A Conjugate-Gradients Based Method for Harmonic Retrieval Problems that Does Not Use Explicit Signal Subspace Computation

by F.S.V. Bazán[1], Ph.L. Toint[2] and M.C. Zambaldi[3]

Report 97/16 November 24, 1997

[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
Current reports available by anonymous ftp from the directory
"pub/reports" on thales.math.fundp.ac.be

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

A Conjugate-Gradients Based Method for Harmonic Retrieval Problems that Does Not Use Explicit Signal Subspace Computation

F. S. V. Bazán   Ph. L. Toint   M. C. Zambaldi

November 24, 1997

Abstract

The harmonic retrieval problem consists of estimating frequencies and decay factors of multiple exponential signals from experimental measurements. At variance with most available algorithms, which are based on the singular value decomposition (SVD) and the explicit identification of the so-called "signal subspace", we present an alternative method in which this explicit identification is not needed. We propose to replace it by an *implicit* estimation using the concept of predictor matrices and algorithmically realized using conjugate gradients. We finally discuss some numerical examples, both synthetic and from a real application, showing that the new method produces results of an accuracy comparable to that of SVD-based subspace techniques, but without requiring a priori knowledge of the dimension of the signal subspace and at a lower computational cost.

1 Introduction

We consider the problem of determining frequencies and corresponding decay factors of multiple exponential signals from observed data. This problem, referred as the *harmonic retrieval problem* (HRP), arises in a variety of areas such as biology, physics, or engineering. It has been the subject of much research for several years (see, for instance, Mayrargue and Jouveau, 1988, Pisarenko, 1973, Kung, Arun and Bhaskar Rao, 1983, Makhoul, 1975 and Van Der Veen, Deprettere and Lee Swindlehurst, 1993). If the signal under observation is a sum of n complex exponentials of the form

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

where r_j and s_j are complex constants such that $s_j \neq s_k$ for $j \neq k$, $s_j = \alpha_j + i\omega_j$, $i = \sqrt{-1}$ and $\alpha_j \leq 0$, then the HRP consists in retrieving the frequencies ω_j and corresponding decay factors α_j from a finite set of data $\{h_k\}_{k=0}^{L}$. In this formulation Δt is called the sampling interval and it is chosen so that $\max |\omega_j| \, \Delta t < \pi$. When the observed data comes from the impulse response of some time invariant physical system, the HRP is closely related to the system realization problem of constructing a state-space model of minimal order having an impulse response function that best match a set of noise corrupted observations. In this latter case, n is the order of the system, the s_j's are constants called *poles* which say much about the normal modes of the system and the r_j's, called *residues*, describe how much each mode participates in the model. The realization problem then consists in determining the above parameters (see De Moor, 1993, Van Der Veen et al., 1993 and Zeiger and Mac-Ewen, 1974).

As in system realization theory, we collect the observed data in a Hankel matrix having the so-called shift-invariance property. Let $H(\ell)$ be an $M \times N$ Hankel matrix indexed by ℓ with column vectors of the form $\hat{h}_\ell = [h_\ell \, h_{\ell+1} \, \ldots, h_{\ell+M-1}]^T$, $\ell \geq 0$, that is,

$$H(\ell) = \begin{bmatrix} h_\ell & h_{\ell+1} & \cdots & h_{\ell+N-1} \\ h_{\ell+1} & h_{\ell+2} & \cdots & h_{\ell+N} \\ \vdots & \vdots & \cdots & \vdots \\ h_{\ell+M-1} & h_{\ell+M} & \cdots & h_{\ell+M+N-2} \end{bmatrix} = [\hat{h}_\ell, \ldots, \hat{h}_{\ell+N-1}].$$

We assume here that $M \geq N > n$. Shift-invariance means that the column space of the shifted Hankel matrix $H(\ell+1)$ is contained in the column space of the original one, see Van Der Veen et al. (1993). We may write $H(\ell)$ in factored form as

$$H(\ell) = VZ^\ell RW, \tag{1.1}$$

where $Z = \mathrm{diag}(z_1, \ldots, z_n)$, $z_j = e^{s_j \Delta t}$, $R = \mathrm{diag}(r_1, \ldots, r_n)$, $V = V(z_1, z_2, \ldots, z_n)$ is an $M \times n$ Vandermonde matrix whose k-th row is the row vector $[z_1^{k-1} \, z_2^{k-1} \cdots z_n^{k-1}]$ and W the transpose of the matrix that consists of the N first rows of V. Then, for all $\ell \geq 0$, $\mathrm{rank}[H(\ell)] = n$ and the column space of $H(\ell)$, denoted by $\mathcal{R}(H(\ell))$, is spanned by the column vectors of matrix V. For future reference, note that our definitions imply the identity

$$We_j = Z^{j-1} e, \tag{1.2}$$

where e_j is the j-th vector of the canonical basis and e is the vector of all ones.

On the other hand, let $V^\uparrow$ the matrix consisting of all rows of V excluding the first and $V^\downarrow$ the matrix of V consisting of all rows excluding the last. Then the form of V ensures that $V^\uparrow = V^\downarrow Z$. Furthermore, the shift invariance-property of the Hankel matrix ensures that, if $\hat{V}$ is any matrix whose columns span $\mathcal{R}(H(\ell))$ and if $\hat{V}^\uparrow$ and $\hat{V}^\downarrow$ are constructed from $\hat{V}$ as $V^\uparrow$ and $V^\downarrow$ were constructed from V, then the information we require to solve the HRP can always be extracted from the eigenvalues of an $n \times n$ matrix T, say, known as a *transition matrix*, which satisfies

$$\hat{V}^\uparrow = \hat{V}^\downarrow T. \tag{1.3}$$

It is well-known that the eigenvalues of T, called from here on *signal eigenvalues*, are $z_i = e^{s_i \Delta t}$, $i = 1, 2 \ldots, n$. The reader is referred to Kung et al. (1983), Li and Lu (1994) and Van Der Veen et al. (1993) for further discussion of this introductory material.

The solution of the noiseless HRP is thus relatively simple: the number of frequencies n is equal to the rank of $H(\ell)$ and the frequencies themselves, together with the associated decay factors, are computed from the eigenvalues of any transition matrix (1.3). The matrix $\hat{V}$ is usually chosen as the matrix whose columns are the right or left singular vectors associated with the non-zero singular values of the Hankel matrix $H(\ell)$. However, in all practical applications the observed signal is a corrupted version of the signal that we would expect under ideal circumstances. As a consequence, the number of frequencies, n, must now be estimated from a perturbed data matrix $\tilde{H}(\ell) = [\tilde{h}_\ell, \dots, \tilde{h}_{\ell+N-1}]$, which is in general of full rank. The strategy is then to look for a break in the pattern of singular values of $\tilde{H}(\ell)$, attributing the larger ones to the signal and the smaller ones to the noise. In other words, we divide $\mathcal{R}(\tilde{H}(\ell))$ into two subspaces: a dominant subspace spanned by the singular vectors associated with the largest singular values, referred to as the *signal subspace*, and the subspace associated with the remaining singular values, called the *noise subspace*. The detection of such a break has been well studied and a number of techniques based on estimates of the signal subspace have been developed: see, for instance, Bazán and Bavastri (1996), De Moor (1993) and Van Der Veen et al. (1993).

In certain cases however, the noise and signal singular values are not obviously separated and a clear break is not evident. The estimation of the signal subspace may then be imprecise and the SVD-based methods which use the position of the break tend to fail, either because non desired frequencies are introduced or because information may be lost if the signal subspace dimension is underestimated. For this reason, methods for solving the HRP that avoid the explicit determination of the rank of $H(\ell)$ are desirable. It is the object of the present paper to propose such a method. The main idea is to explore the shift invariance property of the Hankel matrix from another point of view: we will reconstruct the frequencies and decay factors from the eigenvalues of predictor matrices. If the data is noise-free, these matrices are defined as follows. We say that the $N \times N$ matrix F is a *forward predictor matrix* if, for all $\ell \geq 0$,

$$H(\ell+1) = H(\ell)F \tag{1.4}$$

(see Bazán and Bezerra (1997)). Similarly, B is a *backward predictor matrix* if it satisfies

$$H(\ell-1) = H(\ell)B \tag{1.5}$$

for all $\ell \geq 1$. We shall show how to solve the HRP from well chosen predictor matrices which are constructed from the solution of a linear least squares problem. In our algorithmic proposal, this system is solved by using conjugate gradient iterations. The choice of conjugate gradients is motivated by the observation that the coefficient matrix for this system is the rank deficient Hankel matrix $H(\ell)$; the solution is thus obtained in at most n iterations in the noiseless case (see Bjork, 1996, Hestenes, 1975, or Trefethen and Bau, 1997). We will also verify both theoretically and in practice that the new method is computationally cheaper than comparable SVD-based methods.

The paper is structured as follows. In Section 2, we present some SVD-subspace approaches including results on singular values estimates for the involved Hankel matrices. Section 3 deals with mathematical properties of predictor matrices and includes an error analysis for their eigenvalues, motivated by the fact that frequencies and decay dactors are extracted from these eigenvalues. The proposed algorithm is presented in Section 4. In Section 5, we present a detailed discussion of the stopping rule for the conjugate-gradients least-squares solver, which is aimed at avoiding the presence of undesirable high frequencies in the computed solution. A brief analysis of the computational cost associated with HTLS, a SVD-based method, and our new algorithm is discussed in Section 6. This choice is supported by the excellent reputation of this method amongst practitioners (see Van Huffel, Chen, Decanniere and Van Hecke (1994), for instance) and by the acknowledged value of using a total least-squares approach (see Fierro and Bunch, 1996, and Fierro, Golub, Hansen and O'Leary, 1997). In order to illustrate the effectiveness of our approach, we finally present some preliminary numerical experiments in Section 7, including a comparison with HTLS.

2 Some subspace approaches using the SVD

We start by discussing ways to exploit the singular value decomposition to take advantage of the shift invariance property of the Hankel matrices. The purpose of this discussion is to introduce a result that motivates choosing these (square) matrices as large as possible, in order to enhance the gap between the noise related and signal related singular values. This result will subsequently be commented on in our numerical experiments.

A general subspace approach may be derived as follows. Consider the $M \times N$ Hankel matrix $H(\ell)$ and its shifted value $H(\ell+1)$. From (1.1) we see that $H(\ell)$ and $H(\ell+1)$ can be factorized as

$$H(\ell) = \overline{V}\,\overline{W}, \quad H(\ell+1) = \overline{V}\,Z\,\overline{W},$$

where $\overline{V} = VZ^\ell$ and $\overline{W} = RW$. Since the above factorizations are full-rank, we have that

$$Z = \overline{V}^\dagger H(\ell+1)\overline{W}^\dagger,$$

where † stands for Moore-Penrose pseudo-inverse. If we now consider any other full-rank factorization of $H(\ell)$ of the form $H(\ell) = FG$, we obtain that the matrix

$$T = F^\dagger H(\ell+1)G^\dagger = F^\dagger\overline{V}Z\overline{W}G^\dagger \tag{2.1}$$

is similar to Z because $FG = \overline{V}\,\overline{W}$ (see De Groen and De Moor, 1987). Depending on how we choose F and G, we may then derive different methods for computing signal eigenvalues. In what follows, we present two approaches based on the SVD of the data matrix. In fact, this SVD may be written as

$$H(\ell) = P\Sigma Q^* = [P_s\, P_n] \begin{bmatrix} \Sigma_s & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} Q_s^* \\ Q_n^* \end{bmatrix} = P_s \Sigma_s Q_s^*, \tag{2.2}$$

where P_s are the singular vectors associated with the non-zero singular values, Σ_s is an $n \times n$ diagonal matrix containing the singular values $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_n > 0$, and * denotes the conjugate transpose. The column vectors in P_s span the signal subspace while the column vectors of P_n span the null space of $H(\ell)^*$, which is the noise subspace. Similarly, the columns of Q_s span the row space of $H(\ell)$, called the *short signal subspace* (De Moor, 1993), while columns of Q_n span $\mathcal{N}(H(\ell))$, the nullspace of $H(\ell)$.

The first approach, due Zeiger and Mac-Ewen (1974) and well-known in the realization theory context as the *balanced realization approach*, chooses

$$F = P_s\Sigma_s^{1/2} \text{ and } G = \Sigma_s^{1/2}Q_s,$$

in which case (2.1) becomes

$$T_{\mathrm{ZM}} = \Sigma_s^{-1/2} P_s^* H(\ell+1) Q_s \Sigma_s^{-1/2}. \tag{2.3}$$

A second approach, called OPIA (Optimized Pseudo-Inverse Algorithm), can easily be obtained from (2.3): observe that all but the last column of $H(\ell+1)$ are identical to the last $N-1$ columns of $H(l)$, and therefore that $H(\ell+1) = H(\ell)\mathcal{C}$, where $\mathcal{C}$ is a companion matrix (see Bazán and Bavastri, 1996). The desired transition matrix is then computed using this relation and (2.2) in (2.3), which yields, after some manipulations, the transition matrix

$$T_{\mathrm{BB}} = Q_s^* \mathcal{C} Q_s.$$

This $n \times n$ matrix is similar to T_{ZM}. OPIA was introduced for solving modal parameter identification problems.

We continue our outline of subspace approaches for solving the HRP by presenting two other methods derived from (1.3). If we let $V^{\uparrow}$ be the matrix consisting of all rows of Q_s but the first, and let $V^{\downarrow}$ be the matrix consisting of all rows of Q_s excluding the last, we may then solve (1.3) for T in two different ways. If it is solved in the standard least squares sense, then one obtains the transition matrix,

$$T_{\mathrm{K}} = (V^{\downarrow})^{\dagger} V^{\downarrow}. \tag{2.4}$$

This method was proposed by Kung (1978), and it is known as the HSVD method in the field of NMR (Nuclear Magnetic Resonance) (see, for instance, Chen, Van Huffel, Dowling and de Groat, 1997, or Van Huffel et al., 1994). Kung also shows that the pseudo-inverse in (2.4) need not be computed by matrix inversion. If equation (1.3) is solved in the total least squares sense, then one obtains an algorithm called HTLS, in which signal eigenvalues are extracted from

$$T_{\mathrm{VH}} = -\hat{Q}_{12}(\hat{Q}_{22})^{-1}, \tag{2.5}$$

where $\hat{Q}_{12}$ and $\hat{Q}_{22}$ are $n \times n$ matrices obtained from the SVD of the augmented data matrix

$$[V^{\uparrow} V^{\downarrow}] = \hat{P}\hat{\Sigma}\hat{Q}^*, \quad \text{with} \quad \hat{Q} = \begin{bmatrix} \hat{Q}_{11} & \hat{Q}_{12} \\ \hat{Q}_{21} & \hat{Q}_{22} \end{bmatrix}.$$

HTLS was recently introduced by S. Van Huffel and it is currently used in time-domain analysis of NMR data (see again Chen et al., 1997, or Van Huffel et al., 1994).

The above methods were developed under the assumption that the signal subspace is available. Indeed, if the data is free of noise, calculating this subspace from the SVD of the Hankel matrix is simple: it suffices to choose the subspace spanned by the right or left singular vectors associated with the non-zero singular values of the data matrix. The problem is considerably more difficult when the data is perturbed by noise, that is if

$$\tilde{h}_k = h_k + \eta_k, \quad k = 0, \ldots, L,$$

where η_k represents the noise on the k-th observation. An estimate of the signal space can then be obtained from the SVD of the perturbed Hankel matrix $\tilde{H}(\ell) = H(\ell) + E$, where E is the

Hankel matrix corresponding to the noise, provided a clear gap can be identified within the singular spectrum of $\tilde{H}(\ell)$. This will typically be the case when the smallest nonzero singular value of the unperturbed Hankel matrix, $\sigma_n(H(\ell))$, is large compared to the (spectral) norm of the perturbation E (see De Moor, 1993, Stewart, 1991 or Van Der Veen et al., 1993). Therefore, it is crucial to examine in which cases $\sigma_n(H(\ell))$ is sufficiently large, in particular as a function of MN, the dimension of $H(\ell)$. This is the object of the next theorem, where we have chosen, for notational convenience, M and N to be multiples of the number of frequencies, n, i.e.,

$$M = p \times n \text{ and } N = q \times n \quad (p, q > 1, p, q \in \mathbf{N}).$$

In this case, the $M \times N$ Hankel matrix is denoted by $H_{p,q}(\ell)$.

Theorem 1 *Let V_o be the $n \times n$ Vandermonde matrix whose i, j-th entry is z_i^{j-1}. Also define*

$$\rho = \min_{j=1,\ldots,n} |r_j| \quad \textit{and} \quad \beta = \min_{j=1,\ldots,n} |z_j|.$$

Then for any pair p, q of strictly positive integers,

$$\sigma_n(H_{p,q}(\ell)) \geq \rho\beta^\ell[\sigma_n(V_o)]^2 f(p,q), \tag{2.6}$$

where

$$f(p,q) = \sqrt{1 + \beta^{2n} + \beta^{2n(2)} + \cdots + \beta^{2n(p-1)}}\sqrt{1 + \beta^{2n} + \beta^{2n(2)} + \cdots + \beta^{2n(q-1)}}. \tag{2.7}$$

Moreover, if m is even and $p + q = m$, the above lower bound is maximized as function of p and q for $p = q = m/2$. In this case, one has that

$$\sigma_n(H(\ell)) \geq \rho\beta^\ell[\sigma(V_o)]^2 \frac{(1-\beta^{mn})}{(1-\beta^{2n})}. \tag{2.8}$$

Proof. If we define $D = Z^n = diag(z_1^n, \ldots, z_n^n)$ and

$$V_p = [V_o \; DV_o \; \ldots \; D^{p-1}V_o]^T \quad W_q = [V_o \; DV_o \; \ldots \; D^{q-1}V_o],$$

the Hankel matrix can then be decomposed as

$$H_{p,q}(\ell) = V_p R Z^\ell W_q.$$

Since this is a full-rank decomposition, we obtain that $H_{p,q}(\ell)^\dagger = W_q^\dagger \left(Z^\ell\right)^\dagger R^\dagger V_p^\dagger$. Hence,

$$\|H_{p,q}(\ell)^\dagger\| \leq \|W_q^\dagger\| \|\left(Z^\ell\right)^\dagger\| \|R^\dagger\| \|V_p^\dagger\|.$$

But, this can be rewritten as

$$\sigma_n(H_{p,q}(\ell)) \geq \rho\,\beta^\ell \sigma_n(W_q)\sigma_n(V_p). \tag{2.9}$$

We next estimate $\sigma_n(W_q)$. The Rayleigh-Ritz singular value characterization gives that

$$[\sigma_n(W_q)]^2 = \min_{0 \neq x \in \mathbf{C}^n} \frac{x^* W_q W_q^* x}{x^* x} = \min_{0 \neq x \in \mathbf{C}^n} \sum_{j=1}^{q} \left(\frac{x^* D^{j-1} V_o V_o^* (D^{j-1})^* x}{x^* x} \right). \tag{2.10}$$

But, for all $x \neq 0$ and $y = (D^{j-1})^* x$, this yields that

$$\frac{x^* D^{j-1} V_o V_o^* (D^{j-1})^* x}{x^* x} = \frac{y^* V_o V_o^* y}{y^* y} \cdot \frac{x^* D^{j-1} (D^{j-1})^* x}{x^* x} \geq [\sigma_n(V_o)]^2 \beta^{2n(j-1)}, \tag{2.11}$$

where we used the nonsingularity of V_o and D to obtain the last inequality. Now, if v is the vector in $\mathbf{C}^n$ that minimizes the Rayleigh-Ritz quotient associated with $W_q W_q^*$, we have, using (2.10) and (2.11) successively, that

$$[\sigma_n(W_q)]^2 = \frac{v^* W_q W_q^* v}{v^* v} = \sum_{j=1}^{q} \left(\frac{v^* D^{j-1} V_o V_o^* (D^{j-1})^* v}{v^* v} \right) \geq [\sigma_n(V_o)]^2 \sum_{j=1}^{q} \beta^{2n(j-1)}. \tag{2.12}$$

A similar derivation using V_p instead of W_q yields that

$$[\sigma_n(V_p)]^2 \geq [\sigma_n(V_o)]^2 \sum_{j=1}^{p} \beta^{2m(j-1)}. \tag{2.13}$$

Replacing now (2.12) and (2.13) into (2.9), we deduce the lower bound (2.6)-(2.7). To prove the last statement of the theorem, we define $g(t,s)$ to be the real function

$$g(t,s) = f^2(t,s) = \frac{(1-\beta^{2nt})(1-\beta^{2ns})}{(1-\beta^{2n})(1-\beta^{2n})}, \quad t, s \in \mathbf{R}, \quad 0 \leq t, s \leq m.$$

We then maximize $g(t,s)$ with the constraint that $s + t = m$ and observe that the maximum is attained for $s = t = m/2$. We conclude our proof by substituting this result in (2.7). □

Thus we see that the lower bound (2.6)–(2.7) increases with the dimensions of the Hankel matrix and is maximized when this matrix is square: the closer β to 1, the more pronounced this effect. This property suggests that the smallest nonzero singular value of $H_{p,p}(\ell)$ itself increases with p, since the interlacing property of singular values implies that $\sigma_n(H_{p,p}(\ell)) \geq \sigma_n(H_{q,q}(\ell))$ whenever $p \geq q$. Given the importance of distinguishing the gap within the singular spectrum of the perturbed Hankel matrix, it thus seems useful to choose $p = q$ as large as reasonably possible. This is usually enough to obtain a suitable estimation of the signal subspace, but it may nevertheless happen that the size of the perturbation to the Hankel matrix also grows with its dimension, which makes estimation more difficult. Such a situation will be illustrated in Section 7.

3 Mathematical properties of predictor matrices

As our approach for solving the HRP is based on computing the eigenvalues of predictor matrices, we first investigate their properties. From the definition (1.4), we see that a forward predictor matrix F predicts a new data sample $h_{\ell+N}$ from the knowledge of the preceding N samples $\{h_\ell, h_{\ell+1}, \ldots, h_{\ell+N-1}\}$. In view of (1.3), F plays a similar role to that of the transition matrix T: it is thus reasonable to expect that signal frequencies and decay factors can be extracted from its eigenstructure. Indeed, (1.4) and (1.1) imply that

$$WF = ZF, \tag{3.1}$$

which shows that the rows of W are the left eigenvectors of F corresponding to the signal eigenvalues. For the moment we focus our attention on predictor companion matrices of the form

$$F = [e_2\ e_3\ \cdots\ e_N\ f] = \begin{bmatrix} 0 & 0 & \cdots & 0 & f_1 \\ 1 & 0 & \cdots & 0 & f_2 \\ 0 & 1 & \cdots & 0 & f_3 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & \cdots & f_N \end{bmatrix}$$

whose N-th column vector, f, is a solution of the system of linear equations

$$H(\ell)f = \hat{h}_{\ell+N},$$

or, equivalently, because of (1.1),

$$Wf = Z^N e. \tag{3.2}$$

Although we know the n signal eigenvalues belong to the spectrum of F, the other $(N-n)$ eigenvalues, which we call *extraneous eigenvalues*, depend on how one chooses the vector f, since the system (3.2) has infinitely many solutions because $H(\ell)$ is rank deficient. A well known result of signal analysis is then that, if one chooses the minimal norm solution of (3.2), then the extraneous eigenvalues fall inside the unit circle in the complex plane (see Cybenko, 1982, Kumaresan, 1983, and Bazán and Bezerra (1997), the last of these references providing a proof based on linear algebra arguments). In the case of stable dynamical systems, the decay factors α_i are strictly less than zero and thus both the signal and extraneous eigenvalues fall inside the unit circle. In these circumstances, the task of extracting the signal eigenvalues from the complete spectrum can be difficult and criteria for separating them are necessary. This motivates us to use the backward predictor matrices introduced above, which perform prediction in the reverse order. As for (3.1), the definition (1.5) is equivalent to the system

$$WB = Z^{-1}W. \tag{3.3}$$

Hence we see, as before, that $\{z_1^{-1},\ z_2^{-1}, \ldots, z_n^{-1}\}$ are the eigenvalues of B corresponding to the left eigenvectors given by the rows of W. Let B be a companion backward predictor matrix of the form

$$B = [b\, e_1\ e_2\ \cdots\ e_N - 1] = \begin{bmatrix} b_1 & 1 & 0 & \cdots & 0 \\ b_2 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ b_{N-1} & 0 & \cdots & 0 & 1 \\ b_N & 0 & \cdots & & 0 \end{bmatrix} \tag{3.4}$$

where the first column vector, b, is a solution of the system of linear equations

$$H(\ell+1)b = \hat{h}_\ell, \tag{3.5}$$

which is in turn equivalent to

$$Wb = Z^{-1}e. \tag{3.6}$$

It was shown in Bazán and Bezerra (1997) that, provided the vector b is the minimal norm solution of (3.6), the signal eigenvalues again separate from the extraneous ones in a clear fashion: while the extraneous eigenvalues still lie inside the unit circle, the signal ones now lie outside.

The attentive reader will have noticed that what we propose here is theoretically equivalent to extracting the frequencies and decay factors from the roots of the polynomial associated with the companion matrix, a method widely applied before the mid 1980's (see Makhoul, 1975 or Kumaresan, 1983). Our purpose is to better exploit the tools of numerical linear algebra, both numerically and theoretically, for the extraction of those roots. Numerically, we rely on the technique of balancing followed by the QR algorithm for computing the eigenvalues of the companion matrix, a procedure which was shown to be stable in Toh and Trefethen (1994) and Edelman and Murakami (1995). We then theoretically show that this eigenvalue problem is reasonably well conditioned in our context. In order to analyze this crucial property, we first prove a few useful technical results.

Lemma 2 *Let $\mathcal{P}$ and $\mathcal{Q} = I - \mathcal{P}$ be the orthogonal projectors onto $\mathcal{N}(W)^\perp$ and $\mathcal{N}(W)$ respectively. Define*

$$R^\downarrow = \mathcal{Q}B^\downarrow \tag{3.7}$$

where $B^\downarrow$ is the companion backward predictor matrix whose first column vector b is the minimal norm solution of (3.6). Also let p_i be the i-th column of $\mathcal{P}$. Then,

(a) $p_i = W^\dagger Z^{i-1}e$;

(b) $\|p_i\| < 1, \quad i = 1, 2, \ldots, N$ for all N and is a decreasing function of N;

(c) $\|b\| \leq 1$ provided $\|p_1\| \leq \frac{1}{\sqrt{2}}$;

(d) $\|R^\downarrow\| = 1$;

(e) the nonzero eigenvalues of $R^\downarrow$ are the extraneous eigenvalues of $B^\downarrow$.

Proof: We have that $\mathcal{P} = W^\dagger W$. Since the i-th column of W is $Z^{i-1}e$, we deduce that the i-th column of $\mathcal{P}$, p_i, is equal to $W^\dagger Z^{i-1}e$, $(i = 1, \ldots, N)$. This is property (a).

To prove (b), we first verify that e_i does not belong to either the nullspace of W or to its orthogonal complement. The first of these conclusions immediately follows from (1.2). To see the second, we consider the system $W^* x = e_i$. If we select n equations from this system but the i-th, we obtain a square nonsigular homogeneous system whose unique solution must be $x = 0$. However, this is in contradiction with the i-th equation, which shows that the complete system is incompatible, and thus that e_i does not belong to the range of $W^* = \mathcal{N}(W)^\perp$. Therefore

$$e_i = \mathcal{P}e_i + \mathcal{Q}e_i = p_i + \mathcal{Q}e_i,$$

and none of the two vectors of the last right-hand side are zero. The first part of property (b) is then obtained by taking norms is this last equation and using the orthogonality of the projectors.

To prove the second part, 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, if

$$W_{N+1}^{\dagger *} W_{N+1}^{\dagger} = (W_{N+1} W_{N+1}^{*})^{-1} \stackrel{\text{def}}{=} A_{N+1},$$

then $W_{N+1} W_{N+1}^{*} = A_{N+1}^{-1}$ 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 $W_{N+1} = [W_N \mid Z^N e] = [e \mid Z W_N]$. Applying the Sherman-Morrison formula twice, we derive that

$$A_{N+1} = A_N - \frac{A_N Z^N e e^* Z^{N*} A_N}{1 + e^* Z^{N*} A_N Z^N e} = Z^{-*} A_N Z^{-1} - \frac{Z^{-*} A_N Z^{-1} e e^* Z^{-*} A_N Z^{-1} e}{1 + e^* Z^{-*} A_N Z^{-1} e}. \tag{3.8}$$

Remembering now that

$$\|p_{N+1,i+1}\| = \|W_{N+1}^{\dagger} Z^i e\|, \quad (i = 0, \ldots, N),$$

or, equivalently, that

$$\|p_{N+1,i+1}\|^2 = e^* Z^{i*} A_{N+1} Z^i e,$$

we deduce from the first equality of (3.8) and (1.2) that, for $i = 0, \ldots, N-1$,

$$\begin{aligned}
\|p_{N+1,i+1}\|^2 &= e^* Z^{i*} A_N Z^i e - \frac{e^* Z^{i*} A_N Z^N e e^* Z^{N*} A_N Z^i e}{1 + e^* Z^{N*} A_N Z^N e} \\
&= \|p_{N,i+1}\|^2 - \frac{(e_{i+1}^* W_N^{\dagger} Z^N e)(e^* Z^{N*} W_N^{\dagger *} e_{i+1})}{1 + e^* Z^{N*} A_N Z^N e} \\
&= \|p_{N,i+1}\|^2 - \frac{|f_{N,i+1}|^2}{1 + \|f_N\|^2},
\end{aligned}$$

where $f_N = W_N^{\dagger} Z^N e$ is the minimum norm solution of the system (3.2). This proves the second part of (b).

The proof of (c) is very similar. We first obtain from the equality between the left-hand side and the last right-hand side of (3.8) 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

$$\|p_{N+1,1}\|^2 = \|b_N\|^2 - \frac{\|b_N\|^4}{1 + \|b_N\|^2} = \frac{\|b_N\|^2}{1 + \|b_N\|^2}.$$

Thus, we obtain that

$$\|b_N\|^2 = \frac{\|p_{N+1,1}\|^2}{1 - \|p_{N+1,1}\|^2}.$$

But, using (b), we deduce that $\|p_{N+1,1}\| \leq \|p_{N,1}\|$, and therefore that

$$\|b_N\|^2 \leq \frac{\|p_{N,1}\|^2}{1 - \|p_{N,1}\|^2}.$$

Conclusion (c) then follows.

In order to prove (d), we note that

$$R^{\downarrow} = \mathcal{Q}B^{\downarrow} = [\mathcal{Q}b, \mathcal{Q}e_1, \mathcal{Q}e_2, \ldots, \mathcal{Q}e_{N-1}] = [0, q_1, q_2, \ldots, q_{N-1}],$$

where q_k denotes the k-th column of the projector $\mathcal{Q}$ and $\mathcal{Q}b = 0$ since $b \in \mathcal{N}(W)^{\perp}$. But, observe that, since $\mathcal{Q}$ is an orthogonal projector,

$$R^{\downarrow}R^{\downarrow^*} = q_1q_1^* + q_2q_2^* + \cdots + q_{N-1}q_{N-1}^*,$$

and this matrix can be rewritten as a rank-one modification of the projector $\mathcal{Q}$ of the form

$$R^{\downarrow}R^{\downarrow^*} = \mathcal{Q}\mathcal{Q}^* - q_Nq_N^* = \mathcal{Q} - \lambda\hat{q}_N\hat{q}_N^*,$$

where $\lambda = \|q_N\|^2 \leq 1$, and $\hat{q}_N$ is the unit-length eigenvetor of $\mathcal{Q}$ in the direction of q_N. From this, and because $\text{rank}(R^{\downarrow}) = N - n$, it follows that the singular values of $R^{\downarrow}$, $\sigma_j(R^{\downarrow})$, are

$$\begin{array}{lcl}
\sigma_j(R^{\downarrow}) & = & 1, \quad j = 1, 2, \ldots, N-n-1, \\
\sigma_{N-n}(R^{\downarrow}) & = & \sqrt{1-\lambda}, \\
\sigma_j(R^{\downarrow}) & = & 0, \quad j = N-n+1, \ldots, N.
\end{array}$$

Thus assertion (d) holds.

The proof of (e) can be found in Bazán and Bezerra (1997). □

We now analyze the behaviour of the signal eigenvalues when the first column of $B^{\downarrow}$ undergoes small perturbations. More specifically, we assume the following.

A.1: For N large enough, the minimum norm solution of (3.6) satisfies $\|b\| \leq 1$.

A.2: If $\tilde{b}$ is the computed signal, b its exact value and η the perturbation, then

$$\tilde{b} = b + \eta \;\text{ and }\; \|\eta\| \ll \|b\|.$$

The first of these assumptions is motivated as follows. Lemma 2, (a) and (b), ensures that the norm of p_1 decreases when N increases and is always smaller than one. Thus it is not unreasonable to assume that, for N sufficiently large,

$$\|p_1\| \leq \frac{1}{\sqrt{2}} \approx 0.707,$$

in which case Lemma 2, (c), ensures that A.1 holds. A.2 essentially means that the noise is much smaller than the signal, a practically necessary condition for any successful calculation. Furthermore, since $\|b\| \leq 1$ by A.1, the absolute value of the noise is then very small, which allows a first-order perturbation analysis near zero. This analysis is the object of our next proposition.

Theorem 3 *Let $\lambda = z^{-1}$ be a signal eigenvalue of $B^{\downarrow}$ with associated right and left eigenvectors u and v, respectively, normalized to ensure that $u^*v = 1$. Let $\tilde{B} = B^{\downarrow} + E$ be the companion matrix whose first column vector is $\tilde{b} = b + \eta$ and let $\tilde{\lambda} = \lambda + \delta\lambda$ be the eigenvalue of $\tilde{B}$ that is closest to λ. Then,*

$$|\tilde{\lambda} - \lambda| \leq |u^*\eta| \, \|W^{\dagger}\|(1 + \|(R^{\downarrow} - \lambda I)^{-1}\|) + \mathcal{O}(\|\eta\|^2). \tag{3.9}$$

Proof: Because the algebraic multiplicity of the signal eigenvalues is one by assumption, eigenvalue perturbation theory (see Wilkinson, 1965, page 67) ensures that, for every left eigenvector u with $u^*v = 1$,

$$\delta\lambda = u^* E v + \mathcal{O}(\epsilon^2),$$

where $\epsilon = \|E\| = \|\eta\|$. But, because the structure of $E = [\eta \mid 0]$, we have that

$$\delta\lambda = v_1 u^* \eta + \mathcal{O}(\epsilon^2), \tag{3.10}$$

with v_1 the first component of v. Observe that the columns of W^* are the left eigenvectors of $B^{\downarrow}$ because of (3.3), which means that

$$Wv = e_i,$$

for some i between 1 and n, with associated eigenvalue λ. If now choose $\phi = W^{\dagger} e_i$, $\psi \in \mathcal{N}(W)$, $v = \phi + \psi$ and $u^* = [1\ z^2 \dots z^{N-1}] = e_i^* W$, then

$$u^* v = e_i^* W \phi + e_i^* W \psi = e_i^* e_i = 1$$

and the vectors u and v satisfy the normalization condition of the theorem's statement. We next consider $|v_1| = |\phi_1 + \psi_1|$. Since $\phi_1 = e_1^* W^{\dagger} e_i$, we deduce that

$$|\phi_1| \le \|W^{\dagger}\|. \tag{3.11}$$

On the other hand, decompose $B^{\downarrow} = \mathcal{P} B^{\downarrow} + R^{\downarrow}$, where $R^{\downarrow}$ is defined in (3.7), and notice that

$$\mathcal{P} B^{\downarrow} \phi = W^{\dagger} W B^{\downarrow} W^{\dagger} e_i = W^{\dagger} W Z^{-1} W W^{\dagger} e_i = W^{\dagger} W Z^{-1} e_i = z^{-1} \phi = \lambda \phi \tag{3.12}$$

and therefore that ϕ is a right eigenvector of $\mathcal{P} B^{\downarrow}$ associated with the eigenvalue λ. Hence,

$$\lambda\phi + \lambda\psi = \lambda v = B^{\downarrow} v = \mathcal{P} B^{\downarrow} v + R^{\downarrow} v = \mathcal{P} B^{\downarrow} \phi + R^{\downarrow} \phi + R^{\downarrow} \psi. \tag{3.13}$$

Equations (3.12) and (3.13) together in turn imply that

$$(R^{\downarrow} - \lambda I)\psi = -R^{\downarrow} \phi$$

where I is the identity matrix. Now, since λ is a signal eigenvalue of $B^{\downarrow}$, Lemma 2, (e), ensures that it is not an eigenvalue of $R^{\uparrow}$. It then follows that $(R^{\downarrow} - \lambda I)$ is not singular and hence, using the definition of ϕ, that

$$\psi_1 = -e_1^* (R^{\downarrow} - \lambda I)^{-1} R^{\downarrow} \phi. = -e_1^* (R^{\downarrow} - \lambda I)^{-1} R^{\downarrow} W^{\dagger} e_i.$$

But Lemma 2, (d), gives that $\|R^{\downarrow}\| = 1$, and therefore that

$$|\psi_1| \le \|(R^{\downarrow} - \lambda\, I)^{-1}\|\, \|W^{\dagger}\|. \tag{3.14}$$

Combining (3.11) and (3.14), we then obtain that

$$|v_1| \le \|W^{\dagger}\| (1 + \|(R^{\downarrow} - \lambda I)^{-1}\|). \tag{3.15}$$

The proof is then concluded by taking absolute values in (3.10) and then substituting (3.15) in the result. □

The statement of this theorem suggests the following comments. The first is that

$$\|u\| = \sqrt{1 + |z|^2 + \cdots + |z|^{2(N-1)}}$$

is not a large number because $|z| < 1$. In practice, it is often of the order of ten. On the other hand, η is also generally small because of A.2. Moreover, we note that u or η themselves do not appear in (3.9), but rather their inner product. This implies that only the component of the noise in the short signal subspace plays any role, and, because $u^*\eta \leq \|u\|\|\eta\|$, that this component is typically small. We also know from Theorem 1 that $\|W^\dagger\|$ decreases as N increases. However, nothing is known on $\|(R^\downarrow - \lambda\, I)^{-1}\|$, except that

$$|(R^\downarrow - \lambda\, I)^{-1}\| \leq \frac{1}{|\lambda| - 1}, \tag{3.16}$$

which is only moderately useful since λ may be close to one. But we expect that any increase in $\|(R^\downarrow - \lambda\, I)^{-1}\|$ will be counterbalanced by a corresponding decrease in $\|W^\dagger\|$. Ultimately, the bound in (3.9) thus depends on three factors: the importance of the noise in the short signal subspace , the dimension of the Hankel matrix N and the speed at which $\|W^\dagger\|$ decreases when this dimension grows.

Since the last theorem shows that the eigenvalues of the companion matrix are well conditioned under assumptions that are reasonable in our context, we may thus feel confident that using a stable numerical algorithm will result in satisfactory accuracy.

4 An algorithm based on conjugate gradients

The results of the above propositions as well as the mathematical properties of the ideal Hankel data matrix constitute the core of our proposal for solving the HRP. The main idea is to extract the signal eigenvalues from suitably chosen companion backward predictor matrices of the form (3.4), where the first column b is the minimum norm solution of the system of equation (3.5). We propose to compute b by conjugate gradients (CG) in its specialised version for least-squares solutions (CGLS). Our motivation for this choice is that, in the absence of noise, this algorithm will compute the value of b in at most n iterations in exact arithmetic, where n is, as above, the rank of the Hankel data matrix (in exact arithmetic). Furthermore, the convergence is also very fast when the singular values of the system matrix are clustered around one (see Bjork, 1996, page 288). Of course, convergence in less than n iterations will no longer happen when noise is present and computer arithmetic used, but we will see that a suitable stopping criterion may be designed for the CGLS iteration that will at the same time requires a relatively low number of iterations and produces a satisfactory separation of the signal from the noise. We are aware that this approach exchanges the difficulty of finding a gap in the singular spectrum for that of stopping CGLS: our contention is that the latter is often easier in practice.

We formally state our algorithm, which we call Algorithm CGLP, for Conjugate-Gradients based Linear Prediction Method.

Algorithm CGLP: Given L data points $\tilde{h}_k$, $k = 0, 1, \ldots, L-1$.

Step 1: For some $\ell \in (0, L]$, compute an approximate minimal norm solution of the system

$$\tilde{H}(\ell+1)b = \tilde{h}_\ell$$

using the CGLS algorithm.

Step 2: Build the $\tilde{B}^\downarrow$ according to (3.4) and compute all its eigenvalues $\tilde{\lambda}_j$ that lie outside the unit circle.

Step 3: Extract the decay factors and signal frequencies from

$$\tilde{s}_j = \tilde{\alpha}_j + i\,\tilde{\omega}_j = -\frac{\log \tilde{\lambda}_j}{\Delta t}.$$

End of algorithm.

Of course, we now have to specify the CGLS algorithm, and how it will be terminated. Algorithm CGLS was proposed by Hestenes (1975) and can be stated as follows for the least-squares solution of the system $Ax = c$.

Algorithm CGLS: Given an initial approximation $x^{(0)}$, set

- $r^{(0)} = c - Ax^{(0)}$,
- $p^{(0)} = s^{(0)} = A^* r^{(0)}$,
- $\gamma_0 = \|s^{(0)}\|^2$,
- for $k = 0, 1, 2, \ldots$,

$$\left\lfloor \begin{array}{lcl} q^{(k)} & = & Ap^{(k)}, \\ \alpha_k & = & \gamma_k / \|q^{(k)}\|^2, \\ x^{(k+1)} & = & x^{(k)} + \alpha_k p^{(k)}, \\ r^{(k+1)} & = & r^{(k)} - \alpha_k q^{(k)}, \\ s^{(k+1)} & = & A^* r^{(k+1)}, \\ \gamma_{k+1} & = & \|s^{(k+1)}\|^2, \\ \beta_k & = & \gamma_{k+1}/\gamma_k, \\ p^{(k+1)} & = & s^{(k+1)} + \beta_k p^{(k)}. \end{array} \right.$$

End of algorithm.

The convergence properties of Algorithm CGLS can be studied by observing that the k-th iterate, $x^{(k)}$, can be defined as the unique solution of the problem

$$\min \|Ax - c\| \quad \text{constrained to} \quad x \in \mathcal{K}_k(A^*A, A^*c),$$

where $\mathcal{K}_k(A^*A, A^*c)$ is the Krylov subspace associated to the normal equations, that is

$$\mathcal{K}_k(A^*A, A^*c) = \text{span}\{A^*c, (A^*A)A^*c, \ldots, (A^*A)^{k-1}A^*c\}.$$

We refer the reader to Bjork (1996), page 288, for further detail. Then the following result is known to hold (see Hestenes, 1975).

Theorem 4 *Let $x^{(k)}$ the sequence of CGLS iterates associated with the least-squares problem* $\min \|Ax - c\|$ *. Then, provided that $x^{(0)} \in \mathcal{R}(\mathcal{A}^*)$, e.g. $x^{(0)} = 0$, the sequence $\{x^{(k)}\}$ converges to the pseudo-inverse solution $x^\dagger = A^\dagger c$, and convergence occurs in at most p steps, where $p = rank(A)$.*

If exact arithmetic is used, this result indicates that CGLS is efficient in that it must find a solution of the system (3.5) in at most n iterations. If we now consider that our data is noisy, it is possible to show, when the noise in Gaussian with zero mean (see De Moor (1993)), that, when N is sufficiently large,

$$\tilde{H}(\ell)^*\tilde{H}(\ell) \approx H(\ell)^*H(\ell) + \sigma I \tag{4.1}$$

for some $\sigma > 0$ proportional to the noise variance. This is of interest since we also know (see Golub and Van Loan (1996), page 530) that CGLS converges in at most $n+1$ iterations if (4.1) holds as an equality. Furthermore, Golub and Van Loan's "metatheorem" (page 530) can be rephrased in our framework to say that, "if $\tilde{H}(\ell)^*\tilde{H}(\ell)$ is close to a rank n correction of the identity, then the CGLS method almost converges after $n+1$ steps". Unfortunatley, we may not in general assume that the noise is Gaussian with zero mean, nor can we use exact arithmetic to solve our problem. Since $\tilde{H}(\ell)$ must be considered as a general full-rank matrix (which says that CGLS would require N iterations to converge in exact arithmetic), we may therefore expect that CGLS may not converge as quickly as one hopes, due to both noise and arithmetic inaccuracies. We thus face the problem of stopping the CGLS iterations as soon as its behaviour becomes completely dominated by these inaccuracies.

Nothing is known to the authors on the behaviour of CGLS for perturbed low-rank systems. However, more experience is available for truncated SVD-based techniques. If there exists a clear gap between signal and noise singular values of the coefficient matrix, the problem has been extensively analyzed, and suitable criteria for truncating the singular spectrum have been developed. For instance, Hansen (1987), Hansen (1990) and Hansen (1992) describe the so-called "discrete Picard condition" and suggest to use the "L-curve criterion" when it holds.

The case where the singular values $\tilde{\sigma}_k$ decay gradually with no gap in the singular value spectrum is more difficult to analyze since there is no intuitive way of choosing a truncation parameter for wich the truncated SVD solution provides a satisfactory approximation to b. However, in discrete ill-posed problems, the L-curve criterion may still be used to stop CGLS even if the gap is unclear, and this strategy appears to work often well in practice. The only known explanation of this success is heuristic and uses the observation that, once CGLS has converged to the low-frequency components of the solution, convergence is then superseded by divergence of the iterates' norm in a dramatic form (see Bjork, 1996, Hanke and Hansen, 1993 or Hanke, Nagy and Plemmons, 1993). Unfortunately, this is not yet supported by a theoretical analysis. However, the use of the L-curve criterion is not always without difficulties, and the authors's experience with this technique in the solution of perturbed prediction problems has been to date very unsatisfactory, perhaps because no drastic change in the approximate solution norm was observed at any particular iteration. But numerical examples and model problems reported in Van der Sluis (1992) and Van der Sluis and Van der Vorst (1990) indicate that, even when no distinct gap is present in the singular value spectrum, the CGLS algorithm still manages to extract the components of the solution associated with the largest singular values in very few iterations. This behaviour can in turn be explained by the observation that the approximation of the eigenvalues of A^*A projected on the Krylov subspace $\mathcal{K}(A^*A, A^*b)$ (the so-called Ritz values) still appear to converge to the square of the largest singular values of A, as long as the discrete Picard condition holds for the unperturbed right hand side and the large singular values of A are well separated.

Although heuristic, this is possibly the stronger reason for which CGLS works well for the solution of discrete ill-posed problems, and is also important in our motivation to use this algorithm on our problem.

5 Stopping the conjugate-gradients iteration

We now discuss the crucial decision to stop the CGLS iterations early enough to avoid contamination by the noise but late enough to capture the most important components of the solution of the HRP.

We start by emphasizing that usual residual-based criteria or other convergence tests for stopping the CG method (see for instance, Bruaset, 1995 and the references therein), cannot be used as stopping rules for CGLS in our context, because we are not interested in an exact solution of the least-squares problem (which involves the noise in the data), but rather in an early approximation in which noise contamination is low. In our opinion, no efficient stopping criterion for CGLS could be developed if sufficient information about the problem under analysis (or its solution) is not available. Thus our purpose is not to elaborate a stopping rule that applies for general perturbed rank-defficient least-squares problems. We prefer instead to consider a strategy that enables us to stop CGLS at some iteration where, at least for our purpose, the underlying signal information is captured.

Our strategy is based on two observations. The first is that every approximation of the solution computed during the CGLS procedure may be analyzed for its frequency components: typically, the signal corresponding to the CGLS approximate solutions contains more and more high frequency components as iterations proceed. The second is that these solutions (considered in the frequency domain) diverge surprisingly quickly after the components associated with large singular values have been found. We therefore consider a strategy where the frequency content of each successive CGLS iteration is analyzed and where the decision to stop the algorithm is based on this analysis.

Practically, we perform the frequency analysis by applying the fast Fourier transformation to the CGLS iterates and we examine the number of frequency peaks in the resulting square Fourier spectrum. This spectrum, sometimes called the power density spectrum, has the advantage that it reinforces meaningful peaks relatively to smaller ones. If the data is noise free, then the authors' experience is that the peaks corresponding to the frequency components of the desired signal (the minimum norm solution of the prediction problem) are detected very early on in CGLS and remain unmodified until convergence is obtained. The only observed effect of CGLS iterations in the frequency domain is an alteration of the height and width of these peaks, but not of their number or position. We believe that this behaviour is probably a consequence of the fact that the successive CGLS iterates, which we denote by $b^{(k)}$, all belong to the short signal space $\mathcal{R}[H(\ell)]^*$. On the other hand, this behaviour dramatically changes as soon as noise is present in the data: if, at early iterations, the number of frequency peaks is nearly identical to that observed in the noise free case, it suddenly considerably increases when the noise starts contaminating the computed signal. Examples illustrating this observation will be discussed in the Section 7.

We therefore propose to stop the CGLS iteration as soon as the number of peaks in the power density spectrum of the iterates explodes. This proposal is in accordance with the conclusions of Hanke and Hansen (1993), who suggest that the most appropriate form of stopping CGLS is likely to be via on-line visualization of the behaviour of the iterates. Our contribution here is to focus on the frequency content of these iterates. There are obviously many possible ways to implement such a rule. Our choice has been to retain the first iterate $b^{(k)}$ such that the number of frequency peaks associated with $b^{(k+1)}$ is at least twice that obtained for $b^{(k)}$.

6 Computational cost

The purpose of this section is to analyse the computational complexity of the CGLP method, and to compare it with that of HTLS, a well respected subspace method. Let us consider CGLP first, and examine its main steps successively.

Step 1: This step involves the least-squares estimation of the vector b. It is performed by applying the CGLS algorithm and stopping as soon as the Fourier transform of the iterates produce significantly many more peaks than the previous one (see above). The cost of this step will thus be a linear function of the number of CGLS iterations that are needed. Each such iteration requires $(4MN + 6N + 4M)$ flops, where a flop is an add or a multiply (see Bjork, 1996, page 290). On the other hand, the FFT transform to obtain the power spectrum at each iteration costs $\frac{1}{2}N * \log(N)$ flops (see Golub and Van Loan, 1996, page 190). Thus the total computational cost for Step 1 can therefore be estimated to be

$$Iter * (4MN + 6N + 4M + \frac{1}{2}N\log(N)) \text{ flops } .$$

Step 2: Computing the companion backward predictor matrix does not require any flop, but its eigenvalue analysis involves approximately $10N^3$ flops. This is typically the dominant part of the whole calculation.

Step 3: Since we only need to retain the eigenvalues larger than one in modulus, this involves $3N$ flops for the modulus calculations and approximately n flops plus n evaluations of the complex logarithm to finally obtain the frequencies and decay factors.

This brings us to a grand total of (approximately)

$$Iter * (4MN + 6N + 4M + \frac{1}{2}N\log(N)) + 10N^3 + 3N + n \text{ flops and } n \text{ logarithms.}$$

Let us now consider the cost of the HTLS algorithm.

Step 1: SVD decomposition of the Hankel matrix. This matrix is $M \times N$ and the SVD decomposition (where we only consider the first N columns of the "left" orthogonal matrix) then costs $6MN^2 + 20N^3$ flops using the Chan-SVD. One obtains the factorization $H(\ell) = P\Sigma Q^*$, from which the matrices $V^{\uparrow}$ and $V^{\downarrow}$ are then extracted.

Step 2: One then computes a second SVD decomposition $[V^{\downarrow}V^{\uparrow}] = \hat{P}\hat{\Sigma}\hat{Q}^*$, which costs $12(M-1)n + 20n^3$ flops using the Chan-SVD algorithm.

Step 3: Forming the $T_{\rm VH}$ matrix from $\hat{Q}$ then requires approximately $\frac{4}{3}n^3$ flops, and the eigenvalue analysis of this latter matrix uses approximately $10n^3$.

Step 4: As for CGLP, the frequencies and decay factors are finally extracted at a cost of n flops and n logarithm calculations.

This brings us to an (approximate) total of

$$6MN^2 + 20N^3 + 12(M-1)n^2 + 21.3n^3 + 3N + n \text{ flops and } n \text{ logarithms}$$

for the HTLS algorithm.

To simplify the bounds slightly, we now assume that $M = \alpha N$ ($\alpha \geq 1$). Then the cost ratio becomes

$$\frac{\text{cost for HTLS}}{\text{cost for CGLP}} \approx \frac{26\alpha N^3 + 12\alpha(N-1)n^2 + 21.3n^3 + 3N + n}{10N^3 + Iter * ((4\alpha N^2 + (6+4\alpha)N + \frac{1}{2}N\log(N)) + 3N + n}$$

where we have neglected the cost of the logarithm calculations for both algorithms.

We first consider the case where $\alpha = 1$, that is $M = N$, which corresponds to the conclusions of Theorem 1 and the recomendations of Van Huffel et al. (1994) to choose the Hankel matrix as square as possible. In this case, the ratio of the costs for the two algorithms is given by

$$\frac{\text{cost for HTLS}}{\text{cost for CGLP}} \approx \frac{26N^3 + 12(M-1)n^2 + 21.3n^3 + n}{10N^3 + Iter * (4N^2 + 10N + \frac{1}{2}N\log(N)) + 3N + n}$$

where we have again neglected the logarithms. If we make the further assumption that these costs are dominated by the terms in N^3, we obtain that HTLS is approximately twice as expensive as CGLP. However, this assumption is not entirely realistic in that the other terms are not completely negligible, which decreases this ratio somewhat. If, on the other hand, we use rectangular Hankel matrices (as is often done in practice to include more data while keeping the linear algeabra cost within bounds), then the conclusions are even more favourable for CGLP.

We conclude this section by noting that the relative computational advantage of CGLP can be further enhanced if one uses an eigenvalue calculation algorithm that only computes the eigenvalues outside the unit circle. The flop count for this part of the calculation is then reduced much below the total of $10N^3$ that we have used above. Exactly how much can be gained in this way is difficult to estimate in general, as it depends on the size N of the matrix at hand, and on n, which depends on the nature of the signal itself. We refer the reader to Duff and Scott (1993) for further details on this interesting possibility, and discuss some preliminary numbers in the next section.

7 Numerical experience

We are now in position to describe our numerical experiments, whose purpose is to illustrate the discussions of the preceding sections. We consider three HRP problems, the first two being identification problems for simulated vibrating systems and the third the practical estimation of a building structure modal parameters from a single free decay response function. All computations were performed in MATLAB on an DEC Alphastation 500 (300 MHz) at the FUNDP, Namur, Belgium. The data set for the first two experiments was chosen such that the singular spectrum presents the two extreme situations discussed in the previous section (very clear and very unclear spectral gaps). We used two different implementations of our algorithm. The first (version 1) computes all eigenvalues of the companion matrix (using QR after balancing), and the second (version 2) uses a subspace iteration method for computing only the eigenvalues which correspond to the projection of the matrix onto the last Krylov subspace obtained by the CGLS procedure (we used the CGLS normalized residuals to provide an orthonormal basis for this subspace). This is a simplified version of the algorithm by Duff and Scott (1993), which, in particular, does not use acceleration techniques proposed by these authors. For comparison, analogous numerical experiments using HTLS are also reported. We furthermore include numerical results on eigenvalue sensibility for the first example. In our description, we use the following symbols:

- $Iter$: number of CGLS iterations for obtaining the approximate solution ,
- Nt: the truncation level used by SVD subspace methods (which can also be viewed as an estimate of the number of different frequencies contained in the signal or as the dimension of the signal subspace),
- Nsr: Noise to signal ratio defined by

  $$Nsr = \frac{||\eta||}{||h||}$$

  where η and h are the vectors containing the noise and the signal respectively,
- E_α: relative error of estimated decay factors, that is

  $$E_\alpha = \frac{|\alpha - \tilde{\alpha}|}{|\alpha|},$$
- E_ω: Relative error of estimated frequencies, that is

  $$E_\omega = \frac{|\omega - \tilde{\omega}|}{|\omega|},$$

- $||r^{(k)}||$: CGLS residual norm at step k.

7.1 A first synthetic example

In this example we consider data obtained by sampling the signal defined by

$$h(t) = 0.2\, e^{-0.1t} \sin(7t) + 0.16\, e^{-0.056t} \sin(10t) + 0.12\, e^{-0.09t} \sin(18t). \tag{7.1}$$

This signal contains 6 frequencies, which implies that the associated Hankel data matrix $H(\ell)$ has rank 6. The available data for this experiment consists of 1024 data points obtained by sampling the signal 20 times per second (i.e. $\Delta t = 0.05$).

7.1.1 Signal eigenvalue sensitivity

We first use this signal to illustrate the behaviour of the upper bound (3.9). We start by observing that the signal eigenvalues in this case are: $z_1 = e^{-0.1\Delta t + 7\, i\, \Delta t} = \bar{z}_4$, $z_2 = e^{-0.056\Delta t + 10\, i\, \Delta t} = \bar{z}_5$ and , $z_3 = e^{-0.09\Delta t + 18\, i\, \Delta t} = \bar{z}_6$. We next compute $||W^{\dagger}||$ for several values of N and for $n = 6$, as well as the constants $||(R^{\downarrow} - \lambda I)^{-1}||$ and $||W^{\dagger}||(1 + ||(R^{\downarrow} - \lambda I)^{-1}||)$ for $N = 256$ at each of the different frequencies of the signal. These are shown in Tables 1 and 2, respectively. This last table also shows (in its last row) the error obtained after evaluating the eigenvalues using CGLP on a signal perturbed by Gaussian noise at a level of 5% .

N	6	36	72	128	256
$\lVert W^{\dagger}\rVert$	414.01	0.2088	0.1535	0.1233	0.1054

Table 1: Behaviour of $||W^{\dagger}||$ as a function of N

Signal eigenvalues	$z_1 = \bar{z}_4$	$z_2 = \bar{z}_5$	$z_3 = \bar{z}_6$
$\lVert(R^{\downarrow} - \lambda I)^{-1}\rVert$	93.3710	137.6654	100.9153
$(\lvert\lambda\rvert - 1)^{-1}$	199.5004	356.6431	221.7226
$\lvert u^{*}\eta\rvert$	0.0088	0.0029	0.0020
$\lVert u\rVert\, \lVert\eta\rVert$	0.1637	0.1985	0.1704
bound of (3.9)	0.0872	0.0417	0.0212
$\lvert\tilde{\lambda} - \lambda\rvert$	0.01	0.0056	0.0089

Table 2: Analysis of the eigenvalue error bound for $N = 256$, $||\eta|| = 0.017$.

We note the marked decrease of $||W^{\dagger}||$ for increasing values of N, and the corresponding important role in the overall bound of the product $|u^{*}\eta|\, ||W^{\dagger}||$. We also emphasize that the relative error predicted by the bound is very comparable to the noise to signal ratio (5%). We observe that the accuracy of the eigenvalue as calculated by CGLP is of the same order as that predicted by the bound. Finally, we notice that our comment indicating that the bound (3.16) is only moderately useful is vindicated by the numerical results.

7.1.2 Retrieving frequencies and decay factors

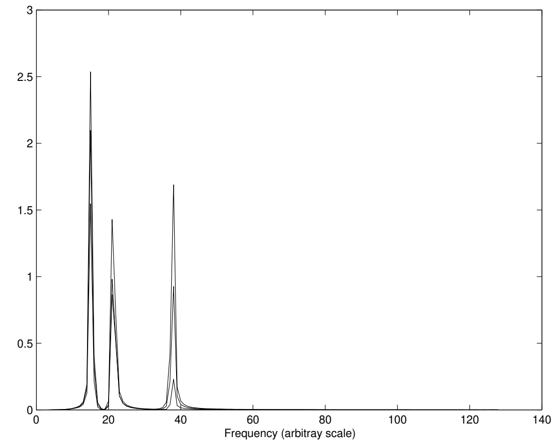
We continue our analysis of example (7.1) by examining how frequencies and decay factors can be retrieved from noise corrupted data. For this analysis, we have perturbed the Hankel matrices (corresponding to our signal) for $N = 256$ by adding zero mean Gaussian noise at 5% and 50 % levels. Results of GCLP method as well as those of HTLS for the cases in which the noise level was of 5% and 50% are presented in Table 3. The SVD-based methods were all truncated for a subspace of dimension 6. Numerical values of norms of CGLS iterates and their corresponding residual norms as well as for some singular values of the corrupted Hankel matrix are presented in Table 4.

Nsr	Method	Mode	α_k	$\tilde{\alpha}_k$	E_α	ω_k	$\tilde{\omega}_k$	E_ω
5%	CGLP	1	0.1000	0.0990	0.0098	7.0000	6.9998	0.0000
		2	0.0560	0.0546	0.0246	10.0002	10.0002	0.0000
		3	0.0900	0.0875	0.0276	18.0019	18.0000	0.0001
	HTLS	1	0.1000	0.1201	0.2008	7.0000	7.0048	0.0007
		2	0.0560	0.0546	0.0244	10.0000	10.0005	0.0000
		3	0.0900	0.0876	0.0272	18.0000	17.9998	0.0000
50%	CGLP	1	0.1000	0.1028	0.0285	7.0086	7.0000	0.0012
		2	0.0560	0.0472	0.1571	9.9957	10.0000	0.0004
		3	0.0900	0.0986	0.0957	17.9949	18.0000	0.0003
	HTLS	1	0.1000	0.1093	0.0928	7.0051	7.0000	0.0007
		2	0.0560	0.0482	0.1392	9.9946	10.0000	0.0005
		3	0.0900	0.0947	0.0525	17.9925	18.0000	0.0004

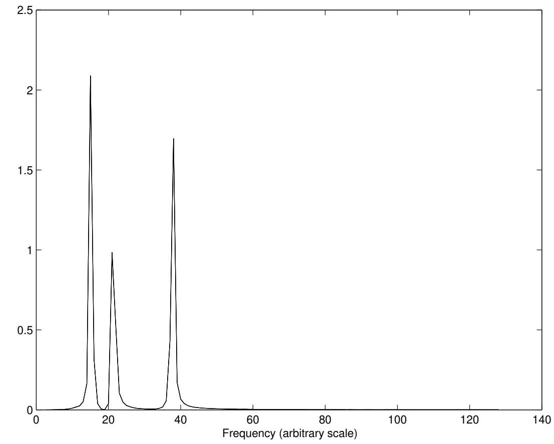
Table 3: Numerical results of Example 1: $N = 256$, $Iter = 6$ and $Nsr = 5\%$ (above) and $Iter = 3$ and $Nsr = 50\%$ (below).

Nsr	k	$\tilde{\sigma}_k$	$\|r^{(k)}\|$	$\|b^{(k)}\|$	k	$\tilde{\sigma}_k$	$\|r^{(k)}\|$	$\|b^{(k)}\|$
5%	1	10.9789	0.6535	0.1974	6	5.9214	0.1192	0.2385
	2	10.8370	0.3097	0.2254	7	0.2418	0.0879	0.5135
	3	9.3635	0.1215	0.2384	8	0.2384	0.0717	0.7521
	4	9.0893	0.1194	0.2385	9	0.2321	0.0608	0.9701
	5	5.9803	0.1194	0.2385	10	0.2237	0.0555	1.1006
50%	1	10.7775	1.0211	0.2015	6	5.5860	0.5580	0.6004
	2	10.6470	0.8583	0.2289	7	2.1642	0.5453	0.6215
	3	9.6981	0.8028	0.2444	8	1.9829	0.4549	0.8102
	4	9.4000	0.7648	0.2716	9	1.6242	0.3986	0.9571
	5	5.9514	0.5790	0.5661	10	1.6130	0.3623	1.0752

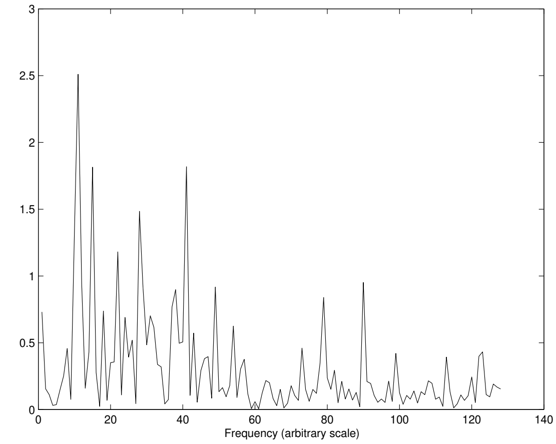
Table 4: Singular values, CGLS residual norms and CGLS iterate norms of Example 1: $N = 256$, $Iter = 6$, $Nsr = 5\%$ (above) and 50% (below).



(a)



(b)



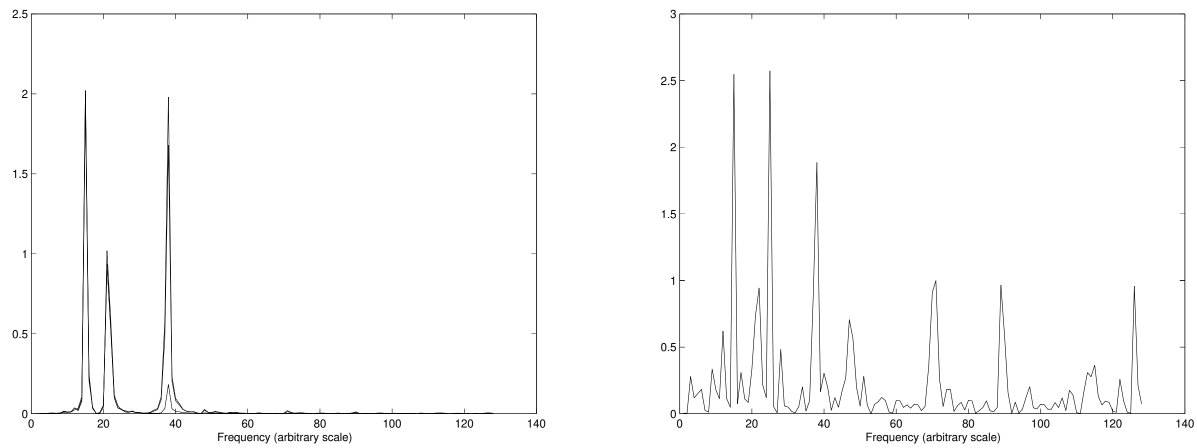
(c)

Figure 1: Power spectral density curve of CGLS iterates of Example 1. Nsr = 5%. (a): k = 1, 2, 3; (b): k=4,5,6; (c): k =7

We first note that Table 4 shows a well defined gap ($\tilde{\sigma}_6 - \tilde{\sigma}_7 \approx 5$) in the singular spectrum of the perturbed Hankel matrix when the noise to signal ratio is small (5%). Note that, if we had chosen $N = 64$, the gap would then be reduced by approximately 2 instead of 5 , in accordance with Theorem 1. We also see in this table that stopping the CGLS algorithm on the basis of the residual norm or of the iterate's norm only appears to be very difficult, even in this case. This is in contrast with the power spectrum behaviour for the same iterates, which is shown in Figure 1. In this figure, the explosion of the number of frequency peaks for $k = 7$ is very apparent. As we implied earlier, the correct number of frequency peaks (3) is detected at the *first* iteration of CGLS, and remains constant up to the point where it explodes.

This example is also ideal, in its low noise (and correspondingly clear spectral gap) version, for comparing CGLP and HTLS. This comparison is provided in Table 3, where we observe that all methods behave satisfactorily. However, if the noise level is increased to 50%, a reasonable approximation to the desired solution can still be constructed by CGLS provided that the iterations are suitably stopped. This is illustrated in Figure 2 and Table 3: one clearly sees there that stopping CGLS for $k = 3$ is adequate. This is in accordance with the expected fact that a higher noise level implies that more of the small singular values are swamped and thus that the spectrum must be truncated earlier. It should be noted here that HTLS is shown in a slightly advantageous light because, although no clear gap can be seen in the singular spectrum, the correct signal subspace dimension (6) was nevertheless employed.

In this example ($N = 256$), version 1 of CGLP (with $Iter = 6$) was 1.5 times faster than HTLS (with $Nt = 6$). Version 2 was 83 times faster than HTLS for the case where $Nsr = 5\%$. This is due the excellent approximation of the signal subpsace produced by CGLS for the moderately noisy case. The quality of this subspace is then better exploited by the eigenvalue calculation. When $Nsr = 50\%$, the acceleration is in the range from 20 to 42, depending on how carefully the eigenvalue algorithm is initialized. In this case, the quality of the signal subspace approximation is further deteriorated by the more important noise.



(a) (b)

Figure 2: Power spectral density curve of CGLS iterates of Example 1. Nsr = 50%. (a): k = 1, 2, 3; (b): k=4

7.2 A second synthetic example

This second example simulates a vibrating system described by a model function containing 8 frequencies where the higher ones are subject to more damping. The contribution of these frequencies in the power spectrum is correspondingly smaller, which makes them more sensitive to the effect of the noise. The signal to be analyzed is defined by

$$h(t) = 0.2\, e^{-0.1t} \sin(7t) + 0.16\, e^{-0.056t} \sin(10t) + 0.12\, e^{-0.09t} \sin(18t) + 0.16 e^{-0.5t} \sin 40t$$

As in Example 1, the sampling interval is given by $\Delta t = 0.05$. We report the experiments for $N = 256$ and (Gaussian, zero mean) noise to signal ratio of 40% .

Method	Mode	α_k	$\tilde{\alpha}_k$	E_α	ω_k	$\tilde{\omega}_k$	E_ω
CGLP	1	0.1000	0.1167	0.1669	6.9971	7.0000	0.0004
	2	0.0560	0.0520	0.0710	9.9982	10.0000	0.0002
	3	0.0900	0.0849	0.0568	18.0023	18.0000	0.0001
	4	0.5000	0.3290	0.3420	40.0502	40.0000	0.0013
HTLS	1	0.1000	0.1166	0.1664	6.9979	7.0000	0.0003
	2	0.0560	0.0535	0.0453	9.9983	10.0000	0.0002
	3	0.0900	0.0807	0.1037	18.0010	18.0000	0.0001
HTLS	1	0.1000	0.1177	0.1767	6.9952	7.0000	0.0007
	2	0.0560	0.0540	0.0352	9.9970	10.0000	0.0003
	3	0.0900	0.0833	0.0743	17.9991	18.0000	0.0001
	4	0.5000	0.3466	0.3068	40.0552	40.0000	0.0014

Table 5: Numerical results of Example 2: $N = 256$, $Iter = 6$ and $Nsr = 40\%$, $Nt = 6$ (middle) and $Nt = 8$ (bottom).

k	$\tilde{\sigma}_k$	$\|r^{(k)}\|$	$\|b^{(k)}\|$	k	$\tilde{\sigma}_k$	$\|r^{(k)}\|$	$\|b^{(k)}\|$
1	11.0968	1.0782	0.1995	6	6.1341	0.5849	0.3965
2	10.9560	0.8414	0.2369	7	2.0991	0.5277	0.4901
3	8.8749	0.7663	0.2567	8	2.0651	0.4278	0.7212
4	8.5804	0.7081	0.2875	9	1.1798	0.3856	0.8371
5	6.1970	0.5985	0.3831	10	1.1768	0.3436	0.9773

Table 6: Singular values, CGLS residual norms and CGLS iterate norms of Example 2: $N = 256$, $Iter = 6$, $Nsr = 40\%$.

The swamping of the peak corresponding to the higher frequencies is apparent in Figure 3, parts (a) and (b). Table 6 also shows that no clear gap can be detected in the singular spectrum. Figure 3, parts (c) and (d) illustrate why we have chosen to stop the CGLS algorithm after 6 iterations. The abscence of a clear gap suggested to experiment with two values ($Nt = 6$ and 8) for the size of the signal space in HTLS. Results of this experiments are shown in Table 5. The first observation is that CGLP and the subspace methods produce similar answers for the three first frequencies when the subspace methods use $Nt = 6$. The fourth frequency is obviously not calculated by the subspace methods for this value of Nt, although it is reasonably approximated by CGLP. If one uses $Nt = 8$ instead, then all frequencies are found by all methods, which are then very comparable.

In this example ($N = 256$), version 1 of CGLP (with $Iter = 6$) was 1.5 times faster than HTLS (with $Nt = 8$). Version 2 was from 20 to 40 times faster than HTLS, again depending on the details of the eigenvalue algorithm initialization. This is very consistent with the speed-ups observed in the first example.

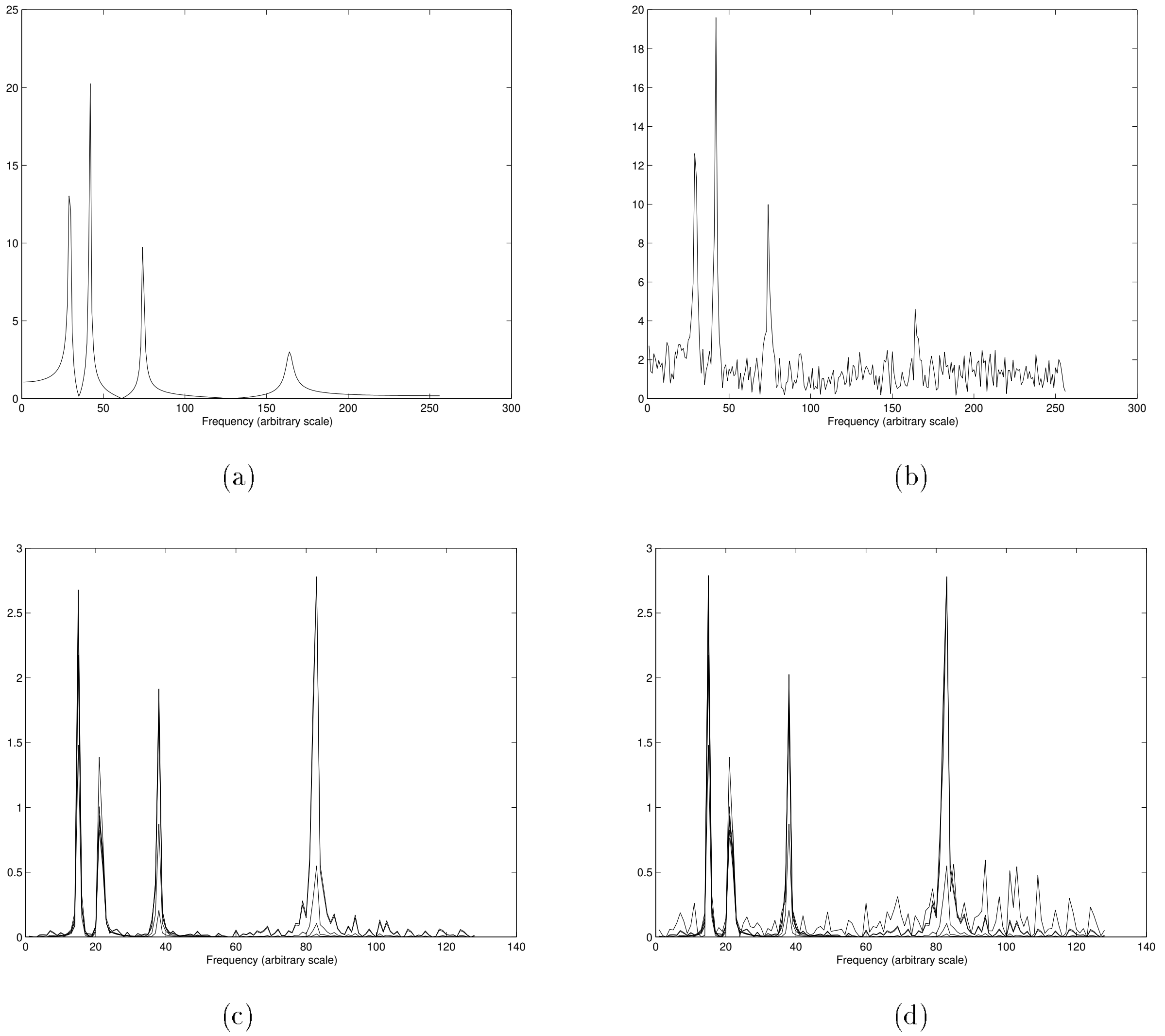
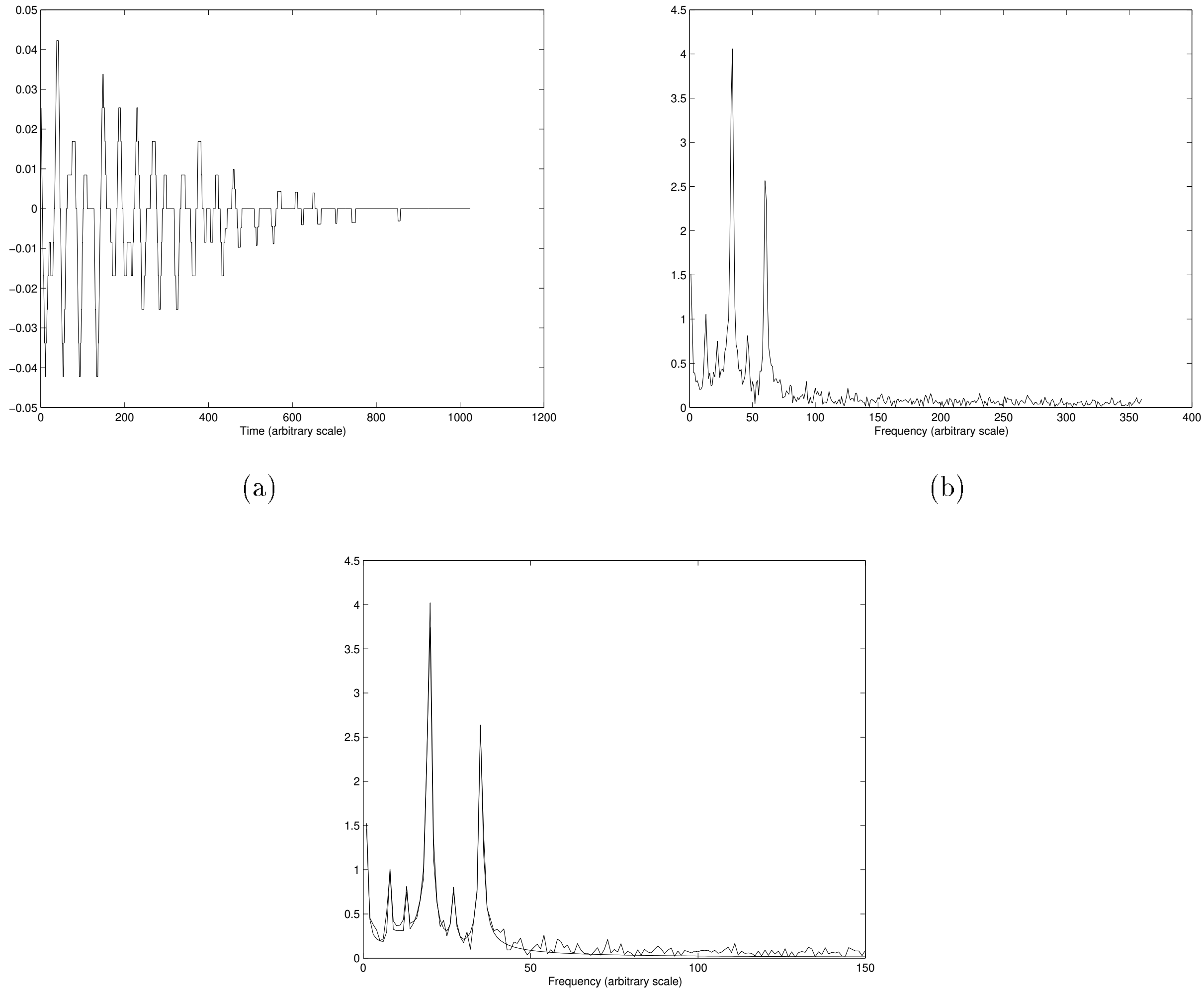


Figure 3: Fourier spectrum of Example 2. (a): clean data, (b): noise data. Power spectral density curve of CGLS iterates of Example 2. Nsr = 40%. (c): k = 1, 2,. . ., 6; (d): k =7

7.3 A example using experimental data

The data set considered in this example was gathered in the Acoustics and Vibration Laboratory of Universidade Federal de Santa Catarina, SC, Brazil. It consists of $L = 1024$ samples of the free decay response associated with a point of a real vibrating structural model of a three levels industrial building, and was recorded at a rate corresponding to $\Delta t = 1/512$. Because no information on the order of the system was available from the singular value spectrum (see Table 7) , subspace methods were applied with $Nt = 8, 10$ and 12. N was chosen to be 360, a relatively large value compared to L. We report our numerical results in Table 8. In this table, the # symbol indicates the presence of real eigenvalues. These eigenvalues were discarded because they have no impact on the vibratory system. The $\xi_k = \alpha_k/|s_k|$ are the damping factors, while the f_k are the estimated frequencies measured in Hz. The prediction problem was solved in this case with only 7 CGLS iterations. We immediately note that CGLP and HTLS do not find the same set of frequencies. However, results reported in Bazán and Bavastri (1996) for a multiple-input multiple-output analysis of the same structure using ICATS[1] indicate that both sets of frequencies are correct. CGLP and HTLS thus appear to have indentified different parts of the spectrum. The quality of the CGLP solution is illustrated in Figure 4 where the Fourier spectrum of both the original data and the fitted data are displayed. Also observe that HTLS only found four correct frequencies when $Nt = 10$, which should, in principle, be enough to recover five.



(a) (b) (c)

Figure 4: (a): Time data. (b): Fourier spectrum of experimental data. (c): Frequency domain fitting using CGLP parameters

In this example ($N = 360$), version 1 of CGLP (with $Iter = 7$) was 1.43 times faster than HTLS (with $Nt = 12$). Version 2 is between 15 and 20 times faster than HTLS.

[1] This software, produced by the Imperial College, London, is a well-trusted engineering tool in modal analysis. See Imperial College London (1995).

k	$\tilde{\sigma}_k$	$\|r^{(k)}\|$	$\|d^{(k)}\|$	k	$\tilde{\sigma}_k$	$\|r^{(k)}\|$	$\|d^{(k)}\|$
1	2.2822	0.1483	0.1379	8	0.4241	0.0478	0.3614
2	2.2275	0.0972	0.1764	9	0.4208	0.0420	0.4782
3	1.4403	0.0833	0.196	10	0.2343	0.0379	0.5971
4	1.4117	0.0681	0.2300	11	0.2340	0.0365	0.6462
5	0.7036	0.0618	0.2485	12	0.2163	0.0350	0.7004
6	0.4643	0.0580	0.2633	13	0.2142	0.0349	0.7030
7	0.4600	0.0508	0.3228	14	0.1564	0.0344	0.7232

Table 7: Singular values, CGLS residual norms and CGLS iterate norms of Example 3: $N = 360$, $Iter = 7$.

Mode	CGLP		HTLS ($Nt = 8$)		HTLS ($Nt = 10$)		HTLS ($Nt = 12$)	
	ξ_k	f_k (Hz)	ξ_k	f_k (Hz)	ξ_k	f_k (Hz)	ξ_k	f_k (Hz)
1	0.0184	4.7736	0.0358	4.6876	0.0543	4.6802	0.0500	4.6412
2	0.0215	8.6289	#	0	#	0	#	0
3	-	-	#	0	#	0	0.0007	11.4961
4	0.0145	13.2791	0.0186	13.4313	0.0193	13.4007	0.0194	13.4166
5	0.0195	18.5578			0.0144	18.5458	0.0139	18.5663
6	0.0094	24.3125	0.0105	24.2784	0.0101	24.2915	0.0101	24.2940

Table 8: Numerical results of Example 3: $N = 360$, $Iter = 7$.

8 Conclusion

We have proposed a new method for solving the HRP which does not require the a priori estimation of the signal subspace dimension. This method is based on the use of the conjugate-gradients algorithm and the extraction of eigenvalues from associated companion matrices. This proposal is supported by new results on the sensitivity of these eigenvalues, and also by the ability of the conjugate-gradient method to quickly extract the dominant part of the signal when coupled with an adequate stopping rule which prevents noise to contaminate the results. Preliminary numerical experiments on synthetic and real data have shown that the method produces results of an accuracy comparable to that of the HTLS algorithm, although at a substantially reduced computational cost. The authors believe that this accuracy, which is remarkable for a pure least-squares technique, is comparable to that of a method using total least-squares because, as shown in Theorem 3, it only depends on the component of the noise in the signal subspace, while total least-squares approaches depend on the error between the complete true and noise perturbed signal subspaces, which is potentially larger. We also showed that the reduction in computational cost is even more substantial when the noise level is not too high, which then permits a very efficient use of subspace iteration methods for computing only the signal eigenvalues. We finally obtained a new result (Theorem 1) on the smallest signal singular values of the noise free Hankel matrix, which tends to favor the choice of large and square such matrices.

It is clear that continued experience with the new algorithm is necessary to fully assess its potential. In particular, further computational work is necessary to fully exploit the potential of subspace iteration methods for eigenvalues, by the use of acceleration techniques and by additional optimization of the complete procedure. A comparison should also be carried out with methods that compute only a part of the singular spectrum of the Hankel matrix: a speed-up of around ten is expected for the corresponding version of HTLS (according to Van Huffel et al., 1994). Finally, one needs more experience on the behaviour of the algorithm when applied to complex data. Further analysis of the sensitivity of the eigenvalue calculation are also of interest and are the subject of ongoing research.

Acknowledgments

This research was performed during a visit of the first author in FUNDP, Namur, Belgium, funded by the CNPq, Brasil, grant 201299/96-8(NV), while the visit of the third author at FUNDP was funded by CNPq, grant 201429/95-O(NV).

The support of the CNPq and of the Department of Mathematics at FUNDP is gratefully acknowledged. Special thanks are due to B. Colson and E. Cornélis for their help and interest in this research.

References

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**, 365–3680, 1996.

F. S. V. Bazán and L. H. Bezerra. On zero location of predictor polynomials. *Numer. Linear Algebra With Applications*, (to appear), 1997.

A. Bjork. *Numerical Methods for least squares problems.* SIAM, Philadelphia, USA, 1996.

A. Magnus Bruaset. *A survey of preconditioned iterative methods.* Pitman Research Notes in Math. series 328. Longman Scientific & Technical, UK, 1995.

H. Chen, S. Van Huffel, E. Dowling, and R. de Groat. Tls based exponential parameter estimation. *in* S. Van Huffel, ed., 'Recent advances in total least-squares techniques and errors-in-variables modelling', Philadelphia, USA, 1997. SIAM.

G. Cybenko. Locations of zeros of predictor polynomials. *IEEE Trans. Aut. Control*, **AC-27**(2), 235–237, February 1982.

P. De Groen and B. De Moor. The fit of a sum of exponentials to noisy data. *Computational of Applied Mathematics*, **20**, 175–187, 1987.

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.

I. S. Duff and J. A. Scott. Computing selected eigenvalues of sparse unsymmetric matrices using subspace iteration. *ACM Transactions on Mathematical Software*, **18**(2), 137–159, June 1993.

A. Edelman and H. Murakami. Polynomial roots from companion matrix eigenvalues. *Mathematics of Computation*, **64**(210), 763–776, 1995.

R. D. Fierro and J. R. Bunch. Perturbation theory for orthogonal projection methods with applications to least-squares and total least-squares. *Linear Algebra and its Applications*, **234**, 71–96, 1996.

R. D. Fierro, G. H. Golub, P. C. Hansen, and D. P. O'Leary. Regularization by truncated total least-squares. *SIAM Journal on Scientific Computing*, **18**(4), 1233–1241, 1997.

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

M. Hanke and P. C. Hansen. Regularization methods for large-scale problems. *Surveys of Math. Industry*, **3**, 253–315, 1993.

M. Hanke, J. Nagy, and R. Plemmons. Preconditioned iterative regularization for ill-posed problems. *in* L. R. ans R. S. Varga, ed., 'Numerical Linear Algebra', pp. 141–163, Berlin, 1993. de Gruyter.

P.C. Hansen. The truncated svd as a method for regularization. *BIT*, **27**, 534–553, 1987.

P.C. Hansen. The discrete picard condition for discrete ill-posed problems. *BIT*, **30**, 658–672, 1990.

P.C. Hansen. Analysis of discrete ill-posed problems by means of the L-curve. *SIAM Review*, **34**, 561–580, 1992.

M. R. Hestenes. Pseudoinverses and conjugate gradients. *Communications of the ACM*, **18**, 40–43, 1975.

Imperial College London. ICATS, Imperial College Analysis, Testing and Software. Mechanical Engineering Department, Exibition Road, London SW7 2BX, 1995.

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.

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', pp. 705–714, Pacific Grove California, 1978.

S. Kung, K. S. Arun, and D. V. Bhaskar Rao. State-space and singular value decomposition-based approximation methods for the harmonic retrieval problem. *Opt. Soc. Amer.*, **73**, 1799–1811, December 1983.

F. Li and Y. Lu. Bias analysis for esprit-type algorithms. *IEEE Trans. Antenn. and Propag.*, **42**(3), 418–423, March 1994.

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

S. Mayrargue and J. P. Jouveau. A new application of SVD to harmonic retrieval. *in* E. F. Deprettere, ed., 'SVD and Signal Processing', pp. 467–472. North-Holland, Amsterdam, 1988.

V. Pisarenko. The retrieval of harmonics from a covariance function. *Geophys. J. Roy. Astron. Soc.*, **33**, 347–366, 1973.

G. W. Stewart. Perturbation theory for the singular value decomposition. *in* R. J. Vaccaro, ed., 'SVD and Signal Processing II: Algorithms analysis and applications', pp. 99–109, New York, 1991. Elsevier.

K. C. Toh and L. N. Trefethen. Pseudozeros of polynomials ans pseudospectra of companion matrices. *Numerische Mathematik*, **68**, 403–425, 1994.

L. N. Trefethen and D Bau. *Numerical Linear Algebra*. SIAM, Philadelphia, USA, 1997.

A. Van der Sluis. The convergence behaviour of conjugate gradients and ritz values in various circumstances. *in* R. Beauwens and P. de Groen, eds, 'Iterative Methods in Linear Algebra', pp. 49–66, New York, 1992. Elsevier.

A. Van der Sluis and H. Van der Vorst. Sirt and cg-type methods for the iterative solution of least squares problems. *Linear Algebra and its Applications*, **130**, 257–302, 1990.

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.

S. Van Huffel, H. Chen, C. Decanniere, and P. Van Hecke . Improved total least squares based algorithm for time-domain nmr data quantification. *J. Magnet. Resonance A*, **110**, 228–237, 1994.

J. H. Wilkinson. *The Algebraic Eigenvalue Problem*. Oxford University Press, Oxford, Great Britain, 1965.

H. P. Zeiger and A. J. Mac-Ewen. Approximate linear realization of given dimension via ho's algorithm. *IEEE Trans.. Aut. Control*, **AC-19**, 153–153, 1974.