Singular value analysis of predictor matrices

Bazán, Fermin S.V; Toint, Philippe

Published in:
Mechanical Systems and Signal Processing

DOI:
10.1006/mssp.2000.1370

Publication date:
2001

Document Version
Early version, also known as pre-print

Publications of the Department of Mathematics
Federal University of Santa Catarina
Florianópolis - SC - Brasil
http://www.mtm.ufsc.br

Singular Value Analysis of Predictor Matrices

by F.S.V. Bazán[1] and Ph. L. Toint[2]

Revised September 22, 2000

To appear in Mechanical Systems and Signal Processing

[1] Department of Mathematics,
Federal University of Santa Catarina,
Florianopolis, Santa Catarina 88040-900, Brasil.
Email: fermin@mtm.ufsc.br

[2] Department of Mathematics,
Facultés Universitaires ND de la Paix,
61, rue de Bruxelles, B-5000 Namur, Belgium.
Email: pht@math.fundp.ac.be

Singular Value Analysis of Predictor Matrices

F. S. V. Bazán *   Ph. L. Toint

Revised September 22, 2000

Abstract

Predictor matrices arise in problems of science and engineering where one is interested in predicting future information from previous ones using linear models. The solution of such problems depends on an accurate estimate of a part of the spectrum (*the signal eigenvalues*) of these matrices. In this paper, singular values of predictor matrices are analyzed and formulae for their computation are derived. By applying a well-known eigenvalue-singular value inequality to our results, we deduce lower and upper bounds on the modulus of signal eigenvalues. These bounds depend on the dimension of the problem and allow us to show that the magnitude of signal eigenvalues is relatively insensitive to small perturbations in the data, provided the signal is slightly damped and the dimension of the problem is large enough. The theory is illustrated by numerical examples including the analysis of a signal arising from experimental measurements.

Key words. Singular values, eigenvalues, linear prediction, time series, exponential modeling

1 Introduction

Predictor matrices often arise in a number of areas such as modal analysis, speech processing, system identification, etc, where prediction of future from previous information, is of primary interest. In many cases, this prediction is computed using linear models. The nature of the information itself depends on the particular application under study and often has the form of discrete time series, commonly known in engineering as *discrete time signals*. More precisely, given a set of real or complex-valued observations $h_\ell, h_{\ell+1}, \ldots, h_{\ell+N-1}$, a linear prediction model assumes that the future value $h_{\ell+N}$ has the form

$$\bar{f}_1 h_\ell + \bar{f}_2 h_{\ell+1} + \cdots + \bar{f}_N h_{\ell N-1} = h_{\ell+N}, \quad \ell \geq 0. \tag{1.1}$$

*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).

In this formulation, N is the *order* of the model and the $\bar{f}$'s, are complex coefficients known as *predictor parameters.* These coefficients reflect intrinsic properties of the signal such as frequencies, plane waves, damping factors, etc, whose estimation from a finite data set $\{h_k\}_{k=0}^{L}$, is an important problem in areas such as system identification and spectral estimation, among others. The linear prediction model is also currently used in other areas such as economy and zoology; see [21] and [6] for details. In this work, we focus on those applications where the data are assumed to be of the form

$$h_k = \sum_{j=1}^{n} r_j (e^{s_j \Delta t})^k = \sum_{j=1}^{n} r_j z_j^k, \quad k = 0, 1, \ldots,$$

where $r_j, s_j \in \mathbb{C}$, $s_j \neq s_k$ for $j \neq k$, $s_j = \alpha_j + i\omega_j$, $\imath = \sqrt{-1}$, and $\alpha_j \leq 0$. In these applications the problem is to estimate the parameters r_j, s_j and the number n (the *signal parameters*), from a finite sampling of the observed signal. We always assume $\omega_j \Delta t < \pi$. The classical approach for the problem is relatively simple in the noiseless case: the parameters s_j are extracted from the roots of the so-called *forward predictor polynomial*

$$P_f(t) = \bar{f}_1 + \bar{f}_2 t + \cdots + \bar{f}_N t^{N-1} - t^N, \tag{1.2}$$

whose coefficients $\bar{f}_j$ are estimated by solving the set of linear prediction equations

$$\begin{bmatrix} h_\ell & h_{\ell+1} & \cdots & h_{\ell+N-1} \\ h_{\ell+1} & h_{\ell+2} & \cdots & h_{\ell+N} \\ \vdots & \vdots & \cdots & \vdots \\ h_{\ell+M-1} & h_{\ell+M} & \cdots & h_{\ell+M+N-2} \end{bmatrix} \begin{bmatrix} \bar{f}_1 \\ \bar{f}_2 \\ \vdots \\ \bar{f}_N \end{bmatrix} = \begin{bmatrix} h_{\ell+N} \\ h_{\ell+N+1} \\ \vdots \\ h_{\ell+M+N-1} \end{bmatrix}, \tag{1.3}$$

where we assume $M \geq N \geq n$, and n is the rank of the coefficient matrix. The underlying idea behind this is that n zeros of $P_f(t)$, known as *signal zeros*, are of the form $z_j = e^{s_j \Delta t}$, a result early proven by R. de Prony for the case $M = N = n$. For details of Prony's method, see Section 9.4 in Hildebrand [11]. Applications of Prony's method are encountered in number of areas such as acoustics, electromagnetics, and structural dynamics, among others, see, e.g., Magda, Strauss and Wei [20], Braun and Ram [7, 8], Kumaresan [16], and Kurka [19]. A new theoretical approach for Prony's method is described in Wei and Majda [25]. Once the signal zeros are available, the problem of estimating the parameters r_j is simple and we do not comment any further about this.

One could also use the linear prediction equations in the reverse order, replacing (1.1) by

$$\bar{b}_1 h_\ell + \bar{b}_2 h_{\ell+1} + \cdots + \bar{b}_N h_{\ell N-1} = h_{\ell-1}, \quad \ell \geq 1. \tag{1.4}$$

Then the parameters s_j are extracted from the zeros of the *backward predictor polynomials*

$$P_b(t) = \bar{b}_N + \bar{b}_{N-1} t + \cdots + \bar{b}_1 t^{N-1} - t^N, \tag{1.5}$$

whose coefficients are estimated as above. In this case, the signal zeros are $z_j^{-1} = e^{-s_j \Delta t}$ [16]. However, as only the signal zeros are of interest, one is faced with the problem of separating them from the other $N-n$ zeros, called *extraneous zeros*, which appear as a consequence of choosing $N > n$ since n is not known in advance. This separation turns out to be difficult since the extraneous zeros depend on how one chooses the coefficients $\bar{f}_j$ from the infinitely many solutions of the system (1.3). Further details about the zeros of predictor polynomials can be found in Kumaresan [16], and also in Cybenko [9], where the problem is examined in the framework of infinite dimensional Hilbert spaces. A more recent approach, where signal zeros are viewed as eigenvalues of predictor matrices (*the signal eigenvalues*), can be found in Bazán and Bezerra [2] and Bezerra and Bazán [4].

The problem, however, becomes difficult when the data are corrupted by noise, since both the rank n and the parameters $\bar{f}_j$ must be estimated from a linear prediction system whose coefficient matrix is generally of full rank, though this can be often circumvented by taking $N \gg n$, see, e.g., [17], [16], [22] and [23]. But, as polynomial root-finding methods are time consuming, new approaches based on estimates of the so-called *signal spaces* (the row or column space of the data matrix) are actually preferred. In these techniques, the signal zeros emerge as eigenvalues of a small $n \times n$ matrix. Methods in this category include Kung's method [18], Eigensystem Realization Algorithm (ERA) of Juang and Pappa [14], HTLS of Van Huffel, Chen, Decanniere and Van Hecke [24], OPIA of Bazán and Bavastri [1], and the matrix pencil method of Hua and Sarkar [13], among others. Many references about both polynomial and subspace approaches can be found in [23]. However, despite the bursting activity in new approaches, little is known about signal eigenvalue sensitivity, an intrinsic component of the problem.

The goal of this paper is to perform a singular value analysis of predictor matrices, the results of which provide insight into the sensitivity of these eigenvalues. Our analysis relies on the fact that, since the eigenvalues λ_j of any matrix $A \in \mathbb{C}^{N \times N}$ satisfies the inequalities

$$\sigma_N \leq |\lambda_j| \leq \sigma_1 \quad j = 1, 2, \cdots, N, \tag{1.6}$$

(see Golub and Van Loan [10], page 318), where σ_N and σ_1 denote the smallest and largest singular value of A, then reliable information about signal eigenvalue sensitivity can be drawn from (1.6), provided these singular values are available. We provide analytic formulae for all singular values of a class of predictor matrices and analyze the asymptotic behavior of the bounds (1.6) for N large. As a consequence, we show that the magnitude of the signal eigenvalues becomes relatively insensitive to small perturbations on the data, provided mild conditions are satisfied.

We first analyze in Section 2 the localization of signal eigenvalues extracted from the spectrum of predictor matrices. The main results are presented in Section 3 where we give an exact description of the annulus (1.6) for the class of predictor matrices obtained by orthogonal projection: we show that an upper bound on the width of this annulus shrinks as the dimension of the problem increases, and that it asymptotically becomes small provided the signal is slightly damped. Finally, Section 4 presents some numerical results which illustrate our theoretical analysis.

2 Predictor matrices and basic results

Predictor matrices are defined as follows in the noiseless case. Let $H(\ell)$ be the $M \times N$ Hankel matrix of the system (1.3). We say that the $N \times N$ matrix F is a *forward predictor matrix* if

$$H(\ell+1) = H(\ell)F, \quad \forall \ell \geq 0. \tag{2.1}$$

Similarly, B is a *backward predictor matrix* if for $\ell \geq 1$, it satisfies

$$H(\ell-1) = H(\ell)B. \tag{2.2}$$

Notice that $H(\ell)$ can be factored as

$$H(\ell) = VZ^{\ell}RW, \tag{2.3}$$

where $Z = \text{diag}(z_1, \ldots, z_n)$, $R = \text{diag}(r_1, \ldots, r_n)$, V is an $M \times n$ Vandermonde matrix with z_k^{j-1} as its (j,k) entry and W the transpose of the matrix that consists of the N first rows of V. Hence, we have that (2.3) is a full-rank factorization of $H(\ell)$ and that for all $\ell \geq 0$, $\text{rank}[H(\ell)] = n$. The column space of $H(\ell)$, denoted by $\mathcal{R}(H(\ell))$, is spanned by the columns of matrix V, while the row space of $H(\ell)$, $\mathcal{R}(H(\ell)^*)$, is spanned by the columns of W^*. Here the symbol * stands for complex conjugate transpose. From (2.3) also follows that $\mathcal{N}(H(\ell))$, the null space of $H(\ell)$, is equal to the null space of W, and that therefore

$$\mathcal{R}(H(\ell)^*) = [\mathcal{N}(W)]^{\perp}. \tag{2.4}$$

Furthermore, if $\mathcal{P}$ denotes the orthogonal projector onto $\mathcal{R}(H(\ell)^*)$, then we have that

$$\mathcal{P} = H(l)^{\dagger}H(l) = W^{\dagger}W = W^*W^{\dagger *}, \tag{2.5}$$

where † stands for the Moore-Penrose pseudo-inverse. The reader is referred to [5] for details on projections and pseudo-inverses. We now observe that (2.1) and (2.2) have infinitely many solutions because $H(\ell)$ is rank deficient. The solutions set of (2.1), S_F, is

$$S_F = \{F \mid F = \widehat{F} + (I - \mathcal{P})X, \quad X \in R^{N \times N}\}, \tag{2.6}$$

where $\widehat{F} = H(l)^{\dagger}H(l+1)$. Similarly, if we set $\widehat{B} = H(l)^{\dagger}H(l-1)$, then the set of backward predictor matrices is

$$S_B = \{B \mid B = \widehat{B} + (I - \mathcal{P})X, \quad X \in R^{N \times N}\}. \tag{2.7}$$

We observe that the signal eigenvalues can be extracted from any forward or backward predictor matrix for, if one substitutes F in (2.1), one has that,

$$WF = ZW, \tag{2.8}$$

which shows that the rows of W are left eigenvectors of F corresponding to the signal eigenvalues. In the same way, (2.2) implies that

$$WB = Z^{-1}W, \tag{2.9}$$

and thus the rows of W are left eigenvectors of B associated with the eigenvalues z_j^{-1}. However, if signal eigenvalues are independent of which predictor matrix is chosen in the class, this is not the case for the extra $N-n$ eigenvalues (*the extraneous eigenvalues*). It turns out that a analysis of the localization of these extraneous eigenvalues is possible, provided we restrict ourselves to a suitable class of predictor matrices. We focus here on two interesting choices: the matrix $\widehat{F}$ (or $\widehat{B}$) and minimal norm predictor matrices with companion structure. This last class covers, if prediction is carried out in the forward direction, predictor matrices of the form

$$F = [e_2 \; e_3 \; \cdots \; e_N \; f],$$

where e_j is the j-th vector of the canonical basis and the last column vector, $f = [f_1 \; f_2 \cdots \; f_N]^T$, is the solution of the linear system

$$H(\ell)\, f = H(\ell+1)e_N. \tag{2.10}$$

of minimal 2-norm. If prediction is carried out in the reverse order instead, backward companion predictor matrices have the form

$$B = [b \; e_1 \; e_2 \; \cdots \; e_{N-1}],$$

where the first column vector, $b = [b_1 \; b_2 \cdots \; b_N]^T$, is the solution of the system

$$H(\ell)\, b = H(\ell-1)e_1, \quad \ell \geq 1 \tag{2.11}$$

of minimal 2-norm. The following result gives information about the eigenvalues of the above predictor matrices.

Theorem 1 *Let $\widehat{F}$ and $\widehat{B}$ be as in (2.6) and (2.7), respectively, and let F and B be the forward and backward minimal norm companion predictor matrices. Then, provided $N > n$, we have that*

(a) F and B are both nonsingular.

(b) $|\hat{\lambda}_k(\cdot)| < 1, \quad k = 1, \dots N-n$ *where* $\hat{\lambda}(\cdot)$ *denotes an extraneous eigenvalue and* $(\cdot)$ *any of the matrices* $\widehat{F}$, $\widehat{B}$, F *or* B.

Proof. We note that to prove (a) for F, it suffices to prove that $e_1^* f \neq 0$. Observe that $f \in \mathcal{R}(H(l)^*)$. We then verify that e_1 does not belong to either $\mathcal{N}(W)$ or to its orthogonal complement. The first of these two claims follows from the fact that $We_1 = e$, where e is the vector in $\mathbb{C}^n$ of all ones. To see the second, we consider the system $W^* x = e_1$. If we select n equations of this system starting from the second one, we obtain a square nonsingular homogeneous system whose unique solution is $x = 0$. However, this is in contradiction with the first equation, which shows that the system is incompatible. Thus $e_1 \notin \mathcal{R}(H(l)^*)$, which ensures that $e_1^* f \neq 0$, as claimed. One can similarly check that B is nonsingular, and thus part (a) of the theorem holds. The proof of (b) involving companion matrices can be found in [2]. We now prove that (b) for $\widehat{F}$ and $\widehat{B}$. In order to see that the extraneous eigenvalues of $\widehat{F}$ fall inside the unit circle, notice that, as $\widehat{F} = H(l)^\dagger H(l+1) = W^\dagger Z W$, it is immediate to see that $\lambda(\widehat{F}) = \{z_1, \dots, z_n\} \cup \{0\}$ (see Horn and Johnson [12], Theorem 1.3.20). A similar argument can be applied for $\widehat{B}$, which concludes the proof. □

As this theorem describes completely the locations of all eigenvalues of the predictor matrices $\widehat{F}$, $\widehat{B}$, F and B, what remains to do is to determine their singular values. We first start with a technical lemma that allows us to compute the singular spectrum of companion predictor matrices. The determination of the singular spectrum of $\widehat{B}$ and $\widehat{F}$ is slightly more involved and is postponed to the next section.

Lemma 2 *Let* u_1, u_2, v_1 *and* v_2 *be vectors in* $\mathbb{C}^N$, $N > 2$, *such that at least one of the inner products* $v_1^* u_1$ *or* $v_2^* u_2$ *is different of* -1. *Suppose that the rank-two modification of the identity given by* $I + u_1 v_1^* + u_2 v_2^*$ *is nonsingular. Then we have that,*

$$\det(I + u_1 v_1^* + u_2 v_2^*) = 1 + v_1^* u_1 + v_2^* u_2 + v_1^* u_1 v_2^* u_2 - v_2^* u_1 v_1^* u_2, \tag{2.12}$$

and that, the associated characteristic polynomial is

$$p(\lambda) = (1-\lambda)^{N-2}[\lambda^2 - (2 + v_1^* u_1 + v_2^* u_2)\lambda + 1 + v_1^* u_1 + v_2^* u_2 + v_2^* u_2 v_1^* u_1 - v_2^* u_1 v_1^* u_2]. \tag{2.13}$$

Proof. We assume, without loss of generality, that $v_1^* u_1 \neq -1$. It then follows that $I + u_1 v_1^*$ is nonsingular since $\det(I + u_1 v_1^*) = 1 + v_1^* u_1 \neq 0$. Hence, using properties of the determinant, we have that

$$\begin{aligned} \det(I + u_1 v_1^* + u_2 v_2^*) &= \det(I + u_1 v_1^*) \det(I + (I + u_1 v_1^*)^{-1} u_2 v_2^*)) \\ &= (1 + v_1^* u_1)(1 + v_2^* (I + u_1 v_1^*)^{-1} u_2), \end{aligned}$$

and the first part of the lemma follows after applying the Sherman-Morrison formula to the last right-hand side. On the other hand, given that

$$p(\lambda) = \det(I + u_1 v_1^* + u_2 v_2^* - \lambda I) = \det((1-\lambda)I + u_1 v_1^* + u_2 v_2^*),$$

since $p(1) = 0$ and $N > 2$, we have that $\lambda = 1$ is an eigenvalue of $I + u_1 v_1^* + u_2 v_2^*$. If $\lambda \neq 1$,

$$p(\lambda) = (1-\lambda)^N \det(I + (1-\lambda)^{-1} u_1 v_1^* + (1-\lambda)^{-1} u_2 v_2^*),$$

and the second part of the lemma is then obtained by applying (2.12) in this equation. □

Thus it suffices to extract the eigenvalues associated with a quadratic polynomial in (2.13) to obtain the eigenvalues of the perturbed matrix, since the remaining ones are equal to one. We illustrate this by considering the problem of computing the singular spectrum of the backward companion matrix B introduced above. In fact, since the singular values of B can be computed from the eigenvalues of BB^*, we observe that

$$BB^* = [b \; e_1 \; \ldots \; e_{N-1}] \begin{bmatrix} b^* \\ e_1^* \\ \vdots \\ e_{N-1}^* \end{bmatrix} = bb^* + e_1 e_1^* + \cdots + e_{N-1} e_{N-1}^*,$$

where b is the minimum 2-norm solution of 2.11, and hence that

$$BB^* = I + bb^* - e_N e_N^*.$$

By applying Lemma 2 to BB^*, with $u_1 = v_1 = b$ and $u_2 = -v_2 = e_N$, we find that the characteristic polynomial of BB^* is $p(\lambda) = (1-\lambda)^{N-2}[\lambda^2 - \lambda(1 + \|b\|^2) + 4|b^* e_N|^2]$. Hence, we have that the singular spectrum of B, $\sigma(B)$, is of the form

$$\sigma(B) = \{\sigma_1(B), 1, 1, \ldots, 1, \sigma_N(B)\},$$

where

$$\sigma_1^2(B), \sigma_N^2(B) = \frac{\|b\|^2 + 1 \pm \sqrt{(\|b\|^2+1)^2 - 4|e_N^* b|^2)}}{2}. \qquad (2.14)$$

This result is not new, (see for instance [15]), but the authors are not aware of a proof along the lines developed here. We now use it to obtain an important eigenvalue bound. Since for each signal eigenvalue λ we have that,

$$\sigma_N(B) \le |\lambda| \le \sigma_1(B) \le \sqrt{\frac{\|b\|^2 + 1 + \sqrt{(\|b\|^2+1)^2 - 4|e_N^* b|^2)}}{2}} \le \sqrt{1 + \|b\|^2}, \qquad (2.15)$$

an interesting upper bound can be immediately derived provided $\|b\|^2$ is small enough (remember in this case $|\lambda| \geq 1$). In our context, as shown in [3], it is fortunate that $\|b\|^2 \approx 0$ in many practical applications, provided the dimension of the problem is sufficiently large. If this is true, then the form of the upper bound indicates that it could be rather tight. Hence, this preliminary analysis suggests that reliable bounds could be obtained provided they only depend on quantities similar to the right-hand side of (2.15). Unfortunately, no lower bound of interest can be obtained from the left inequality because $\sigma_N(B) \approx 0$, which motivates our search for a better lower bound.

3 Signal eigenvalue bounds

In spite of the promising quality of the above upper bound, the link of the signal eigenvalues with the solution of a "large eigenvalue problem" seems to generate a new inconvenience, in that it appears to require that the prediction matrix is sufficiently large. In this section, we shall show that this can be circumvented provided signal eigenvalue bounds are derived by using the singular spectrum of predictor matrices *obtained via orthogonal projections*. We say that the $n \times n$ matrix $F_{\mathcal{P}}$, is a forward predictor matrix obtained by orthogonal projection if, it is of the form

$$F_{\mathcal{P}} = V_1^* F V_1, \tag{3.16}$$

where V_1 denotes any $N \times n$ matrix with orthonormal columns that span $\mathcal{R}(H(\ell)^*)$. The motivation for this definition relies on the fact that the spectrum of $F_{\mathcal{P}}$, $\lambda(F_{\mathcal{P}})$, only contains the n signal eigenvalues, since $\lambda(F_{\mathcal{P}}) = \lambda(\mathcal{P}F) = \lambda(\widehat{F})$ when zero eigenvalues are discarded. Similarly, $B_{\mathcal{P}}$ is a backward predictor matrix obtained by orthogonal projection if it is of the form

$$B_{\mathcal{P}} = V_1^* B V_1. \tag{3.17}$$

The spectrum of $B_{\mathcal{P}}$ then consists of the reciprocal of the signal eigenvalues. Thus, if eigenvalue bounds are derived from the singular spectrum of these matrices, the signal eigenvalues are related to a small $n \times n$ eigenvalue problem. The purpose of this section is therefore to develop a singular value analysis of these matrices and to analyze the corresponding signal eigenvalue bounds.

Theorem 3 *Suppose $\widehat{B}$ is the backward predictor matrix introduced in (2.7). Then the singular spectrum of $\widehat{B}$, $\sigma(\widehat{B})$, satisfies*

$$\sigma(\widehat{B}) = \sigma(B_{\mathcal{P}}) \cup \{0\}, \tag{3.18}$$

where

$$\begin{aligned}\sigma_1^2(B_{\mathcal{P}}) &= \frac{2+\|b\|^2-\|p_N\|^2+\sqrt{(\|b\|^2+\|p_N\|^2)^2-4|e_N^*b|^2}}{2},\\ \sigma_i(B_{\mathcal{P}}) &= 1, \quad (i=2,\dots,n-1),\\ \sigma_n^2(B_{\mathcal{P}}) &= \frac{2+\|b\|^2-\|p_N\|^2-\sqrt{(\|b\|^2+\|p_N\|^2)^2-4|e_N^*b|^2}}{2},\end{aligned} \tag{3.19}$$

where b and p_N are the first and the last column vectors of B and the projector $\mathcal{P}$, respectively.

Proof. From 2.7 we have that $\widehat{B} = \mathcal{P}B = V_1V_1^*B$. Hence,

$$\widehat{B}^*\widehat{B} = B^*V_1V_1^*V_1V_1^*B = (V_1^*B)^*(V_1^*B),$$

and therefore

$$\sigma(\widehat{B}) = \sigma(V_1^*B). \tag{3.20}$$

On the other hand, if we introduce $A = V_1^*B$, then $B_{\mathcal{P}} = AV_1$ and

$$B_{\mathcal{P}}B_{\mathcal{P}}^* = AV_1V_1^*A^* = A\mathcal{P}A^*. \tag{3.21}$$

But, as $\mathcal{P} = V_1V_1^* = W^\dagger W = W^*W^{\dagger *}$ by (2.5), then

$$A^* = B^*V_1 = B^*\mathcal{P}V_1 = B^*W^*W^{\dagger *}V_1 = W^*Z^{-*}W^{\dagger *}V_1,$$

where the last equality is because of (2.9). Hence, since $W^{\dagger *}W^* = I$, where I denotes the $n \times n$ identity matrix, we have, using (2.9) again, that

$$\mathcal{P}A^* = W^*W^{\dagger *}W^*Z^{-*}W^{\dagger *}V_1 = B^*W^*W^{\dagger *}V_1 = B^*V_1 = A^*. \tag{3.22}$$

Using this property in (3.21), we deduce that the singular values of $B_{\mathcal{P}}$ are the singular values of A, and thus the first part of the theorem follows from (3.20). To prove the second statement, notice that

$$AA^* = [V_1^*b, V_1^*e_1, \dots, V_1^*e_{N-1}] \begin{bmatrix} b^*V_1 \\ e_1^*V_2 \\ \vdots \\ e_{N-1}^*V_1 \end{bmatrix} = V_1^*bb^*V_1 + V_1^*e_1e_1^*V_1 + \cdots + V_1^*e_{N-1}e_{N-1}^*V_1,$$

can be rewritten as

$$AA^* = I + xx^* - yy^*,$$

where $x = V_1^* b$, and $y = V_1^* e_N$. By applying Lemma 2 to AA^* with $u_1 = v_1 = x$ and $u_2 = -v_2 = y$, we obtain that the spectrum of this matrix is formed by $n-2$ eigenvalues equal to 1, and the remaining ones, obtained from the roots of the polynomial in (2.13), are given by

$$\lambda_1, \lambda_2 = \frac{2 + \|x\|^2 - \|y\|^2 \pm \sqrt{(\|x\|^2 + \|y\|^2)^2 - 4(x^* y)^2}}{2}.$$

Now, observe that $\|p_N\| = \|V_1 V_1^* e_N\| = \|V_1(V_1^* e_N)\| = \|V_1 y\| = \|y\|$, since V_1 is an isometry, and that

$$\begin{aligned} \|x\|^2 &= b^* V_1 V_1^* b = b^* \mathcal{P} b = b^* b, = \|b\|^2, \\ |x^* y| &= |b^* V_1 V_1^* e_N| = |(b^* \mathcal{P}) e_N| = |b^* e_N| \end{aligned}$$

since $b \in \mathcal{R}[H(\ell)^*]$. We now observe that the largest value of the above roots, λ_1, say, is larger than one because the eigenvalues of B are larger than one in modulus. Assume now that $\lambda_2 > 1$. Then we obtain from its definition that

$$\left(\|x\|^2 - \|y\|^2\right)^2 > (\|x\|^2 + \|y\|^2)^2 - 4(x^* y)^2,$$

which can be simplified to $\|x\|^2 \|y\|^2 < |x^* y|^2$. This is impossible as it contradicts the Cauchy-Schwarz inequality, and we therefore deduce that $\lambda_2 \leq 1$. These roots are therefore $\sigma_1^2(B_{\mathcal{P}})$ and $\sigma_n^2(B_{\mathcal{P}})$, respectively, which concludes the proof. □

Thus, we have obtained an exact description of the singular spectrum of predictor matrices obtained by projection and the annulus (1.6) which provides lower and upper bounds for the signal eigenvalues. However, notice that these bounds are not immediately useful because they are derived from the expressions of the singular values given by the last theorem, which depend themselves on the projector $\mathcal{P}$. In order to overcome this difficulty we prove the following result, where we reintroduce the minimum norm solution f of (2.10).

Theorem 4 *Suppose that $\mathcal{A}_N$ is the annulus defined by*

$$\mathcal{A}_N = \left\{ z \in \mathbb{C} \;\middle|\; \frac{1}{\sqrt{1 + \|f\|^2}} \leq |z| \leq \sqrt{1 + \|b\|^2} \right\}, \tag{3.23}$$

where N is the dimension of the predictor matrix B, then the eigenvalues of $B_{\mathcal{P}}$ belong to $\mathcal{A}_N$.

Proof. We shall prove that both $\sigma_1(B_{\mathcal{P}})$ and $\sigma_n(B_{\mathcal{P}})$ belong to $\mathcal{A}_N$. In fact, using (3.19), we have

$$\sigma_1^2(B_{\mathcal{P}}) = \frac{2 + \|b\|^2 - \|p_N\|^2 + \sqrt{(\|b\|^2 + \|p_N\|^2)^2 - 4|e_N^* b|^2}}{2} \leq 1 + \|b\|^2,$$

which shows that $\sigma_1(B_{\mathcal{P}}) \in \mathcal{A}_N$. To prove that $\sigma_n(B_{\mathcal{P}})$ is not smaller than the inner radius of $\mathcal{A}_N$, we first show that $B_{\mathcal{P}} = F_{\mathcal{P}}^{-1}$: using the definitions of both matrices, (2.5), (2.8), (2.9), and the fact that $WW^\dagger = I$, we have that

$$B_{\mathcal{P}} F_{\mathcal{P}} = V_1^* B V_1 V_1^* F V_1 = V_1^* \mathcal{P} B \mathcal{P} F V_1 = V_1^* W^\dagger W B W^\dagger W F V_1 = V_1^* W^\dagger Z^{-1} W W^\dagger Z W V_1 = I,$$

as claimed. We next observe that this enables us to compute $\sigma_n(B_{\mathcal{P}})$ as

$$\sigma_n(B_{\mathcal{P}}) = 1/\sigma_1(F_{\mathcal{P}}), \tag{3.24}$$

and $\sigma_1(F_{\mathcal{P}})$ can be determined in a way similar to that used for the singular values of the backward predictor matrix. This yields that

$$\sigma_1(F_{\mathcal{P}})^2 = \frac{2 + \|f\|^2 - \|p_1\|^2 + \sqrt{(\|f\|^2 + \|p_1\|^2)^2 - 4|e_1^* f|^2}}{2} \leq 1 + \|f\|^2,$$

where p_1 is the first column of $\mathcal{P}$ and f the minimum norm solution of (2.10). This ensures that the left inequality of (3.23) is satisfied by $\sigma_n(B_{\mathcal{P}})$, which completes the proof.□

We now make the crucial observation that, depending on the dimension of the problem N, the inner and outer radii of $\mathcal{A}_N$ become excellent approximations of $\sigma_1(B_{\mathcal{P}})$ and $\sigma_n(B_{\mathcal{P}})$, respectively. This can be seen as follows. Since $e_N^* b$ is the independent coefficient of the characteristic polynomial associated to the companion matrix B, which is not zero by Theorem 1, then

$$|e_N^* b| = \prod_{k=1}^{N-n} |\widehat{\lambda}_k| \prod_{j=1}^{n} |\lambda_j|, \tag{3.25}$$

where $\widehat{\lambda}_k$ are the so-called extraneous eigenvalues and $\lambda_j = z_j^{-1}$. Hence, as $|\widehat{\lambda}_k| < 1$ by Theorem 1, and, since for N large enough, $|e_N^* b|^2 \approx 0$, from (3.19) we have that $\sigma_1^2(B_{\mathcal{P}}) \approx 1 + \|b\|^2$. A similar reasoning on $|e_1^* f|$ gives that $\sigma_1^2(F_{\mathcal{P}}) \approx 1 + \|f\|^2$, and the quality of these approximations improves when N increases.

Before stating our final result, we introduce two technical lemmas.

Lemma 5 *Define $G_{\mathcal{Q}} = V_2^* B V_2$, where V_2 is an $N \times (N-n)$ matrix whose columns form an orthonormal basis of $\mathcal{N}(H(\ell))$. Then,*

$$\begin{cases} \sigma_j^2(G_{\mathcal{Q}}) & = & 1, \ j = 1, 2 \ldots, N-n-1, \\ \sigma_{N-n}^2(G_{\mathcal{Q}}) & = & 1 - (1 + \|b\|^2)\|q_1\|^2, \end{cases} \tag{3.26}$$

where q_1 is the first column of the orthogonal projector $\mathcal{Q}$ onto $\mathcal{N}(H(\ell))$.

Proof. The proof is in appendix A1.

Lemma 6 *Let b and f be the first and last column vector of B and F respectively. Then*

$$\frac{1+\|b\|^2}{1+\|f\|^2} = \prod_{j=1}^{n} |z_j|^{-2}$$

for $N > n$. Moreover, both $\|b\|$ and $\|f\|$ decrease monotonically when N increases.

Proof. The proof is in appendix A2.

We now return to our main objective and continue analyzing the behavior of the width of $\mathcal{A}_N$ as function of $\|b\|$ and $\|f\|$, and, consequently, of N. Note that, because of Lemma 6, this reduces to analyzing the norm of the forward coefficients $\|f\|$. But, since (2.8) is equivalent to the system $Wf = Z^N e$, where e is the vector in $\mathbb{C}^n$ of all ones, which follows from (2.3), and, as

$$\|f\| = \|W^{\dagger} Z^N e\| \leq \|W^{\dagger}\| \sqrt{n} \beta^N, \tag{3.27}$$

where we set $\beta = \max\{|z_j|, \;\; j = 1, \ldots, n\}$, it suffices to analyze $\|W^{\dagger}\|$ as function of N. We now choose, for notational convenience, $N = p \times n$, $p > 1$. We also write

$$W = [W_0 \;\; DW_0 \;\; \ldots \;\; D^{p-1} W_0],$$

where W_0 is the $n \times n$ Vandermonde matrix whose (j,k) entry is z_j^{k-1}, and $D = Z^n$. Using these definitions, one can then prove that the smallest singular value of W, $\sigma_n(W)$, satisfies

$$\sigma_n(W) \geq \sigma_n(W_o) \left(\sum_{j=1}^{p} \gamma^{2n(j-1)} \right)^{1/2}, \tag{3.28}$$

where $\gamma = \min\{|z_j|, \;\; j = 1, \ldots, n\}$ (see [3], Theorem 1, for details). Hence, we have that

1) if $\alpha_j < 0$, *i.e.* the signal is damped, from (3.27) we have then that

$$\lim_{N \to \infty} \|f\| = 0, \tag{3.29}$$

because $\beta^N \to 0$ and $\|W^{\dagger}\| = 1/\sigma_n(W)$ is bounded as $\gamma < 1$ ensures that the sum in (3.28) is finite;

2) if the signal is undamped instead, *i.e.*, $\alpha_j = 0$, then (3.28) implies that $\|W^{\dagger}\| \to 0$ as $N \to \infty$ because $\beta = \gamma = 1$, and thus once more we obtain the limit (3.29) from (3.27).

But the limit (3.29) and Lemma 6 together give that

$$\lim_{N\to\infty}(1+\|b\|^2) = \prod_{j=1}^{n}|z_j|^{-2}(1+\lim_{N\to\infty}\|f\|^2) = \prod_{j=1}^{n}|z_j|^{-2}.$$

To conclude this discussion, we note that the last part of Lemma 6, this last limit and (3.23), ensure the following result.

Theorem 7 *The annulus that contains the signal eigenvalues associated to $B_{\mathcal{P}}$ has a monotonically decreasing width, i.e. $\mathcal{A}_{N+1} \subseteq \mathcal{A}_N$. Moreover, it is asymptotically described by*

$$\mathcal{A}_\infty = \{z \in \mathbb{C} \,/ 1 \le |z| \ \le \prod_{j=1}^{n}|z_j|^{-1}\}. \tag{3.30}$$

Thus, we have shown that the quality of the signal eigenvalues bounds depends on the speed at which $\|f\|^2$ approaches to zero as the dimension N increases. However, as illustrated in (3.27), this speed depends on the behavior of $\|W^\dagger\|$ as a function of N, which ultimately depends on the nature of the signal itself. The authors' experience is that in most of practical applications, moderate values of N are sufficient to ensure values of the norms of the predictor coefficients smaller than one (see, for instance, the examples discussed in [3]).

Now consider the case of slightly damped signals. For such signals, we know that the signal eigenvalues are relatively close to one, which we have shown to imply, for large N, that the width of the annulus (1.6) is small. Since these radii provide excellent approximations for $\sigma_1(B_{\mathcal{P}})$ and $\sigma_n(B_{\mathcal{P}})$, these singular values must be close to each other. Furthermore, the stability of the singular values (see Golub and Van loan [10], Section 8.3.1) guarantees that a small perturbation of the data will not alter this property. This, in turn, implies that the width of the annulus (1.6) remains small, even after a small perturbation. As a consequence, the magnitude of signal eigenvalues cannot vary by a large amount. Further, since these eigenvalues cannot fall outside of the annulus, this property suggests that the eigenvalues themselves should be insensitive to small perturbations on the data.

3.1 Connection of Predictor Matrices with Certain Subspace-Based Methods

Our goal here is to show that there exists a close relationship between predictor matrices obtained by projection orthogonal and certain matrices used by two known modal parameter identification methods. Specifically, we shall relate the forward predictor matrix of (3.16) with those matrices used by ERA and OPIA, and show that all these matrices share the same eigenvalues. In fact, we notice that OPIA uses an $n \times n$ matrix of the form

$$S = V^*FV, \tag{3.31}$$

where $V \in \mathbb{C}^{N\times n}$ is a matrix whose columns are right singular vectors related to the largest singular values of $H(\ell)$. This shows that the matrix used by OPIA is indeed a predictor matrix obtained by projection. On the other hand, the matrix used by ERA is

$$S_E = \Sigma^{-1/2} U^* H(\ell+1) V \Sigma^{-1/2}, \tag{3.32}$$

where V is as before, U contains the left singular values related to the largest singular values of $H(\ell)$ and Σ a diagonal matrix containing these singular values. Following reference [1] (see relations (17), (20), (22) and (27) therein), it can be proved that $S_E = \Sigma^{1/2} S \Sigma^{-1/2}$. This shows that both S and S_E share the same eigenvalues, and the result continues to hold regardless of whether the signal is perturbed or not. We thus conclude that a unified signal eigenvalue perturbation analysis covering ERA and OPIA using an appropriate predictor matrix is always possible. A report about this subject is in preparation and should appear in a future contribution.

4 Numerical Examples

In this section we present the results of some computer simulations which illustrate the behavior of the bounds (3.23) and the role of $\|W^\dagger\|$. We consider two numerical examples. The first is extracted from the specialized literature of the signal analysis field, and the second is a signal synthesized from experimental measurements. For each example, we compute $\|W^\dagger\|$ and the bounds (3.23), $(1+\|f\|^2)^{-1/2}$ and $(1+\|b\|^2)^{1/2}$, for several values of N.

4.1 Bounds for the Signal Eigenvalues of a Synthetic Signal

This example illustrates the bounds associated with the sampled signal defined by

$$h_k = e^{(-0.01+2\pi 0.20\imath)k} + e^{(-0.02+2\pi 0.22\imath)k}, \quad k = 0, 1, \ldots,$$

whose signal eigenvalues are $z_1 = e^{(-0.01+2\pi 0.20\imath)}$ and $z_2 = e^{(-0.02+2\pi 0.22\imath)}$. In this case, we have that $|z_1| = 0.9900$, $|z_2| = 0.9802$ and the upper bound limit is $\prod_{j=1}^{n} |z_j|^{-1} = 1.035$ (note that $n = 2$). This signal is often used for testing the capability of algorithms to extract frequencies and decay factors from noisy signals, because the two closely spaced frequencies are easily seen as a single one when additive noise is present (see, *e.g*, [13]). The behavior of the bounds as functions of N is displayed in Figure 1-(b). In Figure 1-(a) we show the behavior of $\|W^\dagger\|$ on a logarithmic scale. From these figures we see the marked monotonic decrease of both $\|W^\dagger\|$ and the width of the annulus $\mathcal{A}_N$ for increasing values of N. In this example, we also verify that for $N \geq 60$, our bounds agree with their limiting values 1 and 1.035 up to two decimal places.

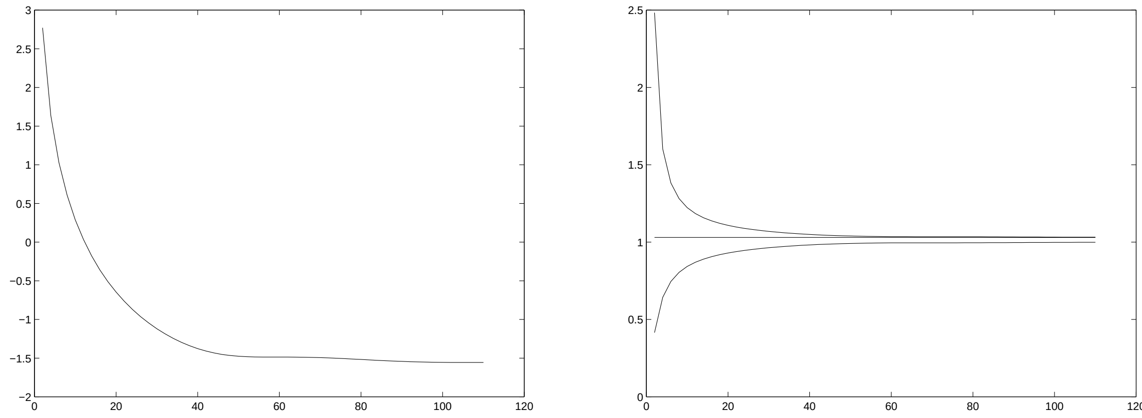


Figure 1: (a) $\log(\|W^\dagger\|)$ as function of N. (b) Bounds 3.23 as function of N

4.2 Bounds for Signal Eigenvalues related to a Mechanical System

In this example we illustrate the behavior of the bounds (3.23) for a synthesized signal obtained from experimental measurements of the free response of a real vibratory system. Full information about the procedure used to synthesize the signal can be found in [3]. For this case, the signal eigenvalues come in complex conjugate pairs and $n = 10$. These eigenvalues are shown in Table 1.

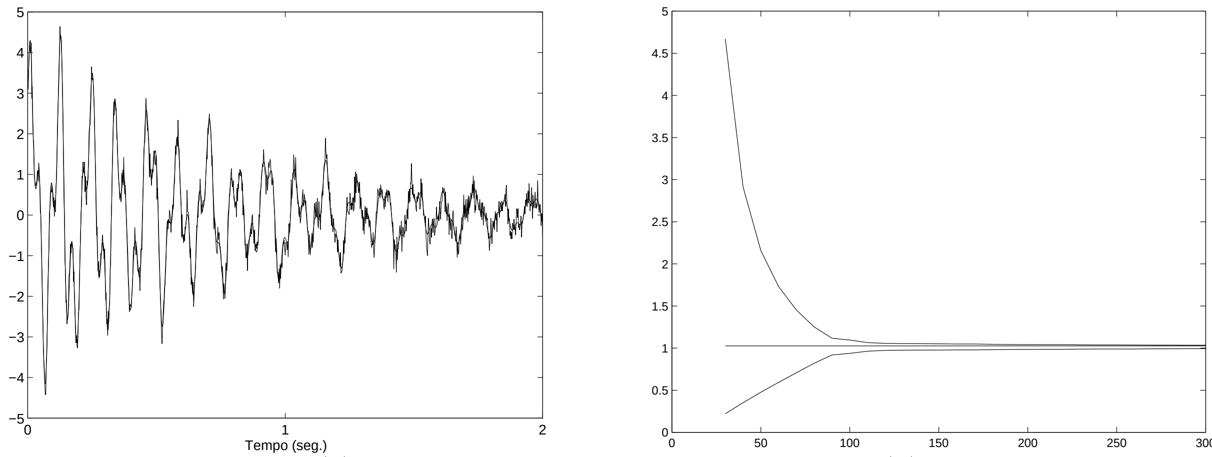
The behavior of the upper and lower bounds (3.23) as functions of N, which we denote here by

j	z_j	$\lvert z_j\rvert$	$\lvert z_j\rvert^{-1}$
1	$0.9699 \pm 0.2248\imath$	0.9956	1.0044
2	$0.9532 \pm 0.2931\imath$	0.9972	1.0028
3	$0.9844 \pm 0.1619\imath$	0.9976	1.0024
4	$0.9921 \pm 0.1055\imath$	0.9977	1.0023
5	$0.9972 \pm 0.0585\imath$	0.9989	1.0011

Table 1: Signal Eigenvalues of synthesized signal and corresponding moduli.

L_N and U_N respectively, is displayed for $N \geq 30$ in Figure 2-(b). The rapid decrease of the width of the annulus $\mathcal{A}_N$ is again very apparent.

Notice that, because $|z_j| \approx 1$, the signal is slightly damped (see Figure 2-(a)). In order to better illustrate the behavior of the bounds as functions of N, we have computed their distances to their corresponding limits, $L_\infty = 1$, and $U_\infty = \prod_j^{10} |z_j|^{-1} = 1.0262$, respectively. These distances as well as the norms $\|W^\dagger\|$ are shown in Table 2 for certain values of N. This table also illustrates the decrease of $\|W^\dagger\|$ as the effect of increasing N. We also note that the bounds agree well with their limits for $N \geq 200$.



(a) (b)

Figure 2: (a) Signal related to a mechanical system. (b) Bounds (3.23) as function of N

N	10	20	30	40	50	60
U_N	415.0088	11.7560	4.6708	2.9155	2.1564	1.7303
$U_N - U_\infty$	413.9826	10.7298	3.6445	1.8893	1.1302	0.7041
L_N	0.0025	0.0873	0.2197	0.3520	0.4759	0.5931
$L_\infty - L_N$	0.9975	0.9127	0.7803	0.6480	0.5241	0.4069
$\|W^\dagger\|$	5.4548×10^9	0.0012×10^9	1.8228×10^4	0.0974×10^4	0.0096×10^4	0.0014×10^4

N	200	220	240	260	280	300
U_N	1.0411	1.0410	1.0379	1.0367	1.0352	1.0338
$U_N - U_\infty$	0.0149	0.0148	0.0117	0.0105	0.0090	0.0076
L_N	0.9857	0.9858	0.9888	0.9899	0.9913	0.9926
$L_\infty - L_N$	0.0143	0.0142	0.0112	0.0101	0.0087	0.0074
$\|W^\dagger\|$	0.1127	0.1104	0.1095	0.1055	0.1044	0.1033

Table 2: Behavior of bounds 3.23 and $\|W^\dagger\|$ as functions of N.

5 Conclusions and perspectives

We have developed a singular value analysis of certain predictor matrices that enabled us to derive a closed form for their singular spectrum. Using these results, we have derived lower and upper bounds for the so-called *signal eigenvalues*, both depending of the dimension of the problem. By analyzing the influence of the dimension on these bounds, we have shown that they can become very tight provided the dimension of the problem is sufficiently large and the signal is slightly damped. This was illustrated with numerical examples including the analysis of the bounds for a signal related to a vibrating structure.

We may anticipate interesting applications of our results to certain subspace-approaches for modal parameter identification problems such as ERA, Kung's method, and OPIA, among others.

In particular, they seem to open a way for a unified signal eigenvalue perturbation analysis covering these methods, provided they can be shown to depend on specific predictor matrices obtained by projection. This challenging development is the object of ongoing research.

Appendix

A Proof of Lemma 5

Notice that b is the minimum norm solution of (2.11) and therefore $b \in \mathcal{R}(H(\ell)^*)$. From this

$$G_{\mathcal{Q}} = V_2^* \, B \, V_2 = [0, \, V_2^* e_1, \ldots, V_2^* e_{N-1}] \, V_2,$$

can be rewritten as

$$G_{\mathcal{Q}} = V_2^* \, \mathcal{J} \, V_2 - uv^*, \tag{A.1}$$

where $\mathcal{J}$ is the permutation matrix $\mathcal{J} = [e_N, e_1, \ldots, e_{N-1}]$, and u^* and v^* are respectively, the last and first row of V_2. The proof of the theorem is then based on the computation of the eigenvalues of $G_{\mathcal{Q}}^* G_{\mathcal{Q}}$. We start then by observing that

$$G_{\mathcal{Q}}^* G_{\mathcal{Q}} = G^* G - G^* uv^* - vu^* G + (u^* u) vv^*, \tag{A.2}$$

where we set $G = V_2^* \mathcal{J} V_2$. Analyzing the first term of the right-hand side we see that

$$G^* G = V_2^* \mathcal{J}^* V_2 V_2^* \mathcal{J} V_2 = V_2^* \mathcal{J}^* (I - V_1 V_1^*) \mathcal{J} V_2 = I - V_2^* \mathcal{J}^* V_1 V_1^* \mathcal{J} V_2, \tag{A.3}$$

where the last equality follows from the the orthogonality of $\mathcal{J}$. On the other hand, observe that $\mathcal{J}^* V_1$ can be rewritten as,

$$\mathcal{J}^* V_1 = \begin{bmatrix} w^* \\ V_1^{\uparrow} \end{bmatrix} = \begin{bmatrix} b^* V_1 \\ V_1^{\uparrow} \end{bmatrix} + \begin{bmatrix} w^* - b^* V_1 \\ 0 \end{bmatrix} = A^* + \begin{bmatrix} w^* - b^* V_1 \\ 0 \end{bmatrix}$$

where $V_1^{\uparrow}$ is the matrix of V_1 consisting of all rows excluding the last, w^* is the last row of V_1, and $A^* = BV_1$. Hence, we have that

$$V_2^* \mathcal{J}^* V_1 = V_2^* \begin{bmatrix} w^* - b^* V_1 \\ 0 \end{bmatrix} = V_2^* e_1 (w^* - b^* V_1) = v(w^* - b^* V_1), \tag{A.4}$$

since $V_2^* A^* = 0$ by (3.22). Substituting this relation in (A.3) yields

$$G^* G = I - (\|w\|^2 - b^* e_N - e_N^* b + \|b\|^2) vv^*. \tag{A.5}$$

We now analyze the second term of the right-hand side of (A.2). Notice that by using $u = V_2^* e_N$ and $v^* = e_1^* V_2$, we have,

$$G^* u = V_2^* \mathcal{J}^* V_2 V_2^* e_N = V_2^* \mathcal{J}^* (I - V_1 V_1^*) e_N = V_2^* \mathcal{J}^* e_N - V_2^* \mathcal{J}^* V_1 V_1^* e_N.$$

But, if one observes that $\mathcal{J}^* e_N = e_1$ and $V_1^* e_N = w$, we have, using (A.4), that

$$G^* u = v - v(w^* - b^* V_1) w = v - v(w^* w - b^* V_1 V_1^* e_N) = v - v(w^* w - b^* e_N),$$

since $b \in \mathcal{R}(H(\ell)^*)$, and hence,

$$G^* u v^* = (1 - \|w\|^2 + b^* e_N) v v^*. \tag{A.6}$$

This, in turn, implies that the third term of the right-hand side of (A.2) is

$$v u^* G = (1 - \|w\|^2 + e_N^* b) v v^*. \tag{A.7}$$

Replacing now (A.7), (A.6) and (A.5) into (A.2) and taking into account that $\|w\|^2 + \|u\|^2 = 1$, because $[w^* u^*]$ is the last row of the orthogonal matrix $[V_1\ V_2]$, we deduce that

$$G_{\mathcal{Q}}^* G_{\mathcal{Q}} = I - (1 + \|b\|^2) v v^*.$$

From this relation, we see that $N - n - 1$ eigenvalues of $G_{\mathcal{Q}}^* G_{\mathcal{Q}}$ are equal to the unity, while the remaining one is $1 - (1 + \|b\|^2)\|v\|^2$. The proof concludes by noting that $\|q_1\| = \|\mathcal{Q} e_1\| = \|V_2 V_2^* e_1\| = \|V_2 v\| = \|v\|$. □

B Proof of Lemma 6

We first derive auxiliary results involving the terms of the ratio

$$\frac{1 + \|b\|^2}{1 + \|f\|^2}$$

as functions of N. For this, we consider two consecutive values of N, and use the subscript $_{[N]}$ to denote the dependence of the considered quantities on N. We start by observing that

$$W_{[N+1]}^{\dagger *} W_{[N+1]}^{\dagger} = (W_{[N+1]} W_{[N+1]}^*)^{-1} \doteq A_{[N+1]} \tag{B.8}$$

and that the $W_{[N+1]} W_{[N+1]}^* = A_{[N+1]}^{-1}$ is a rank-one modification of $A_{[N]}^{-1} = W_{[N]} W_{[N]}^*$, i.e.

$$A_{[N+1]}^{-1} = A_{[N]}^{-1} + Z^N e e^* Z^{N*} = Z A_{[N]}^{-1} Z^* + e e^*,$$

where we used the two representations $W_{[N+1]} = [W_{[N]} \;\; Z^N e] = [e \;\; ZW_{[N]}]$. Applying the Sherman-Morrison formula to each of these forms, we derive that

$$A_{[N+1]} = A_{[N]} - \frac{A_{[N]} Z^N e e^* Z^{N*} A_{[N]}}{1 + e^* Z^{N*} A_{[N]} Z^N e} = Z^{-*} A_{[N]} Z^{-1} - \frac{Z^{-*} A_{[N]} Z^{-1} e e^* Z^{-*} A_{[N]} Z^{-1}}{1 + e^* Z^{-*} A_{[N]} Z^{-1} e}. \tag{B.9}$$

Now, the projector associated with the value N is given by

$$\mathcal{P}_{[N]} = W_{[N]}^{\dagger} W_{[N]} = W_{[N]}^{\dagger} [e \;\; ZW_{[N-1]}] = [W_{[N]}^{\dagger} e \;\; W_{[N]}^{\dagger} Z W_{[N-1]}]$$

because of (2.5) and the definition of $W_{[N]}$. This yields $\|p_{1,[N]}\| = \|W_{[N]}^{\dagger} e\|$. Combining this with (B.8), we obtain $\|p_{1,[N+1]}\|^2 = e^* A_{[N+1]} e$. Using this relation in the first equality of (B.9), we obtain

$$\begin{aligned} \|p_{1,[N+1]}\|^2 &= e^* A_{[N]} e - \frac{e^* A_{[N]} Z^N e e^* Z^{N*} A_{[N]} e}{1 + e^* Z^{N*} A_{[N]} Z^N e} \\ &= \|p_{1,[N]}\|^2 - \frac{(e_1^* W_{[N]}^{\dagger} Z^N e)(e^* Z^{N*} W_{[N]}^{\dagger *} e_1)}{1 + e^* Z^{N*} A_{[N]} Z^N e} \\ &= \|p_{1,[N]}\|^2 - \frac{|e_1^* f_{[N]}|^2}{1 + \|f_{[N]}\|^2}, \end{aligned} \tag{B.10}$$

where $f_{[N]} = W_{[N]}^{\dagger} Z^N e$ is the minimum norm solution of the system (2.10). On the other hand, using the equality between the left-hand side and the last right-hand side of (B.9), we have that

$$e^* A_{[N+1]} e = e^* Z^{-*} A_{[N]} Z^{-1} e - \frac{e^* Z^{-*} A_{[N]} Z^{-1} e e^* Z^{-*} A_{[N]} Z^{-1} e}{1 + e^* Z^{-*} A_{[N]} Z^{-1} e}.$$

This is nothing but

$$\|p_{1,[N+1]}\|^2 = \|b_{[N]}\|^2 - \frac{\|b_{[N]}\|^4}{1 + \|b_{[N]}\|^2} = \frac{\|b_{[N]}\|^2}{1 + \|b_{[N]}\|^2},$$

where $b_{[N]} = W_{[N]}^{\dagger} Z^{-1} e$. This implies that

$$1 + \|b_{[N]}\|^2 = \frac{1}{1 - \|p_{1,[N+1]}\|^2}. \tag{B.11}$$

We now observe that, using the fact that $p_{1,[N]} = e_1 - q_{1,[N]}$, (B.10) can be rewritten as

$$1 - \|p_{1,[N+1]}\|^2 = \|q_{1,[N]}\|^2 + \frac{|e_1^* f_{[N]}|^2}{1 + \|f_{[N]}\|^2},$$

which, combined with (B.11), gives

$$1 = (1 + \|b_{[N]}\|^2)\|q_{1,[N]}\|^2 + \frac{|e_1^* f_{[N]}|^2}{1 + \|f_{[N]}\|^2}(1 + \|b_{[N]}\|^2),$$

or equivalently, by Theorem 5,

$$\sigma^2_{N-n,[N]}(G_Q) = |e_1^* f_{[N]}|^2 \frac{1 + \|b_{[N]}\|^2}{1 + \|f_{[N]}\|^2}. \tag{B.12}$$

The final part of our proof depends on two important observations. The first is that, as $e_1^* f_{[N]}$, is the independent term of the characteristic polynomial of F, that is non zero because of Theorem 2, it is equal to the product of the eigenvalues of F. That is,

$$|e_1^* f_{[N]}| = \prod_{k=1}^{N-n} |\widehat{\lambda}_k| \prod_{j=1}^{n} |z_j|, \tag{B.13}$$

where the $\widehat{\lambda}$'s are the extraneous eigenvalues of F, and the z's are the signal eigenvalues. The second is that the extraneous eigenvalues of B are the conjugate of those of F, as proved in [2], Theorem 3.2. Hence the product of their modulus is equal to the product of the singular values of $V_2^* B V_2 = G_Q$. Using now Theorem 5, we deduce that

$$\sigma^2_{N-r,[N]}(G_Q) = \prod_{k=1}^{N-n} |\widehat{\lambda}_k|^2. \tag{B.14}$$

The first part of the theorem then follows from (B.14), (B.13) and (B.12).

Now, observe that, using the second equality of (B.9) and the definitions of f and b,

$$\begin{aligned}
\|f_{[N+1]}\|^2 &= e^* Z^{N+1*} A_{[N+1]} Z^{N+1} e \\
&= e^* Z^{N*} A_{[N]} Z^N e - \frac{e^* Z^{N*} A_{[N]} Z^{-1} e e^* Z^{-*} A_{[N]} Z^N e}{1 + e^* Z^{-*} A_{[N]} Z^{-1} e} \\
&= \|f_{[N]}\|^2 - \frac{|f_{[N]}^* b_{[N]}|^2}{1 + \|b_{[N]}\|^2},
\end{aligned}$$

which shows that $\|f\|$ decreases monotonically with N. The same conclusion then follows for $\|b\|$ because of the first part of the theorem. □

References

[1] F. S. V. Bazán and C. Bavastri. An optimized pseudo-inverse algorithm (OPIA) for multi-input multi-output modal parameter identification. *Mechanical Systems and Signal Processings*, 10:365–380, 1996.

[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, Ph. L. Toint, and M. C. Zambaldi. A conjugate-gradients based method for harmonic retrieval problems that does not use explicit signal subspace computation. Technical Report 16, Department of Mathematics, FUNDP, Namur, Belgium, November 1997.

[4] L. H. Bezerra and F. S. V. Bazán. Eigenvalue location of generalized companion predictor matrices. *SIAM Journal on Matrix Analysis and Applications*, 19(4):886–897, 1998.

[5] Å. Björk. *Numerical Methods for least squares problems.* SIAM, Philadelphia, USA, 1996.

[6] G. E. Box and G. M. Jenkins. *Time Series Forecasting and Control.* Holden-Day, San Francisco, 1970.

[7] S. Braun and Y. Ram. Determination of structural modes via the prony model: System order and noise induced poles. *Acoust. Soc. Am.*, 81(5):1447–1495, 1987.

[8] S. Braun and Y. Ram. Time and frequency identification methods in over-determined systems. *Mechanical Systems and Signal Processings*, 1(3):245–257, 1987.

[9] G. Cybenko. Restrictions of normal operators, Padé approximation and autoregressive time series. *SIAM Math. Anal.*, 4:753–767, July 1984.

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

[11] F. B. Hildebrand. *Introduction to Numerical Analysis.* McGraw-Hill, New York, 1956.

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

[13] Y. Hua and T. P. Sarkar. Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise. *IEEE Trans. On Acoust. Speech and Signal Processings*, ASSP-38(5):814–824, May 1990.

[14] J. J. Juang and R. S. Pappa. An eigensystem realization algorithm for modal parameter identification and modal reduction. *Journal of Guidance, Control and Dynamics*, 8:620–627, 1985.

[15] Fuadd Kittaneh. Singular values of companion matrices and bouds on zeros of polynomials. *SIAM Journal on Matrix Analysis and Applications*, 16(1):333–340, January 1995.

[16] R. Kumaresan. On the zeros of the linear prediction-error filters for deterministic signals. *IEEE Trans. On Acoust. Speech and Signal Processing*, ASSP-31(1):217–221, February 1983.

[17] R. Kumaresan and D. W. Tufts. Estimating the parameters of exponentially damped sinudoids and pole zero-modeling in noise. *IEEE Trans. On Acoust. Speech and Signal Processing*, ASSP-30:833–840, December 1982.

[18] S. Kung. A new identification and model reduction algorithm via singular value decompositions. In *Proceedings of the* 12^{th} *Asilomar Conference on Circuits Systems and Computers*, pages 705–714, Pacific Grove California, 1978.

[19] P.R. G. Kurka. Modal parameter identification in the time domain from a random force. *Mechanical Systems and Signal Processing*, 4(5):393–404, 1990.

[20] G. Majda, W. A. Strauss, and M. Wei. Computation os exponentials in transient data. *IEEE Trans. On Antennas and Propagation*, 37(10):1035–1043, 1989.

[21] J. Makhoul. Linear prediction: A tutorial review. *Proceedings of the IEEE*, 63(4):561–580, 1975.

[22] M. A. Rahman and K. B. Yu. Total least squares approach for frequency estimation using linear prediction. *IEEE Trans. Acoust. Speech Signal process.*, ASSP-35:1440–1454, 1987.

[23] A. Van Der Veen, E. F. Deprettere, and A. Lee Swindlehurst. Subspace-based signal analysis using singular value decomposition. *Proceedings of the IEEE*, 81(9):1277–1309, September 1993.

[24] S. Van Huffel, H. Chen, C. Decanniere, and P. Van Hecke. Algorithm for time-domain NMR data fitting based on total least squares. *Journal of Magnetic Resonance A*, 110:228–237, 1994.

[25] M. Wei and G. Majda. A new theoretical approach for Prony's method. *Linear Algebra and its Applications*, 136:119–132, 1990.