

i.e.

$$\|(A-\sum_{j=1}^{n} c_j \mathrm{e}^{\mathrm{i}\zeta_j x})^2\|_{L^2(\mathbb{R}^+)} \leqslant 2\sigma_n\|A-\sum_{j=1}^{n} c_j \mathrm{e}^{\mathrm{i}\zeta_j x}\|_{L^2(\mathbb{R}^+)}.$$

Set $F=(A-\chi_{\mathbb{R}^+}\sum_{j=1}^n c_j \mathrm{e}^{\mathrm{i}\zeta_j x})$ and suppose we know that $\int_R^\infty |F|^2 dm \leqslant \epsilon$, where R is not very large and ϵ is small. In terms of applications, this is a realistic estimate, because the functions A will have compact support and we may therefore expect the $\mathrm{e}^{\mathrm{i}\zeta_j x}$'s to decay rather quickly. Put

$$\mu_F(t) = \{x : \lambda(|F(x)|) \geqslant t\}.$$

If $\lambda : \mathbb{R}^+ \to \mathbb{R}^+$ is an increasing C^1-function, recall that

$$\int \lambda(|F|)dm = \int_0^\infty \mu_F(t)\lambda'(t)dt.$$

The above then yields

$$\int_0^\infty \mu_F(t)4t^3 dt \leqslant 4\sigma_n^2 \int \mu_F(t)2tdt \Rightarrow$$

$$0 \leqslant 4\int_0^\infty \mu_F(t)(2\sigma_n^2 - t^2)tdt \leqslant 4\int_0^{\sqrt{2}\sigma_n} \mu_F(t)(2\sigma_n^2)tdt - 4\int_{\sqrt{2}\sigma_n}^\infty \mu_F(t)(t^2-2\sigma_n^2)tdt$$

Now,

$$\int_0^{\sqrt{2}\sigma_n} \mu_F(t)(2\sigma_n^2)tdt = \sigma_n^2 \int \inf(2\sigma_n^2, |F(x)|^2)dx \leqslant \sigma_n^2(2\sigma_n^2 R + \epsilon)$$

and, setting $\lambda(t) = \sup(0, (t-\sqrt{2}\sigma_n)^4)$, we have $4(t^2-2\sigma_n^2)t - \lambda'(t) \geqslant 0$ for $t \geqslant \sqrt{2}\sigma_n$, which implies that

$$4\int_{\sqrt{2}\sigma_n}^{\infty} \mu_F(t)(t^2 - 2\sigma_n^2)tdt \geqslant \int_{\sqrt{2}\sigma_n}^{\infty} \mu_F(t)\lambda'(t)tdt = \int \lambda(|F(x)|)dx.$$

Summing up, we have shown that

$$\int \lambda(|F(x)|)dx \leqslant 4\sigma_n^2(2\sigma_n^2 R + \epsilon)$$

If, for example, we set $\epsilon = \sigma_n^2$ then we achieve

$$\|A - \sum_{j=1}^{n} c_j e^{i\zeta_j x}\|_{L^4(\mathbb{R}^+)} \approx \left(\int_0^{\infty} \lambda\left(\left|A - \chi_{\mathbb{R}^+}\sum_{j=1}^{n} c_j e^{i\zeta_j x}\right|\right)dm\right)^{1/4} \leqslant \sigma_n(8R+1)^{1/4}$$

6. Algorithm 1, continued. We are now in a position to prove that $\zeta_k(N)$ and $c_k(N)$ converge. Recall that we are given a function $A \in C_0^1(\mathbb{R}^+)$ with finite support. For simplicity of notation have assumed that the non-zero singular values of Γ_A are distinct. By Proposition 4.1 and Lemma 4.2, this is then true for $\Gamma_{\mathcal{S}_{1,N}A}$ as well, provided that N is sufficiently large. We will in this section assume that this is the case.

Recall that σ_n denotes the singular values of Γ_A and that $\sigma_n(N)$ denotes the singular values of $\Gamma_{\mathcal{S}_{1,N}A}$. Recall that

$$\lim_N \sigma_n(N) = \sigma_n$$

by Corollary 4.3. We consider n to be fixed. Recall that $\{z_k(N)\}_{k=1}^n$ denotes the zero set in $\mathbb{D}$ to $u_n(N)$, that $\zeta_k(N) = -iN\log z_k(N)$, that $c_k(N)$ were certain coefficients such that, with

$$A_{ap}(N)(x) = \sum_{k=1}^{n} c_k(N)e^{i\zeta_k(N)x},$$

we had

$$\|\Gamma_{\mathcal{S}(A-A_{ap}(N))}\| = \sigma_n(N)$$

by Corollary 3.1.

We let U_n denote the singular vector corresponding to σ_n to Γ_A, $\{\zeta_k\}_{k=1}^n$ denote the corresponding zeroes to $\breve{u}_n$ in $\mathbb{C}^+$ and we let c_k be the coefficients such that $A_{ap}(x) = \sum_{k=1}^n c_k e^{i\zeta_k x}$ satisfies

$$\|\Gamma_{A-A_{ap}}\| = \sigma_n,$$

as guaranteed by Theorem 5.9.

THEOREM 6.1. *If $A \in C_0^1(\mathbb{R}^+)$ is such that Γ_A has distinct non-zero singular values, then*

$$\lim_N \zeta_k(N) = \zeta_k$$

and

$$\lim_N c_k(N) = c_k$$

for all $1 \leqslant k \leqslant n$.

Proof. Recall the definition of $\mathcal{S}_{2,N}^{inv}$ (4.1) and Proposition 4.1. It is clear that $\mathcal{S}_{2,N}^{inv}u_n(N)$ is a singular vector for $\mathcal{S}_{2,N}^{inv}\Gamma_{\mathcal{S}_{1,N}A}\mathcal{S}_{2,N}$. By Proposition 4.1 and Lemma 4.2, it follows that there are numbers $\beta_{n,N} \in \mathbb{T}$ such that

$$\lim_N \mathcal{S}_{2,N}^{inv}\beta_{n,N}u_n(N) = U_n. \tag{6.1}$$

For notational simplicity, we redefine the $u_n(N)$'s so that the above formula holds without $\beta_{n,N}$'s. We will now consider $\mathcal{S}_{2,N}^{inv}u_n(N)$ and U_n as functions on $\mathbb{R}$ that are zero on $\mathbb{R}^-$. It is a standard fact that (6.1) implies that

$$\lim_N \mathcal{F}^{-1}(\mathcal{S}_{2,N}^{inv}u_n(N)) = \widetilde{u_n} \tag{6.2}$$

uniformly on compact subsets of $\mathbb{C}^+$. Note that for $\zeta \in \mathbb{C}^+$ we also have

$$\begin{aligned} \mathcal{F}^{-1}(\mathcal{S}_{2,N}^{inv}u_n(N)) &= \frac{1}{\sqrt{2\pi}}\int_0^\infty (\mathcal{S}_{2,N}^{inv}u_n(N))(x)e^{i\zeta x}dx \approx \\ &\approx \frac{1}{\sqrt{2\pi}N}\sum_{k=0}^{\infty}\sqrt{N}(u_n(N))_k e^{i\zeta k/N} = \frac{1}{\sqrt{2\pi N}}\widetilde{u_n(N)}(e^{i\zeta/N}) \end{aligned} \tag{6.3}$$

because the third term is a Riemann sum approximation of the integral. More precisely, let $K \in \mathbb{C}^+$ be compact and set

$$\delta_{K,N} = \sup\{|1 - e^{i\zeta x}| : (\zeta, x) \in K \times [0, N^{-1}]\}.$$

Note that $\delta_{K,N} \to 0$ as $N \to \infty$. For $\zeta \in K$ we can estimate the difference between the upper and lower part of (6.3) as follows

$$\begin{aligned} &\left|\int_0^\infty (\mathcal{S}_{2,N}^{inv}u_n(N))(x)e^{i\zeta x} - N^{-1}\sum_{k=0}^{\infty}\sqrt{N}(u_n(N))_k e^{i\zeta k/N}\right| \leqslant \\ &\leqslant N^{-1}\sum_k \sup_{x\in[k/N,(k+1)/N]}\left|\sqrt{N}(u_n(N))_k(e^{i\zeta k/N} - e^{i\zeta x})\right| \leqslant \\ &\leqslant N^{-1/2}\sum_k |(u_n(N))_k|\delta_{K,N}|e^{i\zeta k/N}| \leqslant \delta_{K,N}\|u_n(N)\|_{l^2}\sqrt{\sum_k \frac{e^{-2\mathrm{Im}\zeta k/N}}{N}} \leqslant \\ &\leqslant \delta_{K,N}\sqrt{\frac{1}{N(1-e^{-2\mathrm{Im}\zeta/N})}} \leqslant \delta_{K,N}\sqrt{\frac{1}{2\mathrm{Im}\zeta}} \end{aligned}$$

Combining this with (6.2) and (6.3) we conclude that

$$\lim_N \frac{1}{\sqrt{2\pi N}}\widetilde{u_n(N)}(e^{i\zeta/N}) = \widetilde{u_n}(\zeta)$$

uniformly for $\zeta \in K$. By Theorem 2.2 the left hand side has precisely n zeroes for each (sufficiently large) N, and by definition these zeros are precisely the points $\zeta_k(N)$, $k = 1, \ldots, N$. Also the right hand side has precisely n zeroes, by Theorem 5.9, which we have denoted by ζ_k. By the uniform convergence and Vitalis theorem, it follows that we can reindex the $\zeta_k(N)$'s such that

$$\lim_N \zeta_k(N) = \zeta_k, \quad \forall k = 1, \ldots, n,$$

which proves the first statement.

The second statement follows from a careful inspection of how the $c_k(N)$ and c_k's are chosen. As this part is less interesting for our purposes, we omit this part of the proof. □

In summary, for any $n \in \mathbb{N}$ the algorithm gives an approximation A_{ap} of a given function A with n terms with an error that is approximately equal to σ_n, given that the sampling interval $1/N$ is taken small enough. The main drawback is that the error is measured in a rather obscure norm, which is hard to estimate by more common norms, although we offer some results in this direction in Section 5.4. In particular, there seems to be no support for the statement that the size of $\|A - A_{ap}\|_{L^p(\mathbb{R}^+)}$ is bounded by σ_n for any $p \geqslant 1$. The aim of the Section 7 is to provide a slightly more intricate algorithm that does yield an estimate of $\|A - A_{ap}\|_{L^2(\mathbb{R}^+)}$ in terms of certain singular values. Before that, we shall return to the algorithm of Beylkin and Monzón.

6.1. Implementation of Algorithm 1 and the algorithm of Beylkin and Monzón revisited. Let a function $A \in C_0^1(\mathbb{R}^+)$ be given and let us assume that we have determined a suitable sampling interval $1/N$. Also assume that $\operatorname{supp} A = [0, 2]$ and that A is very small on $[1, 2]$. For easy visualization, suppose that $N = 3$ and that $\mathcal{S}_{2,2}A = (5, 4, 3, 2, 1, 0.1, 0.1, 0.1, 0.1, 0, 0 \ldots)$. There are four natural choices of Hankel matrices to use for the approximation algorithm.

$$\gamma_1 = \begin{pmatrix} 5 & 4 & 3 \\ 4 & 3 & 2 \\ 3 & 2 & 1 \end{pmatrix}, \quad \gamma_2 = \begin{pmatrix} 5 & 4 & 3 & 2 & 1 \\ 4 & 3 & 2 & 1 & 0.1 \\ 3 & 2 & 1 & 0.1 & 0.1 \\ 2 & 1 & 0.1 & 0.1 & 0.1 \\ 1 & 0.1 & 0.1 & 0.1 & 0.1 \end{pmatrix},$$

$$\gamma_3 = \begin{pmatrix} 5 & 4 & 3 & 2 & 1 \\ 4 & 3 & 2 & 1 & 0 \\ 3 & 2 & 1 & 0 & 0 \\ 2 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \end{pmatrix}, \quad \gamma_4 = \begin{pmatrix} 5 & 4 & 3 & 2 & 1 & 0.1 & 0.1 & 0.1 & 0.1 \\ 4 & 3 & 2 & 1 & 0.1 & 0.1 & 0.1 & 0.1 & 0 \\ 3 & 2 & 1 & 0.1 & 0.1 & 0.1 & 0.1 & 0 & 0 \\ 2 & 1 & 0.1 & 0.1 & 0.1 & 0.1 & 0 & 0 & 0 \\ 1 & 0.1 & 0.1 & 0.1 & 0.1 & 0 & 0 & 0 & 0 \\ 0.1 & 0.1 & 0.1 & 0.1 & 0 & 0 & 0 & 0 & 0 \\ 0.1 & 0.1 & 0.1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.1 & 0.1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

γ_1 and γ_2 obviously correspond to Beylkin and Monzón's algorithm, whereas the latter 2 correspond to the one presented here. In γ_1 and γ_3 we have chosen to only include the part where the function is big, whereas in γ_2 and γ_4 we have included all available information. γ_3 and γ_4 have the advantage that we can precisely predict the number of zeroes in the unit disc for a given singular vector, and hence decreasing the evaluation time of the algorithm. On the other hand, these matrices are bigger which has the opposite effect.

We let n be fixed. Denote the ζ_k's in $\mathbb{C}^+$ that are obtained from u_n for each of the matrices $\gamma_1, \gamma_2, \gamma_3, \gamma_4$ by $\{\zeta_k^1\}_{k=1}^{n_1}$, $\{\zeta_k^2\}_{k=1}^{n_2}$, $\{\zeta_k^3\}_{k=1}^{n}$, $\{\zeta_k^4\}_{k=1}^{n}$. If we consider the zeroes $\{\zeta_k^4\}$ to be optimal, in the light of Corollary 3.1, then our conclusion is actually that the optimal choice is to work with γ_2, at least if N is large. We base this on running comparative tests with all 4 matrices on a large number of randomly generated smooth functions, where we have observed that

- The "tails" in $[1, 2]$ are important; by using γ_1 the amount of zeroes n_1 varies heavily and their location does not seem to be related to the location of $\{\zeta_k^4\}_{k=1}^{n}$.
- It is very rare that $n_2 \neq n$ and the location of $\{\zeta_k^2\}_{k=1}^{n_2}$ is generally closer to that of $\{\zeta_k^4\}_{k=1}^{n}$ than $\{\zeta_k^3\}_{k=1}^{n}$. Thus γ_2 is preferred over γ_3.
- $\{\zeta_k^2\}_{k=1}^{n_2}$ generally coincides well with $\{\zeta_k^4\}_{k=1}^{n}$, and hence it is not worth the additional execution time that is necessary to work with γ_4.
- The values $\sigma_n^1, \sigma_n^2, \sigma_n^3, \sigma_n^4$ does not vary much, σ_n^1 is typically a factor 2 larger than σ_n^4, and the other 2 are somewhere between.

Thus, for implementation purposes, our recommendation is to actually use γ_2, (i.e. the algorithm by Beylkin and Monzón, keeping in mind that the "tail" of the function is important and with the adjustment in the definition of σ_n as well as in the interpretation of its relation with the error ϵ).

This explains the phenomena 1-3 observed by Beylkin and Monzón (Section 2), but only in the case when the function to be approximated is small in the right endpoint. This however was not a criteria in [3], but on the other hand in all of their examples with a function that is large at both endpoints, they actually first decompose it as a sum of two functions, each being small in one of the endpoints, and then apply their algorithm to each one separately. We thus find it likely that the algorithm primarily works for functions that are small in one endpoint.

7. Algorithm 2. Given a function $A \in C_0^1(\mathbb{R}^+)$, let $\sigma_0(r), \sigma_1(r), \dots$ be the singular values of $\Gamma_{r^{-1/2}A'+r^{1/2}A}$, where $r > 0$ is a free parameter. In the algorithm we are about to present, the error of the approximation with n terms will be bounded by $\sigma_n(r)$, and hence r should be chosen such that $\sigma_n(r)$ is small. The foundation of the new algorithm lies in the following theorem.

THEOREM 7.1. *Let $A \in C_0^1(\mathbb{R}^+)$ be given, let $r > 0$ be fixed and let $\sigma_0(r), \sigma_1(r), \dots$ be the singular values of $r^{1/2}\Gamma_A - r^{-1/2}\Gamma_{A'}$. Let U_n be the corresponding singular vectors. Fix n and assume that $\sigma_n(r)$ is distinct, let $\zeta_1, \dots, \zeta_n$ denote the zeroes of $\widecheck{u_n}$ and take $c_1, \dots, c_n \in \mathbb{C}$ such that*

$$A_{ap}(x) = \sum_{k=1}^{n} c_k e^{i\zeta_k x}$$

is the orthogonal projection of A onto $\operatorname{Span}\{e^{i\zeta_k x}\}$ *in $L^2(\mathbb{R}^+)$. Then*

$$\|A - A_{ap}\|_{L^2(\mathbb{R}^+)} \leqslant \frac{\sigma_n}{\sqrt{2}}.$$

Proof. In order to avoid too much repetition of the material in Section 5, we will only write down the proof for the special case $r = 1$, and in the end indicate how to modify the calculations to obtain the general statement.

Set $B = \sqrt{2}\hat{A}$ and $b = \mathcal{U}^{-1}B$. Note that although we use the same notation as in previous sections, we previously had $b = B \circ \alpha$ instead. By Proposition 5.6 we have that

$$\mathcal{F}^{-1}\mathcal{U}H_b\mathcal{U}^{-1}\mathcal{F}^{-1} = \mathcal{F}^{-1}H_{b\circ\beta}\mathcal{F}^{-1} = \Gamma_{\sqrt{2\pi}^{-1}\mathcal{F}^{-1}(b\circ\beta)}.$$

Now,

$$b \circ \beta = \frac{\sqrt{\pi}(\zeta + i)}{\sqrt{\pi}(\zeta + i)} b \circ \beta = \sqrt{\pi}(\zeta + i)\mathcal{U}(b) = \sqrt{\pi}(\zeta + i)B$$

and

$$\sqrt{2\pi}^{-1}\mathcal{F}^{-1}(\sqrt{\pi}(\zeta + i)B) = \mathcal{F}^{-1}((\zeta + i)\hat{A}) = -iA' + iA$$

so $\sigma_0(1), \sigma_1(1), \dots$ are also the singular values to H_b and moreover, $\widecheck{u_n} = \mathcal{U}^{-1}(\widecheck{u_n})$ are the corresponding singular vectors. Setting $z_j = \beta(\zeta_j)$ we also have that $\{z_j\}_{j=1}^n$ is the zero set of $\widecheck{u_n}$.

The function b itself might be unbounded but as H_b is bounded, we have at least that $b \in H^2_-(\mathbb{D})$. By Nehari's theorem and the AAK-theorem (Theorems 5.2 and 5.5) there are coefficients $\tilde{c}_1, \dots, \tilde{c}_n$ such that

$$\sigma_n(1) = \|H_b - \sum_{j=1}^{n} \tilde{c}_j H_{\overline{zk_{z_j}}}\| = \inf_{\phi \in H^2(\mathbb{D})} \|b - \sum_{j=1}^{n} \tilde{c}_j \overline{zk_{z_j}} - \phi\|_{L^\infty} \geqslant$$

$$\geqslant \inf_{\phi \in H^2(\mathbb{D})} \|P_{H^2_-}(b - \sum_{j=1}^{n} \tilde{c}_j \overline{zk_{z_j}} - \phi)\|_{L^2} = \|b - \sum_{j=1}^{n} \tilde{c}_j \overline{zk_{z_j}}\|_{L^2}.$$

By (5.11) and (5.13) we obtain $\mathcal{U}(\overline{zk_{z_j}}) = \overline{\mathcal{U}(k_{z_j})} = C\overline{k_{\zeta_j}}$, where C is an uninteresting constant. Note that $\mathcal{F}(\chi_{\mathbb{R}^+}e^{i\zeta_j x}) = \overline{k_{\zeta_j}}$ by (5.9) and standard formulas. As $\mathcal{U}$ is unitary, we get

$$\sigma_n(1) \geqslant \|\mathcal{U}(b - \sum_{j=1}^{n} \tilde{c}_j \overline{zk_{z_j}})\|_{L^2} = \|B - \sum_{j=1}^{n} C\tilde{c}_j \overline{k_{\zeta_j}}\|_{L^2} =$$

$$= \sqrt{2}\|\hat{A} - \sum_{j=1}^{n} \frac{C\tilde{c}_j}{\sqrt{2}} \overline{k_{\zeta_j}}\|_{L^2} = \sqrt{2}\|A - \sum_{j=1}^{n} \frac{C\tilde{c}_j}{\sqrt{2}} \chi_{\mathbb{R}^+}e^{i\zeta_j x}\|_{L^2}.$$

This finishes the proof for $r = 1$, because A_{ap} is the orthogonal projection onto $\mathsf{Span}\,\{\chi_{\mathbb{R}^+}e^{i\zeta_j x}\}$ in $L^2(\mathbb{R})$, so clearly

$$\|A - A_{ap}\| \leqslant \|A - \sum_{j=1}^{n} \frac{C\tilde{c}_j}{\sqrt{2}} \chi_{\mathbb{R}^+}e^{i\zeta_j x}\|.$$

The proof for the general case is identical but uses

$$\alpha_r : \mathbb{D} \to \mathbb{C}^+; \quad \alpha_r(z) = ri\frac{1+z}{1-z}$$

and the unitary map

$$\mathcal{U}_r : L^2(\mathbb{T}) \to L^2(\mathbb{R}); \quad \mathcal{U}_r(\phi) = \sqrt{\frac{r}{\pi}}\frac{\phi \circ \beta_r}{\zeta + ri}$$

instead. Subsequently all formulas that are derived from these have to be modified accordingly. We omit the details. □

Given a function $A \in C_0^1(\mathbb{R}^+)$ and a desired approximation error $\epsilon > 0$ in $L^2(R)$, the algorithm thus goes as follows;

1. Find a pair (n, r) with n as small as possible and $\sigma_n(r) < \sqrt{2}\epsilon$. Note that $\sigma_n(r) \sim \sqrt{r}$ as $r \to \infty$ and $\sigma_n(r) \sim 1/\sqrt{r}$ as $r \to 0$, so each curve $\sigma_n(\cdot)$ has a minimum for some finite value of r. To find an optimal value of (n, r) one thus needs to find these minima.
2. Once (n, r) has been chosen, Theorem 7.1 gives the existence of certain nodes $\zeta_1, \ldots, \zeta_n \in \mathbb{C}^+$. Approximate values $\tilde{\zeta}_1, \ldots, \tilde{\zeta}_n$ of these nodes can be efficiently calculated using the same methods as in the first algorithm.
3. Values $c_1, \ldots, c_n$ such that $A_{ap} = \sum_{j=1}^{n} c_j e^{i\zeta_j x}$ is the orthogonal projection onto $\mathsf{Span}\,\{e^{i\zeta_j x}\}$ are easily calculated using the least squares method. By Theorem 7.1 we have

$$\|A - A_{ap}\| \lesssim \frac{\sigma_n(r)}{\sqrt{2}} \leqslant \epsilon.$$

Remark 1. It is interesting to observe that Algorithm 1 appears as the limit of the above algorithm as $r \to \infty$. More precisely, let n be fixed, let $\sigma_n(r)$ be as above and denote the corresponding ζ_k's by $\zeta_k(r)$. Also let σ_n and ζ_k be defined as in Section 6. Using Proposition 4.1 and the methods in the proof of Theorem 6.1, it is easy to see that, (after reordering the terms)

$$\zeta_k(r) \to \zeta_k$$

and

$$\frac{\sigma_n(r)}{\sqrt{r}} \to \sigma_n$$

as $r \to \infty$.

Remark 2. A further way to generalize Algorithms 1 and 2 would clearly be to develop an AAK-theory for Hankel operators Γ_a on weighted spaces, as opposed to $l^2(\mathbb{Z}^+)$ as in the this paper. The only results in this direction are by S. Treil and A. Volberg [15], but their proofs are not constructive and hence not suitable for using as a basis of Algorithms of the type considered here. In [5] we develop a "constructive weighted AAK-theory".

REFERENCES

[1] Adamjan, V. M.; Arov, D. Z.; Krein, M. G. *Analytic properties of the Schmidt pairs of a Hankel operator and the generalized Schur-Takagi problem.* (Russian) Mat. Sb. (N.S.) 86(128) (1971), 34–75.

[2] Andersson, F.; Carlsson, M.; de Hoop, M *Nonlinear approximation of functions by sums of wave packets* To appear in J. Appl. Comput. Harmon. Anal.

[3] Beylkin, G.; Monzón, L. *On approximation of functions with exponential sums.* J. Appl. Comput. Harmon. Anal. 19 (2005), 17–48.

[4] Butz, J. R. *s-numbers of Hankel matrices.* J. Functional Analysis 15 (1974), 297–305.

[5] Carlsson, M. *AAK-theory on weighted spaces.* This issue.

[6] Conway, J. B. *A course in functional analysis.* Second edition. Graduate Texts in Mathematics, 96. Springer-Verlag, New York, 1990.

[7] Duren, P. L. *Theory of H^p spaces.* Pure and Applied Mathematics, Vol. 38 Academic Press, New York-London 1970.

[8] Hartman, P. *On completely continuous Hankel matrices.* Proc. Amer. Math. Soc. 9 1958 862–866.

[9] Horn, R. A.; Johnson, C. R. *Matrix Analysis* Cambrige University Press, Cambrige, 1990.

[10] Koosis, P. *Introduction to H^p spaces.* Second edition. Cambridge Tracts in Mathematics, 115. Cambridge University Press, Cambridge, 1998.

[11] Nehari, Z. *On bounded bilinear forms* Ann. Math., 65 (1957) 153-162.

[12] Nikolski, N. K. *Operators, functions, and systems: an easy reading. Vol. 1. Hardy, Hankel, and Toeplitz.* Mathematical Surveys and Monographs, 92. American Mathematical Society, Providence, RI, 2002.

[13] Peller, V. V. *Hankel operators and their applications.* Springer Monographs in Mathematics. Springer-Verlag, New York, 2003.

[14] Peller, V. V. *An excursion into the theory of Hankel operators.* Holomorphic spaces (Berkeley, CA, 1995), 65–120, Math. Sci. Res. Inst. Publ., 33, Cambridge Univ. Press, Cambridge, 1998.

[15] Treil, S.; Volberg. A. *A fixed point approach to Nehari's problem and its applications.* 165–186, Oper. Theory Adv. Appl., 71, Birkhäuser, Basel, 1994.