17 December 1999

ELSEVIER

Chemical Physics Letters 315 (1999) 135–139

CHEMICAL PHYSICS LETTERS

www.elsevier.nl/locate/cplett

Decimated signal diagonalization for obtaining the complete eigenspectra of large matrices

Dževad Belkić [a], Paul A. Dando [a], Howard S. Taylor [a,*], Jörg Main [b]

[a] *Department of Chemistry, University of Southern California, Los Angeles, CA90089, USA*
[b] *Institut für Theoretische Physik und Synergetik, Universität Stuttgart, D-70550 Stuttgart, Germany*

Received 1 September 1999; in final form 4 October 1999

Abstract

An alternative method for obtaining high and interior eigenvalues of a dense spectrum is presented. The method takes advantage of the accurate, well-tested and fully understood algorithms for the fast Fourier transform to create, in a natural manner, a 'window' containing only a small number of eigenvalues of the spectrum. The method is easy to implement, stable, efficient and accurate. © 1999 Elsevier Science B.V. All rights reserved.

1. The DSD method

In this Letter we address the problem of obtaining all the eigenvalues of a large Hermitian matrix such as occurs, for example, in the computation of the rovibrational eigenspectrum of molecules. Our starting point is the Mandelshtam–Taylor [1] version of the Wall–Neuhauser [2] filter diagonalization (FD) method which has been further developed and interpreted by Chen and Guo [3]. The alternative approach presented here will be called the decimated signal diagonalization (DSD) method. In effect, both FD and DSD introduce windowing techniques that reduce the problem of diagonalizing a large data or signal matrix, the typical elements of which involve a time auto-correlation function of the propagator associated with the matrix to be diagonalized, to one of diagonalizing a number of small data matrices (of dimension of the order of 100–200). Each small matrix represents a window of eigenvalues in the spectrum. The major difference between FD and DSD lies in the manner in which the window is introduced. In FD a basis of energy-localized wavepackets is used to window the spectrum; these wavepackets are constructed from the measured or calculated signal. DSD, on the other hand, windows the low-resolution spectrum, obtained by subjecting the same signal to a fast Fourier transform (FFT). From this windowed low-resolution spectrum, a new band-limited decimated signal (or band-limited time auto-correlation function) is created. This shortened signal can, in principle, be analyzed by any of the standard high-resolution signal processing-type techniques. This new signal is obtained from the windowed low-resolution spectrum by using the inverse FFT. The accuracy, robustness and efficiency of FFT are, by implication, features automatically shared by

* Corresponding author. Fax: +1-213-740-3972; e-mail taylor@chem1.usc.edu

0009-2614/99/$ - see front matter
PII: S0009-2614(99)01215-4

DSD. Of particular importance, however, is the fact that the band-limited decimated signal of DSD preserves, throughout the considered window, the full content of the information contained in the original signal.

As with FD, which is explained in detail in Refs. [1,2], the starting point for DSD is to create a time auto-correlation function $c_n (n = 0, 1, \ldots, N-1)$ at a sequence of times, $t_n = n\tau$, indexed by n with the constant sampling time, τ. As in time propagation followed by Fourier transform, the c_n form a signal which here is not processed by FFT but by either DSD or FD. Henceforth, we shall refer to the initial operator to be diagonalized as the 'Hamiltonian' but note that this can, in principle, be any other dynamical operator. The c_n are created by a Chebyshev iterative propagation involving repeated applications of the Hamiltonian matrix to an initial vector $\Psi \equiv \Psi(t = 0)$. When resonant, as opposed to bound, state eigenvalues are required and where the Hamiltonian has an absorbing potential, a modified Chebyshev recursion procedure [4,5] is employed to obtain the c_n. The construction of the c_n is by far the computationally most intensive part of the calculation.

To resolve all eigenvalues, ω_n, FFT would require a signal of length N, with $N\tau \geq 2\pi/\Delta\omega_{\min}$ where $\Delta\omega_{\min}$ is the *minimum* spacing between eigenvalues in the given window. For 'high-resolution' methods, such as FD and DSD, it can be shown that $N\tau \geq 4\pi/\Delta\omega_{\text{av}}$ is required [1,2], where $\Delta\omega_{\text{av}}$ is the *average* local (in the window) eigenvalue spacing. Usually $\Delta\omega_{\text{av}} > \Delta\omega_{\min}$ with the result that both FD and DSD need fewer c_n than FFT in order to resolve the eigenvalues and hence much work is saved.

In Refs [1,2] it was shown how the problem of extracting the eigenvalue spectrum from the time auto-correlation function can be recast in the form of the generalized eigenvalue problem,

$$\mathbf{U}\boldsymbol{B}_k = u_k \mathbf{S}\boldsymbol{B}_k \,. \tag{1}$$

The elements of the $(N/2)\times(N/2)$ operator matrix $\mathbf{U}$ and overlap matrix $\mathbf{S}$ in the Krylov basis set depend trivially upon the c_n [1]:

$$U_{ij} = c_{i+j+1}\,, \quad S_{ij} = c_{i+j}\,. \tag{2}$$

The eigenvalues, u_k of Eq. (1) and ω_k of the Hamiltonian, are simply related via $u_k = \exp(-\mathrm{i}\omega_k\tau)$.

There are, however, two problems with using Eq. (1): first is that $N/2$ is usually large and second that the matrices are often ill-conditioned in that their rank (i.e., the number of eigenvalues in the spectrum) is usually less than their dimension (half the signal length). To overcome this, techniques such as singular value decomposition (SVD) or Cholesky decomposition, etc., can be used but they fail for matrices of large dimension. Thus, it is essential to break the single large-dimension problem into a number of smaller problems that obtain eigenvalues in windows.

The method used in FD to create, say M, windows that cover the spectrum (i.e., spectral range or bandwidth) is described in Section 2B of Ref. [1]. It suffices to say here that FD recognizes that the U_{ij} are, in fact, matrix elements of the time propagator between Krylov basis functions. The latter are themselves powers of the propagator on an initial vector $\Psi(0)$. These basis vectors are delocalized over ω causing the U_{ij} matrix to be full. In order to set up a window, a grid is chosen in a region of the ω range; this can be a Fourier spaced grid. Formally, at the grid points, localized wavepackets are constructed from linear combinations of the Krylov basis functions. The resulting formulae for the matrix elements of the propagator operator in this local basis must now be evaluated and, of course, depend on linear combinations of the c_n. The local nature of the basis results in a small size for the new matrix elements between grid basis functions inside and outside the window. Thus, a small matrix valid for the window emerges.

The new windowing procedure employed in DSD, and known in the field of signal processing as the beam-space method [6], is sufficiently simple to be best described in words. First we apply a FFT, which scales as $N \log_2 N$, to the $\{c_n\}$. A low-resolution spectrum, which does not reveal dense eigenvalues, will result since, in general, N will not be sufficiently large to resolve them. Again we note that $N\tau \geq 2\pi/\Delta\omega_{\min}$ is needed for FFT which is usually larger than $N\tau \geq 4\pi/\Delta\omega_{\text{av}}$ required by both DSD and FD. Now, using the Fourier grid points (for only at these points is the FFT accurate) as window end

points, the low-resolution spectrum is divided into M windows, each containing at most 200 Fourier grid points. For each window separately we obtain its spectrum by the creation and processing of a new short signal of length $N_D = [N/M]$, where $[x]$ is the integer part of x. We create this signal for every window by ignoring (i.e., setting to zero) the low-resolution spectrum outside the window and shifting that remaining within the window to a position symmetric about the frequency origin. That is, if the center of the window is at $\omega = \omega_0$, then we subtract ω_0 from all frequencies in the window. This new spectrum is then subjected to an inverse FFT to create c_n^{bl}, where the superscript 'bl' refers to 'band-limited' signal.

At this stage, c_n^{bl} still consists of N points (i.e., it is the same length as the original signal, c_n). However, since the window bandwidth has been *reduced* by a factor of M, the sampling time, τ, can be *increased* by a factor of M yielding $\tau_D = M\tau$. In this way, we form a new short signal called the band-limited decimated signal, c_n^{bld}. In other words, we can recreate the spectrum inside the window *exactly* using only the set $\{c_0^{bld} = c_0^{bl}, c_1^{bld} = c_M^{bl}, c_2^{bld} = c_{2M}^{bl}, \ldots\}$ of length N_D. Crucially, *no information within the window is lost during this process;* the new signal is a function of *all* the old signal points, c_n.

In principle, we can create in this way M theoretically equivalent signals each of length $N_D = [N/M]$. These M signals are the sets $\{c_j^{bl}, c_{j+M}^{bl}, c_{j+2M}^{bl}, \ldots\}$, $(j = 0, 1, \ldots, M-1)$. In practice, for a noiseless problem, these signals give the same eigenvalues to machine accuracy. Since any of these signals is perfectly valid for the spectrum in the window of interest, we create a generalized eigenvalue problem of size $(N_D/2) \times (N_D/2)$ exactly as in Eq. (1), except with N now replaced by N_D and the original signal by one of the band-limited decimated signals discussed above. The matrices are, in general, still ill-conditioned but, since their dimension is small, SVD type procedures are effective and the methods described in Section 2B of Ref. [1] can be employed.

The required eigenvalues, ω_k, in the window are then obtained from $u_k = \exp(-iM\omega_k\tau)$ and the central frequency, ω_0, added to the real part of each ω_k to obtain the final result. Note that Eq. (12) of Ref. [1] for the amplitudes, d_k, also holds in the context of the band-limited (or decimated and, therefore, DSD) signal. In DSD this equation gives excellent results, while for FD an averaging procedure is needed [1]. Note also that no matrix element evaluation is needed in DSD as it is in FD; the FFT and inverse FFT actually replace this and are themselves very robust and accurate. An exact analytic formula for the c_n^{bld} in terms of the original c_n can be derived and used to replace the FFT and its inverse. However, we choose not to use this formula in favor of the well-tested and accurate FFT routines that are available in numerical libraries (e.g., NAG, IMSL).

Note that DSD can be viewed as a method which creates its window prior to processing the signal whereas FD accomplishes this during processing. As such, the band-limited signal can be used to reduce the dimension of the sets of ill-conditioned signal processing equations, that occur in any non-linear processing method, so that techniques that overcome ill-conditioning can be effective.

2. Results

As a first test of DSD, we created a signal of the form

$$c_n = \sum_{k=1}^{K} D_k \exp(-n\gamma_k\tau)\cos(nw_k\tau + \phi_k) \qquad (3)$$

generated for $n = 0, 1, \ldots, N$ with $K = 500$ and $\tau = 1$. The signal parameters were each chosen randomly with the real valued frequencies, w_k, and decay rates, γ_k, distributed uniformly in the intervals $[0, \pi]$ and $[0, 0.01]$ respectively. The amplitudes, D_k, were also chosen to be uniformly distributed in $[0, 100]$ and the phase, ϕ_k, in the interval $[0, 2\pi]$. Such a signal can be considered to be typical of the time auto-correlation function one would obtain for a dissipative system.

In Fig. 1 we compare the complex values for $\omega_k = w_k - i\gamma_k$ and $d_k = \frac{1}{2}D_k \exp(i\phi_k)$ obtained using DSD (circles) with the exact values (plus signs) for a single small frequency window with $\mathrm{Re}(\omega) \in [1.5, 1.7]$. We see that, even for $N = 2000$ (which is the minimum value of N needed to ensure that DSD can resolve all eigenvalues in the frequency interval $[-\pi, \pi]$), essentially all eigenvalues, w_k, are con-

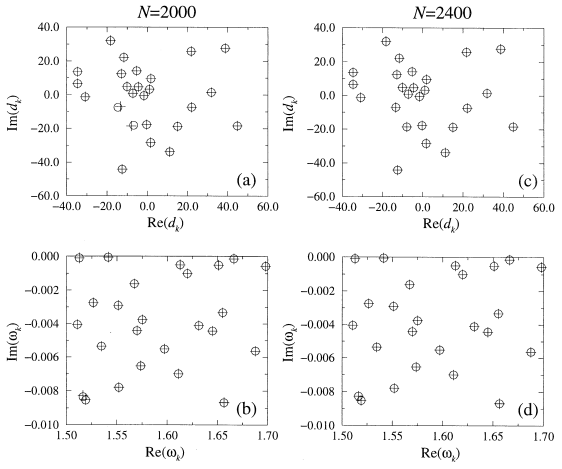

Fig. 1. Comparison between exact parameters of a model signal, Eq. (3), (denoted by +) with those calculated with DSD (denoted by ○). In panels (a) and (b) the calculated values of (complex valued) $\omega_k = w_k - i\gamma_k$ and $d_k = \frac{1}{2}D_k \exp(i\phi_k)$ are obtained using a signal with $N = 2000$; those in panels (c) and (d) with $N = 2400$.

verged and only two of the d_k show any significant deviation from the exact values (see panel (a)). When N is increased by 20%, the calculated results are indistinguishable from the exact values on the scale of the figure. A close inspection of the results reveals that even the two less accurate d_k from panel (a) are obtained to at least 5 significant figures in panel (b).

As a more realistic test of the DSD method for obtaining the eigenvalues of the Hamiltonian of a real, physical system, we also calculated the local mode doublets of SO_2 using the time auto-correlation functions computed and kindly made available to us by H. Guo [7]. Note that, in this case, the time auto-correlation functions are both real and time-symmetric: $c_{-n} = c_n$, $n = 0, \ldots, N-1$. In FD, this symmetry is utilized to construct a real symmetric analogue of the generalized eigenvalue problem, Eq. (1) [1]. In DSD, the situation is slightly different: the band-limiting process is not time-invariant and results, in general, in complex values for the c_n^{bld} and hence also for the matrix elements in Eq. (1). This admitted drawback is, fortunately, a minor one; since the matrices themselves are small, the additional computational effort required to solve Eq. (1) with complex arithmetic remains insignificant when compared with the calculation of the time auto-correla-

Table 1
Comparison between calculated energies in cm^{-1} for local mode doublets of SO_2. Calculations: DSD (decimated signal diagonalization, this Letter) and FD (filter diagonalization, H. Guo, Ref. [7] and private communication). Calculated energies were obtained using $N = 180\,000$ values of the auto-correlation function c_n generated by the Chebyshev wavepacket propagation method and kindly provided to us by H. Guo. The labels n_1, n_2 and n_3 denote the quantum numbers for symmetric stretch, bend and antisymmetric stretch, respectively

n_1	n_2	n_3	Parity	DSD	FD
9	0	0	even	10102.41	10102.41
8	0	1	odd	10252.16	10252.16
10	0	0	even	11188.29	11188.29
9	0	1	odd	11327.15	11327.15
11	0	0	even	12266.73	12266.73
10	0	1	odd	12395.30	12395.30
12	0	0	even	13337.64	13337.64
11	0	1	odd	13455.26	13455.26
13	0	0	even	14400.87	14400.87
12	0	1	odd	14506.83	14506.83
14	0	0	even	15456.27	15456.27
13	0	1	odd	15549.76	15549.76
15	0	0	even	16503.49	16503.49
14	0	1	odd	16583.68	16583.68
16	0	0	even	17542.06	17542.06
15	0	1	odd	17608.16	17608.16
17	0	0	even	18571.17	18571.17
16	0	1	odd	18622.48	18622.48
18	0	0	even	19589.48	19589.49
17	0	1	odd	19626.62	19626.62
19	0	0	even	20594.89	20594.89
18	0	1	odd	20618.60	20618.60
20	0	0	even	21584.79	21584.79
19	0	1	odd	21597.97	21597.97
21	0	0	even	22557.23	22557.23
20	0	1	odd	22564.31	22564.31
22	0	0	even	23511.93	23511.93
21	0	1	odd	23514.84	23514.84
23	0	0	even	24449.46	24449.47
22	0	1	odd	24450.94	24450.94

tion function itself. However, we note that to obtain accurately all of the energies presented in Table 1, it is necessary to utilize the 'negative time' information in the computed c_n. This can be achieved efficiently either by including all $2N-1$ of these c_n in the construction of the low-resolution spectrum by standard FFT or by replacing this step with a fast cosine Fourier transform. Note, however, that the *inverse* FFT used to compute the band-limited signal from the windowed low-resolution spectrum cannot be replaced in this manner.

In Table 1 we compare the energies obtained using DSD with those using FD presented in Ref. [7]. We see that DSD reproduces the FD results almost exactly, with the only exception being a small difference of $0.01\,\mathrm{cm}^{-1}$ for level $(23,0,0)$.

3. Conclusion

The problem of obtaining high and interior eigenvalues of a dense spectrum is studied and an alternative method called decimated signal diagonalization (DSD) is proposed. The DSD method creates, in a natural way, a window in the low-resolution spectrum by taking advantage of the well-known, accurate and robust fast Fourier transform algorithms that are widely available and well-tested.

Acknowledgements

This work was undertaken with the partial support of the DOE (Contract number 53–4815–248), the National Science Foundation (NSF Grant number PHY–9802534) and the donors of the Petroleum Research Fund. The authors thank H. Guo for supplying his signal for SO_2, together with the associated eigenvalues computed by means of FD. V.A. Mandelshtam and H. Guo are thanked for their critical reading of the manuscript. J.M. thanks H.S. Taylor for his kind hospitality during his visit to the University of Southern California where this work was initiated.

References

[1] V.A. Mandelshtam, H.S. Taylor, J. Chem. Phys. 107 (1997) 6756; see also erratum, ibid. 109 (1998) 4128.
[2] M.R. Wall, D. Neuhauser, J. Chem. Phys. 102 (1995) 8011.
[3] R. Chen, H. Guo, J. Chem. Phys. 111 (1999) 464.
[4] V.A. Mandelshtam, H.S. Taylor, J. Chem. Phys. 103 (1995) 2903.
[5] T.P. Grozdanov, V.A. Mandelshtam, H.S. Taylor, J. Chem. Phys. 103 (1995) 7990.
[6] S.D. Silverstein, M.D. Zoltowski, Digital Signal Processing 1 (1991) 161.
[7] G. Ma, H. Guo, J. Chem. Phys. 111 (1999) 4032.