On the 2-condition number of infinite Hankel matrices of finite rank

Bazan, Fermin S.V.; Toint, Philippe

Published in:
Systems and Control Letters

Publication date:
2000

Document Version
Early version, also known as pre-print

Link to publication

Citation for pulished version (HARVARD):
Bazan, FSV & Toint, P 2000, 'On the 2-condition number of infinite Hankel matrices of finite rank', *Systems and Control Letters*, vol. 41, no. 5.

Conditioning of Infinite Hankel Matrices of Finite Rank

F. S. V. Bazán * Ph. L. Toint †

Revised May 14, 2000

Abstract

Let H be an infinite Hankel matrix with h_{i+j-2} as its (i,j)-entry, $h_k = \sum_{l=1}^{n} r_l\, z_l^k$, $k = 0, 1, \ldots,$ $|z_l| < 1$, and $r_l, z_l \in \mathbb{C}$. We derive upper bounds for the 2-condition number of H as functions of n, r_l and z_l, which show that the Hankel matrix H becomes well-conditioned whenever the z's are close to the unit circle but not extremely close to each other. Numerical results which illustrate the theory are provided.

Key words. Infinite Hankel matrices, singular values, condition number, exponential modeling.

1 Introduction

Let H be an infinite Hankel matrix whose (i,j)-entry is h_{i+j-2}, that is

$$H = \begin{bmatrix} h_0 & h_1 & h_2 & \cdots \\ h_1 & h_2 & h_3 & \cdots \\ h_2 & h_3 & \cdots & \cdots \\ \vdots & \vdots & \cdots & \vdots \end{bmatrix}, \tag{1.1}$$

where $\{h_k\}_{k=0}^{\infty}$ denotes a complex-valued sampled signal composed of n exponentials

$$h_k = \sum_{l=1}^{n} r_l\, z_l^k, \quad k = 0, 1, \ldots, \tag{1.2}$$

where r_l, z_l are complex constants and the z's are known as *modes or poles.* We assume $|z_l| < 1$ and $z_k \neq z_l$ for $k \neq l$. Then H has rank n and can be factorized as (see, e.g., Gragg and Reichell [10])

$$H = W\, R\, W^T, \tag{1.3}$$

*Departamento de Matemática, Universidade Federal de Santa Catarina, Florianópolis, Santa Catarina 88040-900, Brasil (fermin@mtm.ufsc.br). Part of this research was supported by CNPq, Brasil, grant 300487/94 - 0(NV).

†Department of Mathematics Facultés Universitaires ND de la Paix, 61 rue de Bruxelles, B-5000, Namur, Belgium (pht@math.fundp.ac.be)

where $R = \mathrm{diag}(r_1, r_2, \ldots, r_n)$ and W^T the infinite Vandermonde matrix

$$W^T = \begin{bmatrix} 1 & z_1 & z_1^2 & z_1^3 & \cdots \\ 1 & z_2 & z_2^2 & z_2^3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & z_n & z_n^2 & z_n^3 & \cdots \end{bmatrix}_{n\times\infty} . \tag{1.4}$$

Let ℓ^2 be the Hilbert space of infinite complex column vectors with finite 2-norm. Then H gives rise to a bounded linear operator of finite rank on ℓ^2 and its Moore-Penrose pseudo inverse can be defined as the unique bounded operator $H^\dagger$ satisfying the conditions

$$HH^\dagger H = H, \;\; H^\dagger H H^\dagger = H^\dagger, \;\; (HH^\dagger)^* = HH^\dagger, \;\; (H^\dagger H)^* = H^\dagger H. \tag{1.5}$$

The star symbol stands for the adjoint of the operator. The singular values of H, which we denote by $\sigma_i(H)$, are the square roots of the eigenvalues of H^*H. They satisfy

$$\sigma_1(H) \geq \sigma_2(H) \geq \cdots \geq \sigma_n(H) > 0 = \sigma_{n+1}(H) = \sigma_{n+2}(H) = \cdots, \tag{1.6}$$

$\sigma_1(H) = \|H\|_2$, and $\sigma_n(H) = \|H^\dagger\|_2^{-1}$, where $\|\cdot\|$ stands for the operator norm. Similarly, the infinite Vandermonde matrix W gives rise to another bounded linear operator of rank n, from $\mathbb{C}^n$ in ℓ^2, and its pseudo inverse $W^\dagger$ also can be defined via the above conditions. Taking this observation into account we verify using (1.5) for W, that

$$W^\dagger = (W^*W)^{-1}W^*. \tag{1.7}$$

From this relation it follows that $W^{T^\dagger} R^{-1} W^\dagger$ satisfies (1.5), which ensures that

$$H^\dagger = W^{T^\dagger} R^{-1} W^\dagger. \tag{1.8}$$

All the above properties of pseudo-inverses are consequences of the well-developed Hilbert space theory for pseudo-inverses of bounded linear operators with closed range, which clearly also hold for matrices in $\mathbb{C}^{M\times N}$. The reader is referred to Ben-Israel and Greville [4] for details about pseudo-inverses and, in particular, to Theorems 2 and 3 of Chapter 8 therein.

The 2-condition number of H, $k_2(H)$, is defined as the ratio of the largest to the smallest non zero singular value of H, that is $k_2(H) = \sigma_1(H)/\sigma_n(H)$, and it is of interest in areas such as signal processing, where Hankel matrices of growing dimension are often used [5, 6, 13, 14, 7].

Our goal in this paper is to analyze under which conditions over n, z_l, and r_l, the Hankel matrix H becomes well-conditioned. To achieve our goal, we first modify (1.3) as

$$H = UU^T, \quad U = WR^{1/2}, \tag{1.9}$$

where $R^{1/2}$ is an arbitrarily chosen square root of R, and then we obtain that

$$k_2(H) \leq \|U\|_2^2 \|U^\dagger\|_2^2 = [k_2(U)]^2. \tag{1.10}$$

This relation enables us to derive bounds for $k_2(H)$ by bounding $k_2(U)$. Notice that inequality (1.10) holds as an equality when the modes z_l and the weights r_l are positive real numbers. Of course, using (1.9) it is easy to see that, discarding zero eigenvalues, the spectrum of H^*H, $\lambda(H^*H)$, satisfies

$$\lambda(H^*H) = \lambda(G\overline{G}), \quad G = U^*U, \tag{1.11}$$

where the bar stands for complex conjugation. Thus if both z_l and r_l are positive real numbers, then $\lambda(H^*H) = \lambda(G^2)$, in which case the inequality in (1.10) becomes an equality. The positive singular values $\sigma_i(H)$ $(i = 1 : n)$ can be computed from the $n \times n$ eigenvalue problem related to $G\overline{G}$. However, we stress that our main goal is not to compute the singular values $\sigma_i(H)$ but rather to derive informative bounds on $k_2(H)$. Thus all our conclusions about $k_2(H)$ shall arise from analyzing their bounds. We state our results by slightly modifying an analysis by Bazán [1] who provided an upper bound for $k_2(W)$. As a by product, we obtain a bound on $k_2(U)$ which improves that of reference [1], when all weights are equal to one. Further, we derive an upper bound on $k_2(H)$ in terms of n, r_l and z_l, whose quality essentially depends on the separation of the z_l inside the unit circle. In particular, we show that if the z's are close to the unit circle but not extremely close to each other, then the related Hankel matrix becomes well-conditioned, provided n is not very large.

The paper is organized as follows. In Section 2 we consider finite sections of the Hankel matrix as Krylov matrices and some basic results arising from this identification are described. Our upper bounds for $k_2(H)$ are presented in Section 3. Because the smallest non zero singular value of finite sections of H plays the role of a *threshold value* for separating signal from noise, in several signal processing applications, an application of our results illustrating that role is presented in Section 4. We discuss the choice of the dimension of the finite Hankel matrix that guarantees a satisfying separation of signal from noise. This is numerically illustrated in Section 5. Section 6 finally presents some conclusions.

2 Basic results

In what follows the singular values of a matrix A are denoted by $\sigma_i(A)$ and are arranged so that $\sigma_1(A) \geq \sigma_2(A) \geq \cdots \geq \sigma_n(A)$. Also, as usual, $\|A\|_2$ and $\|A\|_F$ denote the spectral and the Frobenius norm of A, respectively. Leading $M \times N$ principal submatrices of the Hankel matrix H play a important role in our developments, and are denoted by $H_{M\times N}$; when $M = N$ they are simply denoted by H_N. A first consequence of the matrix Hankel structure is that if $N \geq n$, then H_N inherits its rank from H and also its Vandermonde decomposition (1.3), that is, $\text{rank}(H_N) = n$ and

$$H_N = W_N\, R\, W_N^T, \tag{2.1}$$

where W_N is the matrix consisting of the first N rows of W, and R is as in (1.3). It is well known that the entries of H satisfy a recurrence relation of order n of the form

$$h_k = f_{n-1}h_{k-1} + f_{n-2}h_{k-2} + \cdots + f_0 h_{k-n}, \quad k = n, n+1, \cdots \tag{2.2}$$

which generates the entire signal once the set of n initial values $\{h_0, h_1, \ldots, h_{n-1}\}$ are given (see Gantmacher [8], vol. 2, p. 207). Furthermore, the modes z_l generating the entries in 1.2 are the roots of the polynomial

$$p_n(z) = z^n - f_{n-1}z^{n-1} - \cdots - f_1 z - f_0. \tag{2.3}$$

The coefficients f_i are uniquely determined from the recurrence relation and are referred to as *predictor parameters*. Recurrence relations of type (2.2) of order $N > n$ are also possible. In this case however, the predictor parameters are not uniquely determined since they are computed from the rank-deficient underdetermined linear system

$$H_N\, f \doteq [\mathbf{h}_0\, \mathbf{h}_1\ \cdots\ \mathbf{h}_{N-1}] f = \mathbf{h}_N, \tag{2.4}$$

where

$$\mathbf{h}_i \doteq [h_i\, h_{i+1}\ \cdots\ h_{i+N-1}]^T,\ \ i = 0, \ldots, N\ \text{ and }\ f = [f_0\, f_1\ \cdots\ f_{N-1}]^T.$$

Despite this, the modes z_l still can be extracted from the roots of any polynomial $p_N(z)$ whose coefficients satisfy (2.4) [2, 15]. Let $\mathcal{C}$ denote the companion matrix corresponding to the polynomial

$$P_N(z) = z^N - f_{N-1}z^{N-1} - \cdots - f_1 z - f_0,$$

whose coefficients are the components of the minimum 2-norm solution of (2.4), i.e.

$$\mathcal{C} = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & 0 & 1 \\ f_0 & f_1 & \cdots & \cdots & f_{N-1} \end{bmatrix}. \tag{2.5}$$

The fact that the signal modes can be extracted from the zeros of $P_N(z)$ implies that there are n eigenvalues of $\mathcal{C}$ exactly coinciding with the n modes z_l. More precisely, it can be proved that

$$\mathcal{C}W_N = W_N Z, \tag{2.6}$$

where $Z = \operatorname{diag}(z_1, z_2, \ldots, z_n)$, which means that W_N is a matrix of right eigenvectors of C with associated eigenvalues z_l. We may now rephrase the property that the coefficients f_i are sufficient to predict future values of the signal (see (2.2) in terms of the matrix C: given two successive columns vectors of H_N, $\mathbf{h}_i$ and $\mathbf{h}_{i+1}$, then

$$\mathbf{h}_{i+1} = \mathcal{C}\,\mathbf{h}_i,\ i \geq 0.$$

From this observation, it follows that the Hankel matrix H_N can be regarded as a Krylov matrix generated by C, that is

$$H_N = [\mathbf{h}_0\, \mathbf{h}_1\ \cdots\ \mathbf{h}_{N-1}] = [\mathbf{h}_0\ \ \mathcal{C}\mathbf{h}_0\ \ \cdots\ \mathcal{C}^{N-1}\mathbf{h}_0],$$

and that, since $\operatorname{rank}(H_N) = n$ for $N \geq n$, the associated Krylov subspace, which we denote by $\mathcal{H}_N$, is *invariant* under $\mathcal{C}$. Let V be an $N \times n$ matrix with orthonormal columns spanning $\mathcal{H}_N$ and consider $\mathcal{C}_{\mathcal{P}}$ to be the $n \times n$ matrix defined by

$$\mathcal{C}_{\mathcal{P}} = V^*\mathcal{C}V. \tag{2.7}$$

The bounds on $k_2(H)$ that we shall derive crucially depend on the eigenvalue and singular value spectra of $\mathcal{C}_{\mathcal{P}}$. In what follows we briefly describe a result early obtained by Bazán [1] that characterizes that spectra. Note that, since the columns of both V and W_N span $\mathcal{H}_N$, 2.6 ensures that Z is an eigenvalue matrix of $C_{\mathcal{P}}$ with V^*W_N as right eigenvector matrix, i.e.

$$\mathcal{C}_{\mathcal{P}} = (V^*W_N)Z(V^*W_N)^{-1}. \tag{2.8}$$

On the other, it can be proved that

$$\mathcal{C}_{\mathcal{P}}^*\mathcal{C}_{\mathcal{P}} = V^*\mathcal{C}^*\mathcal{C}V = I + xx^* - yy^*,$$

where I denotes the identity matrix of order n, $x = V^*f^{T*}$ and $y = V^*e_1$. From this it is not difficult to check that the singular values of $\mathcal{C}_{\mathcal{P}}$ verify (see Theorem 3 in [1])

$$\begin{aligned}
\sigma_1^2(\mathcal{C}_{\mathcal{P}}) &= \frac{2 + \|f\|_2^2 - \|p_1\|_2^2 + \sqrt{(\|f\|_2^2 + \|p_1\|_2^2)^2 - 4|f_0|^2}}{2}, \\
\sigma_i^2(\mathcal{C}_{\mathcal{P}}) &= 1, \;\; i = 2, n-1, \\
\sigma_n^2(\mathcal{C}_{\mathcal{P}}) &= \frac{2 + \|f\|_2^2 - \|p_1\|_2^2 - \sqrt{(\|f\|_2^2 + \|p_1\|_2^2)^2 - 4|f_0|^2}}{2},
\end{aligned} \tag{2.9}$$

where p_1 is the first column of the orthogonal projector onto $\mathcal{H}_N$. Hence it follows that

$$1 \leq \sigma_1^2(\mathcal{C}_{\mathcal{P}}) \leq 1 + \|f\|_2^2, \tag{2.10}$$

and thus $\sigma_1^2(\mathcal{C}_{\mathcal{P}}) \to 1$ as $N \to \infty$ since $\|f\|_2 \to 0$ as $N \to \infty$ (see [3] again). On the other hand, since $\det(\mathcal{C}_{\mathcal{P}}^*\mathcal{C}_{\mathcal{P}}) = [\det(\mathcal{C}_{\mathcal{P}})]^2 = \prod_{l=1}^n |z_l|^2$, it follows that

$$\sigma_1^2(\mathcal{C}_{\mathcal{P}})\sigma_n^2(\mathcal{C}_{\mathcal{P}}) = \prod_{l=1}^n |z_l|^2. \tag{2.11}$$

This relation implies that $\sigma_n^2(\mathcal{C}_{\mathcal{P}}) \to \prod_{l=1}^n |z_l|^2$ as $N \to \infty$. All above results are summarized in the following theorem.

Theorem 1 *Suppose $\mathcal{C}_{\mathcal{P}}$ is the matrix defined in (2.7). Then it admits a spectral decomposition given by (2.8), and its singular values $\sigma_i(\mathcal{C}_{\mathcal{P}})$ satisfy (2.9). Furthermore*

$$\lim_{N\to\infty} \sigma_1(C_{\mathcal{P}}) = 1, \quad \text{and} \quad \lim_{N\to\infty} \sigma_n(C_{\mathcal{P}}) = \prod_{i=1}^n |z_i|. \tag{2.12}$$

3 Upper bounds for $k_2(H)$

We start by deriving bound $k_2(U)$, where U is the scaled Vandermonde matrix defined in (1.9). To this end, we first analyze the finite dimensional case for $k_2(U_N)$, where $U_N = W_N R^{1/2}$, with W_N defined in (2.1) and R as in (1.9).

Theorem 2 *Let p and q be indices between 1 and n, such that $\|U_N e_i\|$, $i = 1, \ldots, n$, is maximum for $i = p$ and minimum for $i = q$, where U_N is as above. Define $\alpha = |z_p|$, $\rho = |r_p|$, $\beta = |z_q|$, $\gamma = |r_q|$,* and $\mathrm{k_R} = \rho/\gamma$. *Also define,*

$$\delta = \min_{\substack{1 \le i,j \le n \\ i \ne j}} |z_i - z_j|, \tag{3.1}$$

$$\phi_N = \sqrt{\frac{1 + \alpha^2 + \cdots + \alpha^{2(N-1)}}{1 + \beta^2 + \cdots + \beta^{2(N-1)}}}, \tag{3.2}$$

and

$$D_N^2 = (\sigma_1^2(\mathcal{C}_\mathcal{P}) + \cdots + \sigma_n^2(\mathcal{C}_\mathcal{P})) - (|z_1|^2 + \cdots + |z_n|^2), \tag{3.3}$$

with $\sigma_i(\mathcal{C}_\mathcal{P})$ as in (2.9). Then, for all $N \ge n \ge 2$, the 2-condition number of U_N, $k_2(U_N) = \|U_N\|_2 \|U_N^\dagger\|_2$, satisfies

$$k_2(U_N) \le \frac{1}{2}\left(\eta + \sqrt{\eta^2 - 4}\right), \tag{3.4}$$

where

$$\eta = \sqrt{k_R}\left[1 + \frac{D_N^2}{(n-1)\delta^2}\right]^{\frac{n-1}{2}} \frac{n}{2}\left(\phi_N + \frac{1}{k_R}\phi_N^{-1}\right) - n + 2. \tag{3.5}$$

Proof: Our proof relies on the crucial observation that the conditioning of the eigenvalue problem related to matrix $\mathcal{C}_\mathcal{P}$ (see Theorem 1) is essentially governed by $k_2(U_N)$. Let $u_i = V^* U_N^{\dagger *} e_i$ and $v_i = V^* U_N e_i$ where e_i is the i-th canonical basis vector in $\mathbb{C}^n$. It then follows that they are left and right eigenvectors of $\mathcal{C}_\mathcal{P}$ corresponding to the eigenvalue z_i, respectively, and satisfy $u_i^* v_i = 1$. It is easy to see that $\|v_i\|_2^2 = \|U_N e_i\|_2^2$, and $\|u_i\|_2^2 = \|U_N^{\dagger *} e_i\|_2^2$, because VV^* is the orthogonal projector onto $\mathcal{H}_N$ and the columns of U_N span this subspace. This yields

$$\|U_N\|_F^2 = \sum_{i=1}^{n} \|v_i\|_2^2 \quad \text{and} \quad \|U_N^\dagger\|_F^2 = \sum_{i=1}^{n} \|u_i\|_2^2.$$

Using these observations, we obtain from Theorem 5 by Smith [12], that

$$\|u_i\|_2 = \frac{1}{|s_i|}\frac{1}{\|v_i\|_2} \le \left[1 + \frac{D_N^2}{(n-1)\delta_i^2}\right]^{\frac{n-1}{2}} \frac{1}{\|v_i\|_2}, \quad i = 1, 2, \ldots, n,$$

with $\delta_i = \min_{1 \le j \le n} |z_i - z_j|$, $i \ne j$, and where $|s_i|^{-1} = \|u_i\|\|v_i\|$ is the condition number of the eigenvalue z_i (see [16], page 69). From this, it follows that

$$\|U_N^\dagger\|_F^2 \le \sum_{i=1}^{n}\left[1 + \frac{D_N^2}{(n-1)\delta_i^2}\right]^{n-1} \frac{1}{\|v_i\|_2^2} \le \left[1 + \frac{D_N^2}{(n-1)\delta^2}\right]^{n-1} \sum_{i=1}^{n} \frac{1}{\|v_i\|_2^2}, \tag{3.6}$$

where δ is defined by (3.1). This implies that

$$\|U_N^\dagger\|_F^2 \|U_N\|_F^2 \le \left[1 + \frac{D_N^2}{(n-1)\delta^2}\right]^{n-1} \sum_{i=1}^{n} \|v_i\|_2^2 \sum_{i=1}^{n} \frac{1}{\|v_i\|_2^2}. \tag{3.7}$$

Now note that, if we define $Q = \text{diag}(\|v_1\|_2^2, \|v_2\|_2^2, \ldots, \|v_n\|^2)$, the product of the two sums in the right-hand-side of this inequality may be rewritten as

$$\sum_{i=1}^{n} \|v_i\|_2^2 \sum_{i=1}^{n} \frac{1}{\|v_i\|_2^2} = e^* Q e \, e^* Q^{-1} e, \quad e = (1, \ldots, 1)^T \in \mathbb{R}^n, \tag{3.8}$$

and that this quantity may be bounded from above by using Kantorovic's inequality (Horn and Johnson [11], pag. 444). This implies that

$$e^* Q e \, e^* Q^{-1} e \leq \left[\frac{n}{2} \left(\frac{v_{max}}{v_{min}} + \frac{v_{min}}{v_{max}} \right) \right]^2,$$

where $v_{max} = \max \|v_i\|_2$ and $v_{min} = \min \|v_i\|_2$. Using the previous inequality and the notations introduced in the theorem, and since

$$\|v_i\|_2^2 = \|W_N R^{1/2} e_i\|_2^2 = |r_i|(1 + |z_i|^2 + \cdots + |z_i|^{2(N-1)}),$$

inequality (3.7) gives

$$k_F(U_N) \leq \sqrt{k_R} \left[1 + \frac{D_N^2}{(n-1)\delta^2} \right]^{\frac{n-1}{2}} \frac{n}{2} (\phi_N + \frac{1}{k_R} \phi_N^{-1}). \tag{3.9}$$

We now recall a result about Jordan condition numbers (see [12], Theorem 1) which states that, if $A = X \Lambda X^{-1}$ is a spectral decomposition of $A \in \mathbb{C}^{n \times n}$ and all eigenvalues of A are simple, then

$$n - 2 + k_2(X) + [k_2(X)]^{-1} \leq k_F(X). \tag{3.10}$$

This inequality continues to hold if we substitute X by U_N and can be seen as follows. First notice that as we can always write $\mathcal{C}_{\mathcal{P}} = (V^* U_N) Z (V^* U_N)^{-1}$ (see (2.8)), then (3.10) holds for $X = V^* U_N$. Now given that $X^* X = U_N^* V V^* U_N = U_N^* U_N$, since the columns of U_N span $\mathcal{H}_N$, it follows that X and U_N have the same singular values and therefore $k_2(X) = k_2(U_N)$, as desired.

Solving inequality (3.10) for $k_2(X) = k_2(U_N)$ we obtain

$$k_2(U_N) \leq \frac{1}{2} \left[k_F(U_N) - n + 2 + \sqrt{(k_F(U_N) - n + 2)^2 - 4} \right].$$

The proof concludes by substituting (3.9) in this inequality. □

An immediate consequence of the above analysis is that if $R = I$, bound (3.4) becomes a bound on $k(W_N)$, which improves one derived by Bazán [1]. Despite this, it is worth noting both bounds strongly depend on D_N^2 and δ: small values of the bounds are ensured only when $D_N^2 \ll (n-1)\delta^2$. While δ^2 measures the separation of the modes z_l inside the unit circle, D_N^2, known as the *departure from normality*, measures how near is $\mathcal{C}_{\mathcal{P}}$ from being a normal matrix (see, for instance, [9] pag. 314). Numerical examples showing the dependence of the bound on those quantities are presented and discussed in Section 5. Another consequence is given in the following corollary.

Corollary 1 *Suppose $N \geq n$. Then*

$$k_2(H_N) \leq \frac{1}{4}\left(\eta + \sqrt{\eta^2 - 4}\right)^2, \tag{3.11}$$

where η is as above.

Proof: It suffices noting that $k_2(H_N) \leq [k_2(U_N)]^2$, which holds by (2.1), and then applying Theorem 2 in this inequality.

To derive our bounds for $k_2(H)$ we need the following auxiliary result.

Lemma 1 *Let U_N be the $N \times n$ scaled Vandermonde matrix as above. Define two sequences of real numbers $\{a_N\}$, $\{b_N\}$, $N \geq n$, by $a_N = \|U_N^\dagger\|_2$ and $b_N = k_2(U_N) = \|U_N\|_2 \|U_N^\dagger\|_2$. Then,*

(a) a_N decreases monotonically with N;

(b) $\lim_{N\to\infty} b_N = k_2(U)$.

Proof: The proof of part (a) is immediate and we shall only prove that (b). In fact, let $\{\widehat{U}_N\}$ be the sequence of infinite matrices with zeros everywhere except in the first N rows whose entries coincide with those of U_N, i.e,

$$\widehat{U}_N = \begin{bmatrix} U_N \\ 0 \\ \vdots \end{bmatrix}.$$

Obviously, $\|\widehat{U}_N\|_2 = \|U_N\|_2$ and $\|\widehat{U}_N^\dagger\|_2 = \|U_N^\dagger\|_2$. Next, partition U as

$$U = \begin{bmatrix} U_N \\ C \end{bmatrix} = \widehat{U}_N + \begin{bmatrix} 0 \\ C \end{bmatrix}, \tag{3.12}$$

where $C = [Z^N e \;\; Z^{N+1} e \;\; \cdots\,]^T R^{1/2}$ with e as in (3.8), and note that $U^\dagger = (U^*U)^{-1}U^*$ can be rewritten as

$$U_N^* U_N U^\dagger + C^* C U^\dagger = [U_N^* \;\; 0 \cdots] + [0 \;\; C^*].$$

This can be rewritten again as

$$U^\dagger + (U_N U_N^*)^{-1}(C^*C)U^\dagger = \widehat{U}_N^\dagger + (U_N U_N^*)^{-1}\;[0 \;\; C^*],$$

which, using the fact that $U_N^\dagger U_N^{\dagger *} = (U_N^* U_N)^{-1}$, yields

$$U^\dagger - \widehat{U}_N^\dagger = -U_N^\dagger U_N^{\dagger *}(C^*C)U^\dagger + U_N^\dagger U_N^{\dagger *}\;[0 \;\; C^*].$$

Hence, taking into account that both $\|U^\dagger\|_2$ and $\|U_N^\dagger\|_2$ are bounded ($U^\dagger$ is of finite rank and $\|U_N^\dagger\|_2$ decreases by part (a)), that $\|C^*C\|_2 = \|C\|_2^2$ and that

$$\|C\|_2 \leq \sqrt{n}\,\sqrt{\max|r_l|}(\max|z_l|)^N \|U\|_2 \to 0$$

as $N \to \infty$, because $\max |z_l| < 1$, we deduce that

$$\|U^\dagger - \widehat{U}_N^\dagger\|_2 \le \|U^\dagger\|_2 \|C\|_2^2 \|U_N^\dagger\|_2^2 + \|C\|_2 \|U_N^\dagger\|_2^2 \;\to\; 0$$

when $N \to \infty$, thus implying that $\widehat{U}_N^\dagger \to U^\dagger$. Hence $\|\widehat{U}_N^\dagger\|_2$ converges to $\|U^\dagger\|_2$. Now since $\|\widehat{U}_N\|_2 \to \|U\|_2$, we have that

$$b_N = \|U_N\|_2 \|U_N^\dagger\|_2 = \|\widehat{U}_N\|_2 \|\widehat{U}_N^\dagger\|_2$$

converges to $k_2(U)$, as requested. $\square$

Theorem 3 *Let H be the infinite Hankel matrix introduced in (1.1). Let α, β, ρ, γ, k_R and δ be as in Theorem 2. Then the 2-condition number of H satisfies*

$$k_2(H) \le \frac{1}{4}\left(\widehat{\eta} + \sqrt{\widehat{\eta}^2 - 4}\right)^2, \tag{3.13}$$

where

$$\widehat{\eta} = \sqrt{k_R}\left[1 + \frac{n - 1 + \prod_{i=1}^n |z_i|^2 - \sum_{i=1}^n |z_i|^2}{(n-1)\delta^2}\right]^{\frac{n-1}{2}} \frac{n}{2}\left[\sqrt{\frac{1-\beta^2}{1-\alpha^2}} + \frac{1}{k_R}\sqrt{\frac{1-\alpha^2}{1-\beta^2}}\right] - n + 2. \tag{3.14}$$

Proof: Since by (1.10) and Lemma 1,

$$k_2(H) \le [k_2(U)]^2 = \lim_{N\to\infty} [k_2(U_N)]^2, \tag{3.15}$$

it is sufficient to take limit in (3.5) when N tends to infinity. To compute this limit, note that

$$D_\infty = \lim_{N\to\infty} D_N^2 = n - 1 + \prod_{i=1}^n |z_i|^2 - \sum_{i=1}^n |z_i|^2, \quad \text{and} \quad \lim_{N\to\infty} \phi_N = \sqrt{\frac{1-\beta^2}{1-\alpha^2}}, \tag{3.16}$$

the first because of (3.3) and (2.12), and the second because of (3.2) and the fact that α and β are both smaller than one. The desired result is then obtained by substituting (3.16) into (3.15). $\square$

Theorem 3 shows that the quality of the upper bound depends on the closeness of $|z_l|$ to one and on the separation of the modes themselves inside the unit circle.

Corollary 2 *Define $\widehat{\alpha} = \max |z_l|$, $\widehat{\beta} = \min |z_l|$. Assume the modes z_l satisfy $1 - \widehat{\beta}^2 \le \delta^2$, $\widehat{\alpha}^n \le \widehat{\beta}$. Then, for $n \ge 2$*

$$k_2(H) \le \left[\sqrt{k_R}\, 2^{\frac{n-3}{2}} n \left(\sqrt{\frac{1-\beta^2}{1-\alpha^2}} + \frac{1}{k_R}\sqrt{\frac{1-\alpha^2}{1-\beta^2}}\right) - n + 2\right]^2. \tag{3.17}$$

Proof: The proof is simples and follows from noting that $k_2(H) \leq \widehat{\eta}^2$. $\square$

Bound (3.17) is no better than (3.13) and may overestimate $k_2(H)$ for n large. Despite this, it gives relevant information regarding the dependence of the bound on the distribution of the modes and their separations. The smaller the separations, the closer to the unit circle the modes must lie in order to obtain moderate values for $k_2(H)$. Consequently, for well-separated modes satisfying $|z_l| \approx 1$, we deduce that H should be well-conditioned, provided n is not very large. The assumption $\widehat{\alpha} \approx \widehat{\beta} \approx 1$ occurs in practical applications involving slightly damped signals.

Remark

Others bounds on $k_2(H)$ can be obtained by combining the inequality

$$k_2(H) \leq k_2(R)[k_2(W)]^2, \tag{3.18}$$

which follows from (1.3), with bounds on $k_2(W)$. The quality of the bounds so derived, however, will depend on the sharpness of the bounds on $k_2(W)$. A bound resulting from this procedure is that obtained by using Bazań's bound for $k_2(W)$ [1]. The obtained bound, however, should not improve (3.13), as $k_2(H) \leq [k_2(U)]^2 \leq k_2(R)[k_2(W)]^2$. A numerical comparison illustrating this fact is presented in Section 5.

4 A Signal Processing Application

In several signal processing applications one is interested in retrieving parameters such as frequencies, plane waves, dampings, etc, from a finite set of perturbed data. This data arises as $\tilde{h}_k = h_k + \epsilon_k$, $k = 0, 1, \ldots, L-1$, where h_k is an unknown sampled signal of the form (1.2), where $z_l = e^{(d_l + i\omega_l)\Delta t}$, $i = \sqrt{-1}$, $d_l < 0$, $\omega_l \in \mathbb{R}$, Δt is the sampling interval, and ϵ_k the noise. The aim is to compute, as accurately as possible, estimates of r_l, d_l and ω_l, even if the data is relatively noisy. We refer the reader to Van Huffel [14] for a variety of applications where this problem is relevant.

Most of methods to the problem start by filling the available data in a Hankel matrix $H_{M\times N}$, and can be separated into two large groups: methods that extract the parameters from the roots of *large* polynomials, as described in Section 2, and methods based on estimates of the so-called *signal subspace* (the column or row space of $H_{M\times N}$). Crucial for these approaches is the detection of n (the rank of H) and the estimation of the chosen signal subspace, both informations being extracted from a full rank Hankel matrix: $\widetilde{H}_{M\times N} = H_{M\times N} + E$, where E contains the noise. In practice, this is accomplished by a heuristic criterion for looking for a break in the pattern of singular values of $\widetilde{H}_{M\times N}$, attributing the larger ones to the signal and the smaller ones to the noise. Similarly, the subspace spanned by the singular vectors associated with the set of large singular values is used as an estimate of the signal subspace. For a discussion concerning conditions on the noise matrix E that allow to recover the row signal space from the SVD of $\tilde{H}_{M\times N}$, see De Moor [6].

We are now interested in discussing the best choice of M and N for the purpose of separating the n signal singular values from those associated with the noise, in the situation where the data is filled in a Hankel matrix $\widetilde{H}_{M\times N}$ such that $M + N = L + 1$, where L is fixed, $M, N \geq n$, and E is treated simply algebraically, i.e., with no assumption on the nature of the noise. As singular value theory ensures that this separation is best carried out when $\|E\|_2 \ll \sigma_n(H_{M\times N})$, we could focus on an analysis of the pair (M, N) that maximizes $\sigma_n(H_{M\times N})$ as a function of the dimensions. This maximization is also of interest because bounds on the quality of the approximate signal subspace estimated from noisy measurements typically depends on expressions of the type $\|E\|_2/\sigma_n(H_{M\times N})$ (see, e.g., Theorem 2.1 in [7]). However, given the difficulty of the problem, we restrict ourselves to discussing the choice of M, N that maximizes a lower bound on $\sigma_n(H_{M\times N})$ instead. We start by noting that, since $H_{M\times N} = U_M U_N^T$, it follows that $H_{M\times N}^\dagger = {U_N^T}^\dagger U_M^\dagger$, and thus

$$\sigma_n(H_{M\times N}) = \|H_{M\times N}^\dagger\|_2{}^{-1} \geq \|U_N^\dagger\|_2^{-1}\|U_M^\dagger\|_2^{-1} = \sigma_n(U_N)\sigma_n(U_M).$$

Imposing the constraint that $M + N = L + 1$, this inequality becomes

$$\sigma_n(H_{M\times N}) \geq \sigma_n(U_M)\,\sigma_n(U_{L+1-M}) \;\text{ for }\; \mathrm{n \leq M \leq L + 1 - n}. \tag{4.1}$$

Observe next that for $M \approx L + 1 - n$ (which implies $N \approx n$), this bound is small, since $\sigma_n(U_{L+1-M}) \approx 0$ as U_{L+1-M} is almost a scaled square Vandermonde matrix, which is generally ill-conditioned. This is in contrast with the fact that $\sigma_n(U_M)$ increases with M (ensured by Lemma 1). This balancing effect between M and N suggests that the Hankel matrix should not be chosen too overdetermined. By symmetry, the same reasoning applies in the case $M \approx n$ (which implies $N \approx L + 1 - n$), i.e, $H_{M\times N}$ should not be chosen too underdetermined. Hence, if the aim is to maximize the bound 4.1, then the Hankel matrix should not be chosen *too rectangular*. The following theorem states sufficient conditions that enable us to choose M, N in order to maximize that bound.

Theorem 4 *Let $H_{M\times N}$ be the leading submatrix of the infinite Hankel matrix H, with $M + N = L + 1 = 2T$, and L is a given odd integer. Assume*

$$\sigma_n(U_j) - 2\sigma_n(U_T) + \sigma_n(U_{2T-j}) \leq 0, \;\; j = n, \ldots, 2T - n. \tag{4.2}$$

Then the bound (4.1) on $\sigma_n(H_{M\times N})$ is maximized when $M = N = T$. Furthermore,

$$\sigma_n(H_M) \geq \left[1 + \frac{D_M^2}{(n-1)\delta^2}\right]^{1-n} \frac{\gamma}{n}\frac{1 - e^{2d\Delta t M}}{1 - e^{2d\Delta t}}, \tag{4.3}$$

where $d = d_q$, $\gamma = |r_q|$, with q an integer chosen so that $\|U_N e_i\|$ $(i = 1, \ldots, n)$ in (3.6), is minimum for $l = q$.

Proof: Rewriting (4.2) as $\sigma_n(U_T) \geq \frac{1}{2}\left[\sigma_n(U_j) + \sigma_n(U_{2T-j})\right]$, $j = n, \ldots, 2T - n$, we obtain

$$\sigma_n^2(U_T) \geq \frac{1}{4}[\sigma_n^2(U_j) + 2\sigma_n(U_j)\sigma_n(U_{2T-j}) + \sigma_n^2(U_{2T-j})], \; j = n, \ldots, 2T - n. \tag{4.4}$$

Also, since $\sigma_n(U_{2T-j}) - \sigma_n(U_j) \geq 0$, it is clear that $\sigma_n^2(U_j) + \sigma_n^2(U_{2T-j}) \geq 2\sigma_n(U_j)\sigma_n(U_{2T-j})$. Substituting this inequality into (4.4) we obtain

$$\sigma_n^2(U_T) \geq \sigma_n(U_j)\sigma_n(U_{2T-j}), \;\; j = n, \ldots, 2T - n,$$

which shows that the bound is maximized when $M = N = (L+1)/2$, as claimed.

Estimate (4.3) is an immediate consequence of (3.6) where we use the well-known property $\|U_M^\dagger\|_2 \leq \|U_M^\dagger\|_F$. $\square$

The sense of condition (4.2) is that $\sigma_n(U_j)$ is required to increase rapidly initially (i.e. for $n \leq j \leq T$), but then (for $j \geq T$) the rate of increase must be slower, which seems to be a property often obtained in practice.

Theorem 4 suggests that if the bound we have just analyzed approximates "well" $\sigma_n(H_{M\times N})$, then a similar behavior of the n-th singular value itself is to be expected. Thus if we assume the noise is not high enough to dominate the signal, then the best gap between $\sigma_n(\widetilde{H}_{M\times N})$ and $\sigma_{n+1}(\widetilde{H}_{M\times N})$, say, $g_n(M,N) = \sigma_n(\widetilde{H}_{M\times N}) - \sigma_{n+1}(\widetilde{H}_{M\times N})$, is likely to happen for $M = N$. Consequently, choosing square Hankel matrices seems reasonable. However, as we have no control over $\sigma_n(\widetilde{H}_{M\times N})$ and $\sigma_{n+1}(\widetilde{H}_{M\times N})$, the choice $M = N$ may not be optimal in all cases but the relaxed rule $M \approx N$ may be convenient.

5 Numerical Examples

In order to illustrate the observations concerning the role of $\sigma_n(H_{M\times N})$ in separating signals from noise, as well as the behavior of the bounds on $k_2(H)$ and the condition number itself, we have carried out a number of numerical experiments of which the most relevant are presented below. We shall consider bound (3.13) and that obtained from (3.18) where we use Bazán's bound for $k_2(W)$. These are denoted respectively by $B(k_2(H))$ and $\breve{B}(k_2(H)$. The condition number $k_2(H)$ was computed by using singular values extracted from (1.11).

Example 1: Hankel matrix related to a vibratory system

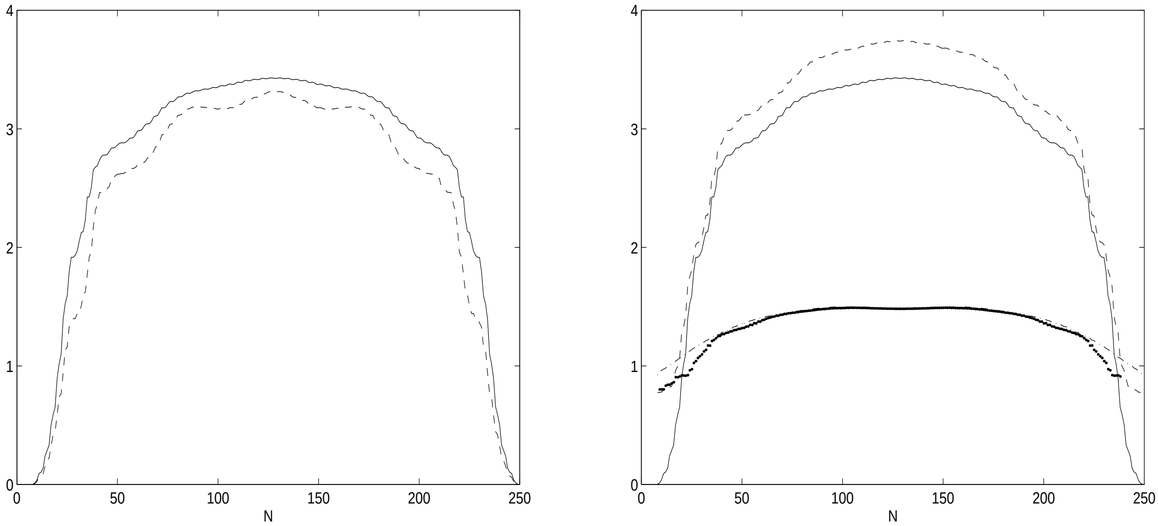
We consider a simulated mechanical system whose impulse response is defined by

$$h(t) = 0.2e^{-0.06t}\sin(25t) + 0.16e^{-0.056t}\sin(27t) + 0.12e^{-0.09t}\sin(18t) + 0.15e^{-0.2t}\sin(15t).$$

As this is a real signal, its sampled version (1.2) comprises 8 exponentials, which implies that the associated Hankel matrix H has rank 8 (i.e. $n = 8$). To illustrate the role that $\sigma_8(H_{M\times N})$ plays in the crucial problem of choosing a pair (M,N) that maximizes the gap $g_8(M,N)$, we compute $\sigma_8(H_{M\times N})$, $\sigma_8(\widetilde{H}_{M\times N})$, $\sigma_9(\widetilde{H}_{M\times N})$ and $\|E\|$ for all pairs (M,N) such that $M+N = 256$ (i.e. $L = 255$), where we use zero-mean Gaussian noise and $\Delta t = 0.05$.

Results corresponding to a noise level $\|\mathbf{e}\|_2/\|\mathbf{h}\|_2 \approx 40\%$ (standard deviation 0.05), where $\mathbf{h}$ and $\mathbf{e}$ are vectors respectively containing the pure signal and noise, are shown in Figure 1-(b). Notice in this figure that the choice of M, N that yields the best gap $g_8(M,N)$ is $M = N$. This not only agrees with our theory but also emphasizes the importance of choosing the dimensions well: if we chose $M = 30$, no clear gap appears. The behavior of both $\sigma_8(H_{M\times N})$ and its lower bound (4.1), displayed in Figure 1-(a), also illustrates what was predicted in theory: the maximum value for both occurs at $M = N = 128$.

In the second part of this experiment we compute $B(k_2(H))$, $\breve{B}(k_2(H))$, as well as those corresponding to the finite Hankel matrix H_N for several values of N, as expressed in Corollary 1.



(a) (b)

Figure 1: (a): $\sigma_n(H_{M\times N})$ (solid line) and its lower bound (4.1) (dashed line). (b): $\sigma_n(H_{M\times N})$ (solid line), $\sigma_n(\widetilde{H}_{M\times N})$ (dashed line), $\sigma_{n+1}(\widetilde{H}_{M\times N})$ (dotted-line), and $\|E\|$ (dashed-dotted line) as functions of (M, N) constrained to $M + N = 256$

As a result we obtain

$$B(k_2(H)) = 31.1090, \quad \text{and} \quad \breve{B}(k_2(H)) = 205.0791,$$

which show the superiority of our bound compared with that derived from (3.18). Values of the bounds on $k_2(H_N)$ are displayed in Figure 2. What is interesting here is that the bounds on $k_2(H_N)$ approach $B(k_2(H))$ relatively well when $N \geq 130$ (for $N = 132$, the bound is 31.2887). This is because the number D_N^2 for those values of N is very small, thus ensuring the condition $D_N^2 \ll (n-1)\delta^2$ in (3.5), which in turn, enforces reasonable values for the bound (3.11). For this example $(n-1)\delta^2 = 0.0695$, $D_{132}^2 = 0.0057$ and $D_\infty^2 = 0.0027$. The number $k_2(H_N)$ itself varies much with N. While its maximum value is about 1.9691×10^4, which occurs at $N = 8$, it rapidly decreases reaching a minimum close to 1.88 at $N = 36$, and then starts to slightly increase with small oscillations until convergence is reached. For this example we obtain

$$k_2(H) = 4.4977.$$

Example 2: Hankel matrix related to a nuclear magnetic resonance (NMR) signal

This example is that of a signal composed of 5 complex exponentials, representing a typical ^{31}P NMR signal [13]. The signal parameters as well as the separations of the signal modes, δ_i, are presented in Table 1. In contrast with the signal of the previous example, the signal in this case is more damped and its Fourier spectrum features closely overlapping peaks (see Figure 3). This means the signal is very sensitive to noise because of the two closely spaced modes.

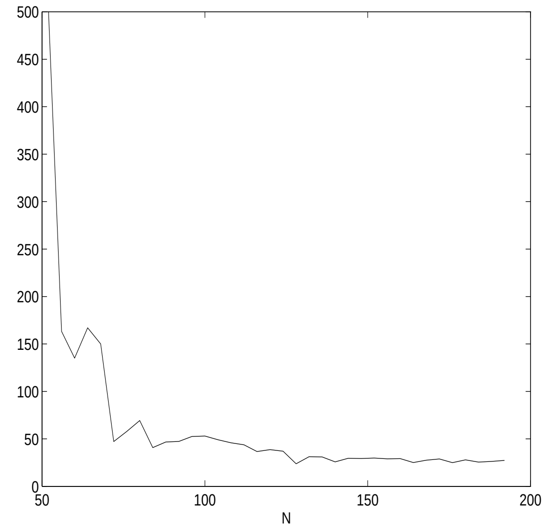


Figure 2: Bound of Corollary 1 on $k_2(H_N)$ as a function of N

We perform the same analysis as for Example 1. The available data for analyzing the gap $g_5(M, N)$ consists of 239 noisy samples ($L = 239$) obtained at a rate of 10kHz (i.e. $\Delta t = 0.0001s$) where we use a zero-mean Gaussian noise with unit standard deviation. The results again confirm our theory and are shown in Figure 4. Notice that in this case however, the maximal gap seems to happen slightly on the left (and thus on the right, by symmetry) of $\frac{1}{2}(L+1) = 120$, which is where $\sigma_5(H_{M\times N})$ is maximum.

Mode	r_l	d_l	ω_l (Hz)	z_l	$\lvert z_l \rvert$	δ_i^2
1	5.8921 + $\imath$1.5788	208	-1379	0.6342 - $\imath$0.7463	0.9794	0.1787
2	9.5627 + $\imath$2.5623	256	-685	0.8858 - $\imath$0.4067	0.9747	0.0643
3	5.7956 + $\imath$1.5529	197	-271	0.9663 - $\imath$0.1661	0.9805	0.0643
4	2.7046 + $\imath$0.7247	117	353	0.9642 + $\imath$0.2174	0.9884	0.0100
5	16.4207 + $\imath$4.3999	808	478	0.8811 + $\imath$0.2729	0.9224	0.0100

Table 1: Signal parameters of a ^{31}P NMR signal.

That the signal is sensitive to noise is easily verified: for $N = 5$, $k_2(H_N) = 1.7819 \times 10^6$. However, even if for that N the Hankel matrix is ill-conditioned, this no longer occurs for increasing values of the dimension (we obtain $k_2(H_N) = 3.4588$ for $N = 100$). The condition number $k_2(H)$, bound (3.13), and bound (3.18) reach the values

$$k_2(H) = 3.2637, \quad B(k_2(H)) = 65.4466, \text{ and } \breve{B}(k_2(H)) = 594.4352,$$

which illustrate once more the superiority of our bound compared with that of (3.18). Explanation for the "low" value of $B(k_2(H))$ again involves the behavior of D_N^2, the separation of the modes z_l and the size of the modes themselves. For this example $(n-1)\delta^2 = 0.0399$, $D_{100}^2 = 0.0310$, and $D_\infty^2 = 0.0298$.

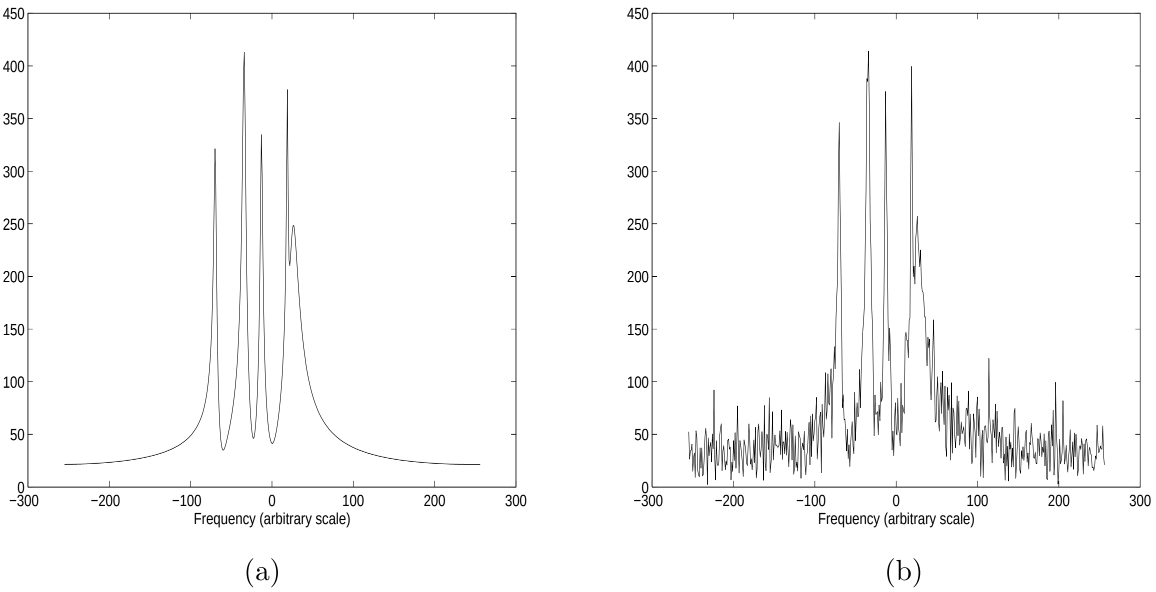


Figure 3: (a): Fourier spectrum of NMR pure signal. (b): Fourier spectrum of noisy NMR signal.

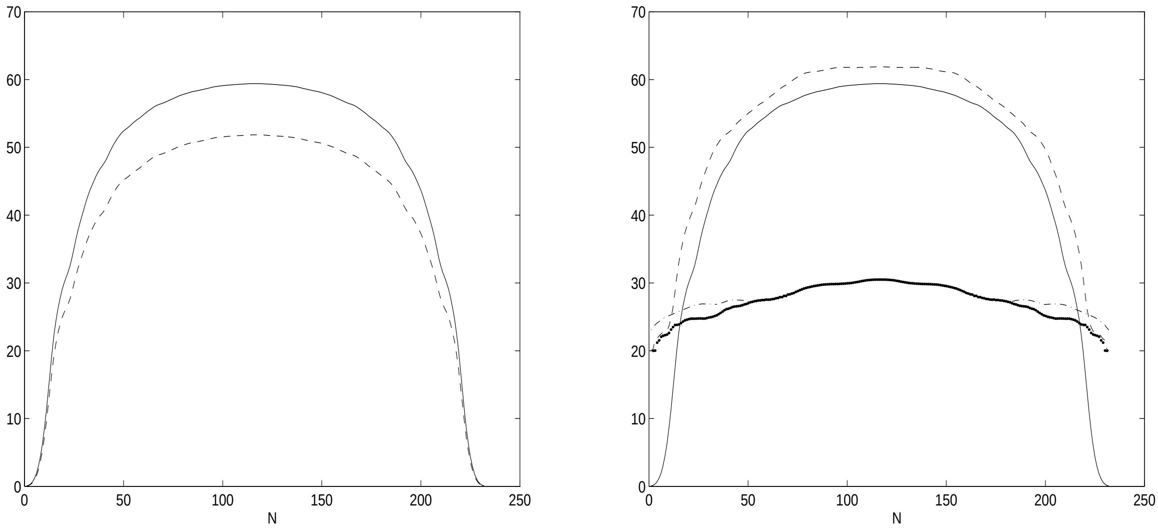
For the purpose of observing how the n-th singular value of the perturbed and unperturbed Hankel matrix behave when the noise is no longer zero-mean Gaussian, the same experiment was carried out with a noise uniformly distributed in an interval $[-t, t]$, where t is chosen to yield approximately the same noise level as in the above example. This experiment was motivated by the fact that this noise model is preferred for certain applications (Cummings and Pike [5]). The corresponding results are not presented because the behavior of the singular values and $\|E\|_2$ is practically identical to that illustrated in Figure 1-(b).

6 Conclusions

We have conducted an analysis on the conditioning of infinite Hankel matrices whose entries are samples of complex valued signals and expressed the results under the form of upper bounds. The bounds involve intrinsic characteristics of the signal such as, the number of spectral components n, the amount of damping, and the closeness of the signal modes. In particular, we have proved that $k_2(H)$ becomes moderate provided the signal is slightly damped, with the effect strengthened when the signal modes are not extremely close to each other and n is small. As the number of spectral components in signal processing is typically not very large, at least in several applications, we conclude that the associated Hankel matrices should be well-conditioned.

Moreover, given a finite set of samples of the pure signal, we have analyzed a lower bound for the smallest non zero singular value of finite Hankel matrices containing the data, which suggests that this singular value is maximized when these matrices are approximately square. This is of interest in signal processing applications where that singular value plays a crucial role in the separation of signal from noise. Numerical examples taken from modal analysis and NMR illustrate the theory.

The authors are aware that further research is desirable for the case where the signal damping is stronger: our bounds could be pessimistic in this case if the modes are not well separated, but numerical experiments indicate that the conditioning of the Hankel matrix remains acceptable.



(a) (b)

Figure 4: (a): $\sigma_n(H_{M\times N})$ (solid line) and its lower bound (4.1) (dashed line). (b): $\sigma_n(H_{M\times N})$ (solid line), $\sigma_n(\widetilde{H}_{M\times N})$ (dashed line), $\sigma_{n+1}(\widetilde{H}_{M\times N})$ (dotted-line), and $\|E\|$ (dashed-dotted line) all as functions of (M, N) constrained to $M + N = 240$

References

[1] F. S. V. Bazán. Conditioning of Rectangular Vandermonde Matrices with nodes in the Unit Disk. *SIAM J. Matrix Analysis and Applications*, vol. 21(2): 679-693, 2000.

[2] F. S. V. Bazán and L. H. Bezerra. On Zero Location of Predictor Polynomials. *Numer. Linear Algebra With Applications*, 4(6): 459-468, 1997.

[3] F. S. V. Bazán and Ph. L. Toint. Singular values of predictor matrices and signal eigenvalue bounds. Technical Report 5, Department of Mathematics, FUNDP, Namur, Belgium, February 1998.

[4] A. Ben-Israel and Thomas N. E. Greville. Generalized Inverses: Theory and Applications. J. Wiley and Sons, New York, 1974.

[5] H. Z. Cummings and E. R. Pike. Photon Correlation and Light Beating Spectroscopy. Plenum-Press, New-York, 1974.

[6] B. De Moor. The singular Value decomposition and long and Short Spaces of Noisy Matrices. *IEEE Trans. On Signal Proc.*, 41(9): 2826–2838, September, 1993.

[7] R. Fierro, L Vanhamme, and S. Van HuffelH. Total Least Squares Algorithms Based on Rank-Revealing Complete Orthogonal Decompositions. In S. Van Huffel Editor, *Recent Advances in Total Least-Squares Techniques and Errors-in-Variables Modeling*. SIAM, Philadelphia, USA, 1997.

[8] F. R. Gantmacher. Theory of Matrices. Vol. 2, Chelsea, New-York, 1959.

[9] G. H. Golub and C. F. Van Loan. Matrix Computations. Johns Hopkins University Press, Baltimore, third edition, 1996.

[10] W. B. Gragg and L. Reichel. On Singular Values of Hankel operators of finite rank. *Linear Algebra and Its Applications*, 121: 53-70, 1989.

[11] R. A. Horn and C. R. Johnson. Matrix analysis. Cambridge University Press, Cambridge, UK, 1990.

[12] R. A. Smith. The condition numbers of the matrix eigenvalue. *Numerische Mathematik*, 10: 232-240,1967.

[13] S. Van Huffel. Enhanced resolution based on minimum variance estimation and exponential data modeling. *Signal Processing*, 33: 333–355, 1993.

[14] S. Van Huffel, editor. Recent advances in Total Least-Squares Techniques and Errors-in-Variables Modeling. SIAM, Philadelphia, USA, 1977.

[15] M. Wei and G. Majda. A New Theoretical Approach for Prony's Method. *Linear Algebra and Its Applications*, 136: 119-132, 1990.

[16] J. H. Wilkinson. The Algebraic Eigenvalue Problem. Oxford University Press, Oxford, Great Britain, 1965.