Fast Identification of Impulse Response Modes via Krylov Space Methods[1]

Daniel L. Boley[2], Franklin T. Luk[3], David Vandevoorde[3]

Abstract

Consider an underlying signal which is a sum of r exponentials plus noise. We present a novel combination of fast techniques which enables us to determine all the underlying modes in only $O(r^2)$ operations, while filtering out most of the noise. Almost all previously known methods applied to noisy signals require at least $O(r^3)$ operations.

1 Introduction

Let $\{h_k\}_{k=1}^{\infty}$ denote a complex-valued signal, and let H represent the associated infinite Hankel matrix whose (i,j)-element is defined by $H_{ij} = h_{i+j-1}$:

$$H = \begin{pmatrix} h_1 & h_2 & h_3 & h_4 & \cdot^{\cdot} \\ h_2 & h_3 & h_4 & h_5 & \cdot^{\cdot} \\ h_3 & h_4 & h_5 & h_6 & \cdot^{\cdot} \\ h_4 & h_5 & h_6 & h_7 & \cdot^{\cdot} \\ \cdot^{\cdot} & \cdot^{\cdot} & \cdot^{\cdot} & \cdot^{\cdot} & \cdot^{\cdot} \end{pmatrix} \tag{1}$$

This matrix is symmetric (not Hermitian if complex):

$$H^T = H.$$

Throughout this paper, the notation M^T denotes the transpose of M, as distinct from the more usual conjugate transpose denoted by M^H. Suppose that the underlying signal is a sum of r exponentials, i.e., for $k = 1, 2, \ldots,$

$$h_k = \sum_{i=1}^{r} \lambda_i^k d_i, \tag{2}$$

where the λ_i's are distinct complex numbers. Then the Hankel matrix H will have rank r; in general, the value of r is not known. The matrix H admits the factorization:

$$H = V^T D V,$$

where D is diagonal and V is Vandermonde:

$$D \triangleq \mathrm{diag}(d_1, d_2, \ldots, d_r)$$

and

$$V \triangleq \begin{pmatrix} 1 & \lambda_1 & \lambda_1^2 & \lambda_1^3 & \cdots \\ 1 & \lambda_2 & \lambda_2^2 & \lambda_2^3 & \cdots \\ 1 & \lambda_3 & \lambda_3^2 & \lambda_3^3 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \\ 1 & \lambda_r & \lambda_r^2 & \lambda_r^3 & \cdots \end{pmatrix}. \tag{3}$$

We stress that a diagonal decomposition is possible only if the λ_j's are distinct. Vandevoorde [15] discusses the general case when the modes are not distinct. If this decomposition can be computed quickly, then this can be used to compute a fast decomposition of a noisy signal with the view of extracting the important "signal" modes. The purpose of this work is to indicate how a fast computation may be accomplished.

The paper is organized as follows. Section 2 describes a Krylov method that reduces the given Hankel matrix to a tridiagonal form. In Section 3 we explain how the overall computation can be performed very quickly, using only $O(r^2)$ operations. Finally, an example is presented in Section 4 to illustrate the details of our approach.

2 Krylov Sequence

The key idea behind our method is to note that the columns of H can be thought of the Krylov sequence generated by the so-called "shift-up" matrix.

Assume that the matrix H of (1) has rank r. By a theorem of Gantmacher [7, vol. 2, p. 207], the signal satisfies a recurrence relation of length r:

$$h_k = a_{r-1} h_{k-1} + a_{r-2} h_{k-2} + \cdots + a_0 h_{k-r}, \tag{4}$$

which generates the entire signal once the r initial values $\{h_1, h_2, \ldots, h_r\}$ are fixed. The recurrence (4) is a difference equation which can be used to solve for the vector:

$$\mathbf{a} = (a_0, a_1, \cdots, a_{r-1})^T,$$

after the next r values $\{h_{r+1}, h_{r+2}, \ldots, h_{2r}\}$ have become known.

Let C denote the companion matrix corresponding to the polynomial:

$$p(\lambda) \triangleq \lambda^r - a_{r-1}\lambda^{r-1} - \cdots - a_1\lambda - a_0; \tag{5}$$

[1]This research was partially supported by NSF Grants CCR-9405380 & CCR-9628786

[2]Department of Computer Science and Engineering, University of Minnesota, Minneapolis MN 55455, USA

[3]Department of Computer Science, Rensselaer Polytechnic Institute, Troy NY 12180, USA

that is,

$$C \triangleq \begin{pmatrix} 0 & 1 & & & & \\ & 0 & 1 & & & \\ & & \ddots & \ddots & & \\ & & & \ddots & \ddots & \\ & & & & 0 & 1 \\ a_0 & a_1 & a_2 & \cdots & a_{r-2} & a_{r-1} \end{pmatrix}. \quad (6)$$

We show that the first r rows of H can be regarded as a Krylov sequence generated by C. Let

$$\mathbf{h}_k \triangleq \begin{pmatrix} h_k \\ h_{k+1} \\ \vdots \\ h_{k+r-1} \end{pmatrix} \quad (7)$$

denote the first r entries in the k-th column of H; e.g.,

$$\mathbf{h}_1 \triangleq \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ h_r \end{pmatrix}.$$

The first r rows of H can be written as

$$H_{1:r,1:\infty} = (\, \mathbf{h}_1 \quad \mathbf{h}_2 \quad \mathbf{h}_3 \quad \cdots \,);$$

so

$$H_{1:r,1:\infty} = (\, \mathbf{h}_1 \quad C\mathbf{h}_1 \quad C^2\mathbf{h}_1 \quad \cdots \,). \quad (8)$$

Since every row is a linear combination of the first r rows, we have that (8) implies

$$H_{1:\infty,1:\infty} = \left(\, \mathbf{h}_1^{(\infty)} \quad Z\mathbf{h}_1^{(\infty)} \quad Z^2\mathbf{h}_1^{(\infty)} \quad \cdots \right), \quad (9)$$

where $\mathbf{h}_1^{(\infty)}$ denotes the first column of the infinite Hankel matrix H, and

$$Z \triangleq \begin{pmatrix} 0 & 1 & & & & & \cdots \\ & 0 & 1 & & & & \cdots \\ & & \ddots & \ddots & & & \cdots \\ & & & \ddots & \ddots & & \cdots \\ & & & & 0 & 1 & \cdots \\ & & & & & 0 & \ddots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

is the "shift-up" matrix. The consequence of (9) is that the $(r+1)$-dimensional Krylov space generated by expanding (8) by r steps can be computed by using (9) and shifting in the entries.

Our algorithms will be based on the application of the Lanczos algorithm to the companion matrix C and the initial vector $\mathbf{h}_1$, yielding the expansion:

$$\begin{cases} CX &= XT \\ C^TY &= Y\widetilde{T}, \end{cases} \quad (10)$$

where the matrices T and $\widetilde{T}$ are tridiagonal and the matrix Y^TX is diagonal. The first column of X is exactly $\mathbf{h}_1$ and the first column of Y is chosen to be

$$\mathbf{e}_1 \triangleq (1, 0, \cdots, 0)^T.$$

The subsequent columns $\mathbf{x}_k$ of X are generated by the recurrence, for $k = 1, 2, \cdots$:

$$t_{k,k+1} \cdot \mathbf{x}_{k+1} = C\mathbf{x}_k - t_{k,k} \cdot \mathbf{x}_k - t_{k,k-1} \cdot \mathbf{x}_{k-1}, \quad (11)$$

where $t_{k,k}$ and $t_{k,k-1}$ are scalars computed to enforce the biorthogonality conditions :

$$\begin{cases} \mathbf{y}_k^T\mathbf{x}_{k+1} &= 0 \\ \mathbf{y}_{k-1}^T\mathbf{x}_{k+1} &= 0, \end{cases}$$

and $t_{k,k+1}$ is an arbitrary scaling factor, usually set so that the resulting vector $\mathbf{x}_{k+1}$ has unit length. These scalars are assembled into the tridiagonal matrix T. The columns $\mathbf{y}_k$ of Y are generated by an analogous recurrence using C^T and generating the entries of $\widetilde{T}$. Implementation details for this standard Lanczos algorithm can be found in [8, §9.4.3].

Notice that the two relations (8) and (9) generate identical vectors, when limited to the first r entries of the first $(r+1)$ vectors. Hence we can obtain the identical Lanczos expansion using Z with starting vector $\mathbf{h}_1^{(\infty)}$, where only the first $(2r-1)$ entries of this latter vector are known. The unknown entries beyond the $(2r-1)$-st are shifted in, but do not affect the coefficients T or $\widetilde{T}$ until the $(r+1)$-st step, which is beyond the step where we wish to terminate the algorithm. Hence we can replace the C in (11) with Z. The equivalence between this algorithm and the Berlekamp-Massey algorithm, as well as a symmetrized version [13] of this algorithm, are discussed in [2]. From this equivalence, it is shown that the complexity of this algorithm is $O(r)$ flops per step, a total of $O(r^2)$ flops for the r steps we require.

This Lanczos method will be used to determine the eigenvalues of C. Suppose $\lambda_1, \lambda_2, \ldots, \lambda_r$ denote the roots of the polynomial $p(\lambda)$ of (5), which we will assume for the purpose of this paper to be simple. The general case has been treated in [15, 3]. If the expansion (10) is carried out until the size of T becomes $r \times r$, then the eigenvalues of C match those of T. The result will be mathematically equivalent to Prony's method [14], which consists of solving for the coefficients of the polynomial (5) using the Yule-Walker equations

$$(\, \mathbf{h}_1 \quad \mathbf{h}_2 \quad \cdots \quad \mathbf{h}_r \,)\, \mathbf{a} = \mathbf{h}_{r+1},$$

and then finding the roots of the polynomial (5) determined by the solution $\mathbf{a}$. However, it is well known that the roots of a polynomial can be very sensitive to the coefficients of the polynomial, especially when they are not well separated; by using the Lanczos method we are able to find those roots without forming the polynomial at all.

3 Fast Vandermonde Decomposition

We sketch our algorithm to compute the Vandermonde decomposition fast. Many of the individual pieces to the algorithms are "off-the-shelf" methods, some are more experimental, and some have received very little discussion. Most details can be found in [15].

We begin with an outline of the basic steps:

1. Use a Lanczos algorithm to generate the tridiagonal matrix T.
2. Compute the "modes" generating the signal, i.e., the eigenvalues of T.
3. Compute the diagonal matrix D:

   $$D = V^{-T} H V^{-1},$$

   where V is the Vandermonde matrix (3) generated by the eigenvalues in step 2. The diagonal structure of D follows from the theory developed in the previous section.
4. Select the most "important" rows of V and entries of D.

We fill in the main details for these steps. For step 1, we use a variant of the Lanczos algorithm discussed in the previous section. However, for a signal corrupted by noise, we must take a sample much longer than the noise-free rank of the signal (the number of modes generating the underlying noise-free signal). In this case, we can carry out the Lanczos algorithm to as many steps as desired (up to the length of the signal sample we have).

The Lanczos algorithm can suffer a breakdown situation if some column $\mathbf{x}_k$ is orthogonal to the corresponding column $\mathbf{y}_k$. In this situation, a "look-head" variety of the Lanczos has been developed in [6], where the biorthogonality conditions are relaxed to some extent, but the result is that the matrices T and $\widetilde{T}$ are no longer tridiagonal. In the context of the analysis of signals corrupted by noise, this is a rare event, but if it should happen, one possible solution is to apply the nonsymmetric Lanczos process to T itself to generate a new "T", starting with a random starting vector. The cost will still be $O(r^2)$.

Once the tridiagonal matrix has been generated, the task is to find its eigenvalues in step 2. The QR-type algorithm is a standard algorithm for the general non-Hermitian eigenproblem. It is based on the iteration:

$$\begin{cases} QR & = T^{\text{old}} - sI \\ T^{\text{new}} - sI & = RQ, \end{cases} \tag{12}$$

where Q, R are respectively unitary and upper triangular and s is a shift to accelerate convergence. Most implementations include many extras for efficiency and robustness which space does not permit us to discuss here. However, this "standard" algorithm does not preserve any tridiagonal structure present in the iterate T^{old}, unless T^{old} is Hermitian.

There are two variants of the QR-type algorithm that can be applied here that can preserve the tridiagonal structure even for non-Hermitian matrices. One possibility is the complex symmetric QR algorithm proposed in [5], for which the matrix T must be symmetrized (unless we use the symmetrized Lanczos algorithm in step 1). Even when T is real, if the signs of the corresponding superdiagonal and subdiagonal entries of T are opposite, then the symmetrized matrix will be complex. The resulting QR algorithm is a direct analog of the ordinary Hermitian QR method, but using complex orthogonal rotations and complex symmetric matrices instead of unitary rotations and Hermitian matrices, respectively. The resulting Q in (12) is "complex orthogonal," meaning

$$Q^T Q = I,$$

as opposed to "unitary," meaning

$$Q^H Q = I.$$

It follows that the symmetry and the tridiagonal structure of T^{old} are preserved.

The other option is to use the LR algorithm [16], which is based on the LU factorization without pivoting to preserve the tridiagonal structure. In this algorithm the Q in (12) is replaced by a lower triangular L computed using Gaussian elimination without pivoting. The LR algorithm can break down because of a zero pivot during the Gaussian elimination, but if a random shift is applied when this occurs, the process can still exhibit very rapid convergence. If T is real, an implicit double-shift LR algorithm can in principle be carried out in real arithmetic [16]. Both algorithms require linear time for each iteration in a manner very similar to the Hermitian analog, and the number of iterations is generally $O(r)$ in a manner very similar to the QR algorithm usually employed. So the total cost will be $O(r^2)$ for both methods. The relative merits between these alternative algorithms have not been studied in detail.

In step 3, we must find the diagonal matrix D. Because of the structure of V, the diagonal entries of D appear in the first column of the matrix DV:

$$DV = V^{-T} H.$$

This first column is the solution $\mathbf{d}$ to the Vandermonde system:

$$V^T \mathbf{d} = \mathbf{h}_1.$$

This can be solved with a fast Vandermonde solver [1], where Higham [9, p. 438] recommends ordering the eigenvalues with the Leja ordering to achieve numerical stability in the algorithm in spite of the possible ill-conditioned nature of V. A simple derivation of the fast algorithm can be found in [8, §4.6.2]. It is based on the fact that an implicit "UL" decomposition of the matrix V can be computed in $O(r^2)$ time using divided differences [10, ch. 6]. A more recent $O(r \log^2 r)$ algorithm has been proposed in [12], and all these algorithms including this last one have been extended to the case of confluent Vandermonde matrices (arising when several λ_i's coincide).

4 Analysis of a Signal

Consider a signal $\{h_k\}$ which suffers from the presence of noise. How can we recover the principal modes that generate the signal?

A popular method by Kung [11] based on the singular value decomposition (SVD) is known to be an effective method for finding modes, but it suffers from the need to carry out both an SVD and a matrix eigensolution, each costing $O(r^3)$ operations. A second popular approach is to form the Hankel matrix generated by the signal, and then proceed to find a nearby Hankel matrix of a lower rank [4]. The Vandermonde decomposition of this nearby low-rank Hankel matrix yields the parameters in (2). The method of [4] iterates until it converges to a nearby Hankel matrix. Unfortunately, this method requires the repeated use of the SVD and hence costs up to $O(r^3)$ operations per iteration.

We indicated in Section 3 how the Vandermonde decomposition can be computed quickly. An obvious way to obtain a nearby Hankel matrix of a lower rank is to set to zero all the diagonal entries in D that are smaller than a certain tolerance. Although this crude method does not always yield the best approximation, a judicious combination of this approach with other criteria can yield a good result. The biggest advantage of this technique is its low total cost of $O(r^2)$ operations, instead of the $O(r^3)$ operations per step of any SVD-based iterative approach as described in the previous paragraph.

We illustrate our fast approach here. Start with a signal generated by five modes, shown by circles on the complex plane in Figure 1, to which has been added white noise with a signal-to-noise ratio of 5.4dB. We then form the 128×128 Hankel matrix H and compute its Vandermonde decomposition:

$$H = V^T D V.$$

Figure 2 shows the absolute values of the diagonal entries of D in descending order. It turns out that select-

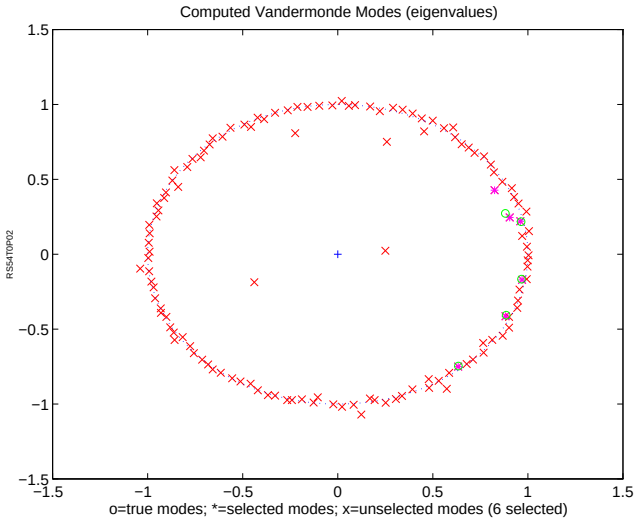

Figure 1: Original noise-free modes (o) and those computed from the noisy tridiagonal matrix discussed in Section 3 (* & x).

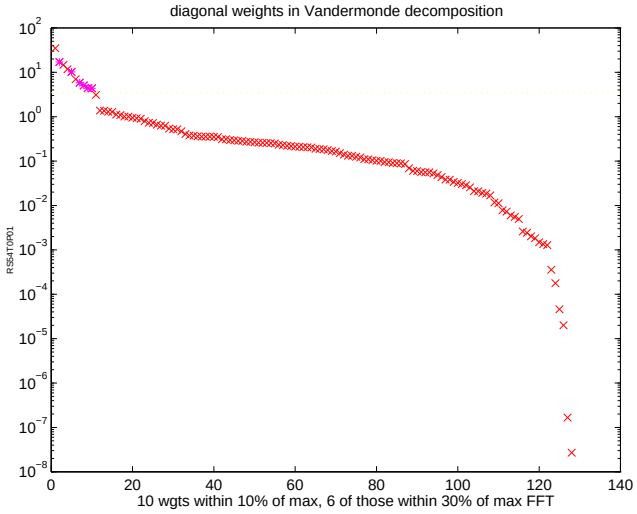

Figure 2: Diagonal entries from Vandermonde decomposition

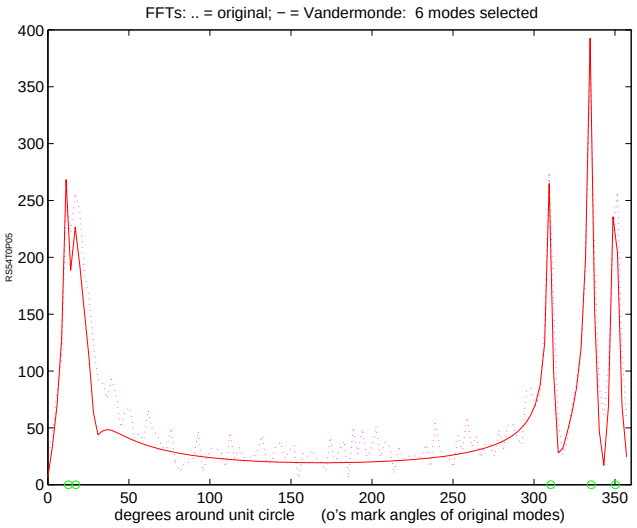

Figure 3: Discrete Fourier Transform (DFT) of the original signal (dotted) and the reconstructed reduced-order signal (solid). Small circles mark the angles corresponding the original modes.

ing the modes corresponding to the five largest values of D does not yield satisfactory results, but we can almost recover the correct modes by the following simple procedure. We select the modes corresponding to the largest entries in D (also called weights), specifically those that are within 10% of the largest entry (in absolute value); in this case ten modes were selected. Then we choose a subset of these ten using a second criterion based on the Discrete Fourier Transform (DFT) of the signal. The DFT of the original signal is shown by the dotted line in Figure 3. As most of the modes lie relatively close to the unit circle, their argument (angle on the complex plane) maps to the horizontal axis of Figure 3. In fact, we have marked the angles corresponding to the five original "unknown" modes by means of circles along the x-axis. This leads to our second criterion, viz., select those modes for which the DFT is larger than a certain threshold (in this case 30%) of the largest value in the DFT (in absolute value). This selection criterion is applied only to those modes that survived the first selection process. In this example, out of the ten modes only six survived the second selection process. These final six modes are marked by $*$'s in Figure 1, and the resulting DFT using these six modes is shown by the solid line in Figure 3. We remark that one can still distinguish the two close peaks in this DFT corresponding to the two very close original modes. Although our fast method has worked so well on this example, we should emphasize that the choice of criteria requires further study. Indeed, several more sophisticated selection criteria are presented and analyzed in [15].

References

[1] Å. Björck and V. Pereyra. Solution of Vandermonde system of equations. *Math. Comp.*, 24:893–903, 1970.

[2] D. L. Boley, T. J. Lee, and F. T. Luk. The Lanczos algorithm and Hankel matrix factorization. *Lin. Alg. & Appl.*, 172:109–133, 1992.

[3] D. L. Boley, F. T. Luk, and D. Vandevoorde. Factorization of Hankel matrices. In *Workshop on Scientific Computing 97*, March 1997. Hong Kong, to appear.

[4] J. Cadzow. Signal enhancement – a composite property mapping algorithm. *IEEE Trans. Acoust., Speech., Sig. Proc.*, 36:49–62, 1988.

[5] J. K. Cullum and R. A. Willoughby. A QL procedure for computing the eigenvalues of complex symmetric tridiagonal matrices. *SIAM J. Matrix Anal. Appl.*, 17(1):83–109, 1996.

[6] R. W. Freund, M. H. Gutknecht, and N. M. Nachtigal. An implementation of the look-ahead Lanczos algorithm for non-hermitian matrices. *SIAM J. Sci. Comp.*, 14(1):137–158, January 1993.

[7] F. R. Gantmacher. *Theory of Matrices.* Chelsea, New York, 1959.

[8] G. H. Golub and C. F. Van Loan. *Matrix Computations.* Johns Hopkins Univ. Press, 3rd edition, 1996.

[9] N. J. Higham. *Accuracy and Stability of Numerical Algorithms.* SIAM, Philadelphia, 1996.

[10] D. Kincaid and W. Cheney. *Numerical Analysis.* Brooks/Cole, 2nd edition, 1996.

[11] S. Y. Kung. A new identification and model reduction algorithm via singular value decompositions. In *Proc. 12-th Asilomar Conf. Circ., Syst. Comp.*, pages 705–714, November 1984.

[12] H. Lu. Fast solution of confluent Vandermonde linear systems. *SIAM J. Matr. Anal.*, 15:1277–1289, 1994.

[13] J. L. Phillips. The triangular decomposition of Hankel matrices. *Math. Comp.*, 25(115):599–602, 1971.

[14] R. Prony. Essai expérimental et analytique sur les lois de la dilatabilité et sur celles de la force expansive de la vapeur de l'eau et de la vapeur de l'alkool, à différentes températures. *J. de l'École Polytechnique*, 1:24–76, 1795.

[15] D. Vandevoorde. *A Fast Exponential Decomposition Algorithm and its Applications to Structured Matrices.* Ph.D. thesis, Computer Science Department, Rensselaer Polytechnic Institute, 1996.

[16] D. Watkins and L. Elsner. Chasing algorithms for the eigenvalue problem. *SIAM J. Matr. Anal.*, 12:374–384, 1991.