Krylov-space algorithms for time-dependent Hartree–Fock and density functional computations

Vladimir Chernyak, Michael F. Schulz, and Shaul Mukamel
Department of Chemistry and Rochester Theory Center for Optical Science and Engineering, University of Rochester, P.O. Box 270216, Rochester, New York 14627-0216

Sergei Tretiak
Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545

Eugene V. Tsiper
Department of Physics, State University of New York at Stony Brook, Stony Brook, New York 11794-3800

(Received 28 January 2000; accepted 7 April 2000)

A fast, low memory cost, Krylov-space-based algorithm is proposed for the diagonalization of large Hamiltonian matrices required in time-dependent Hartree–Fock (TDHF) and adiabatic time-dependent density-functional theory (TDDFT) computations of electronic excitations. A deflection procedure based on the symplectic structure of the TDHF equations is introduced and its capability to find higher eigenmodes of the linearized TDHF operator for a given numerical accuracy is demonstrated. The algorithm may be immediately applied to the formally-identical adiabatic TDDFT equations. © *2000 American Institute of Physics.* [S0021-9606(00)30425-1]

I. INTRODUCTION

The prediction of photophysical and photochemical processes in molecules requires accurate computations of excited state surfaces.[1,2] Excited-states computations are numerically much more expensive compared to their ground-state counterparts. However, the first-principles modeling of excited state dynamics requires substantially reduced information on the excited state many-electron wave functions. Keeping redundant information greatly increases the numerical effort in standard *ab initio* computations. It is, therefore, desirable to develop approaches aimed at the direct computation of relevant quantities, eliminating unnecessary information from the outset.[3,4]

The quantities of interest, i.e., excited-state adiabatic surfaces and nonadiabatic coupling terms can be obtained from the purely electronic optical response functions calculated for different but fixed molecular geometries. Two approaches are widely used for the direct computation of the electronic response, while avoiding the explicit calculation of excited states: the time-dependent variational principle (TDVP),[5,6] and the time-dependent density-functional theory (TDDFT)[7–11] in the Kohn–Sham (KS) form.[12,13] In both cases one follows the dynamics of a certain reduced set of parameters representing the system driven by an external field. In the TDVP these parameters describe a trial many-body wave function whereas in the TDDFT they describe a set of KS orbitals.

The time-dependent Hartree–Fock (TDHF) approach provides a powerful tool for studying the optical response of conjugated molecules.[14–16] The TDHF equations are based on the TDVP where the trial wave functions belong to the space M of single Slater determinants.[5,15] Since the TDDFT can be formulated as the dynamics of a single Slater determinant based on the KS orbitals, the two approaches follow the dynamics of a similar quantity: a single Slater determinant that can be unambiguously described by an idempotent single-electron density matrix ρ (with $\rho^2=\rho$).[17,18] The TDHF and TDDFT yield different dynamical equations for $\rho(\tau)$. A conceptual difference between the two is related to the interpretation of $\rho(\tau)$. In the TDHF, $\rho(\tau)$ is viewed as an approximation for the actual single-electron density matrix,[5] whereas in the TDDFT $\rho(\tau)$ is an auxiliary quantity constrained to reproduce the correct electronic charge distribution at all times.[12,13] The TDDFT is formally exact. However, in practice it yields only approximate results since exact expressions for the exchange-correlation energy $E_{\mathrm{xc}}[n(\mathbf{r})]$ and the corresponding potential $v_{\mathrm{xc}}(\mathbf{r},[n])$ in the KS scheme are not available. An advantage of the TDHF approach is that $\rho(\tau)$ provides not only the electron charge distribution but the optically-induced coherences (changes in chemical bond order) as well. The latter have been shown to be essential for understanding optical properties of conjugated molecules and for the first-principles derivation of simple models for the photoinduced dynamics (e.g., the Frenkel–exciton model).[16] A close resemblance between the TDHF and the TDDFT (especially its adiabatic version) has been established recently using a formulation of the KS density functional theory (DFT) based on the density matrix ρ rather than on the KS orbitals.[6] This formal similarity makes it possible to apply the same algorithm for solving the equations for the optical response in both cases.

The computational bottleneck in applying the TDHF and TDDFT to large molecules[7,19] is the diagonalization of a large $(N^2/4)\times(N^2/4)$ Hamiltonian matrix L, N being the number of single-electron orbitals.[15,20] This difficulty can be overcome by applying fast Krylov-space algorithms for the diagonalization of large matrices.[21,22] The Krylov-space is defined as the space spanned on the vectors $L^n v_0$ with $n=0,1,\ldots,N'$, v_0 being some initial vector. In this article we discuss three types of Krylov-space based algorithms. The Lanczos algorithm computes effectively the lowest eigenvalue and the corresponding eigenvector of a large matrix.[21]

Since the matrices L that need to be diagonalized in the TDHF or adiabatic TDDFT approaches are non-Hermitian, the standard Lanczos algorithm is not applicable, and the oblique Lanczos algorithm (OLA) should be used.[21] The second, Davidson's algorithm (DA), finds several lowest-frequency eigenmodes of the Hamiltonian matrices which appear in the TDHF and adiabatic TDDFT which are ''Hermitian'' with respect to an antisymmetric ''scalar product.''[23–25] A third method for computing the lowest frequency eigenmode of a large Hamiltonian matrix is based on the iterative density matrix spectral moments algorithm (IDSMA).[15,20] All three algorithms show $\sim N^2$ scaling of memory and $\sim N^3$ of computational time, resulting from an $N\times N$ matrix multiplications. However, the scaling prefactors are different. The Davidson type algorithms, especially the recently improved versions,[7,26,27] are extremely fast but memory-intensive, since one needs to keep all the previous iterations for the eigenmodes throughout the iteration procedure. Consider for example the computation of the lowest eigenmode of a matrix using the Davidson iteration in a 200 dimension Krylov space (default maximum dimension in the GAUSSIAN 98).[26] To improve the accuracy we need to calculate the 201st trial Krylov vector, which should be orthogonal to all others. This requires the storing of all previous 200 vectors. On the other hand, to compute the 201st vector in the oblique Lanczos procedure we only need the 200th and the 199th vectors: by orthogonalizing the 201st to the 200th and 199th, it automatically becomes orthogonal to all previous vectors. We only need to keep in memory two vectors rather than 200. This is the memory advantage of Lanczos over Davidson's. However, the Lanczos may require a 400 Krylov-space dimension to obtain an approximate eigenvalue with the same accuracy as Davidson's in a 200 space-dimension, Davidson's algorithm thus converges faster and requires generally less computational effort than the oblique Lanczos.

The OLA is slower but much less memory consuming compared to the Davidson algorithm. In addition it is not stable for some initial vectors v_0 and it should be restarted with a different value of v_0 once it diverges. The IDSMA is stable, has low memory requirements, but is 2–4 times slower than the Lanczos. None of these three algorithms is, therefore, a universal method of choice, and they could all be most suitable for specific applications.

In this article we extend the OLA and IDSMA to obtain a low memory cost algorithm for computing the lowest few eigenmodes. Since both algorithms converge to the lowest eigenmode, the higher eigenmodes can be obtained successively by finding the lowest mode in the subspace orthogonal to that spanned by the lower modes already found. Our previous applications of this orthogonalization procedure (using the IDSMA and the TDHF equations) showed that the procedure is unstable, leading to the accumulation of numerical error for the higher modes. To overcome this difficulty we implement here a deflection procedure that involves the antisymmetric ''scalar product'' (known as the symplectic structure).[5]

Numerical applications are made for the TDHF, but our results apply to the TDDFT as well. The oblique Lanczos algorithm with the deflection procedure (OLA-D) will be compared with the IDSMA with the orthogonalization procedure (IDSMA-O). We find the former to be superior. We argue that this reflects the superiority of the deflection over the orthogonalization method for the higher eigenvalues, rather than the advantages of the OLA compared to the IDSMA for the lowest mode.

This article is organized as follows. In Sec. II we introduce the deflection procedure. In Sec. III we review the linear problem that arises in solving the linearized TDHF equations and briefly discuss the applicability of the Krylov-space-based algorithms. In Sec. IV we present numerical calculations for two molecules, the $NC_5H_6^+$ protonated Schiff base, and poly (p-phenylene) vinylene (PPV) with five phenyl rings. Details of the implementation of the oblique Lanczos algorithm to the linear TDHF problem are given in Appendices A–C.

II. COMPUTING SEVERAL LOWEST EIGENMODES: THE DEFLECTION PROCEDURE

In this section we describe the deflection procedure for computing several lowest eigenmodes of the Hamiltonian matrix L. The calculation makes use of the symplectic structure, defined as an antisymmetric scalar product in the space of the TDHF modes:[5,15]

$$\langle \xi,\gamma\rangle \equiv \mathrm{Tr}\{\gamma^+[\xi,\bar{\rho}]\}, \tag{2.1}$$

where $\bar{\rho}$ is the ground state single-electron density matrix. L is a J-Hermitian matrix: for two eigenmodes ξ_μ and ξ_ν of L with eigenvalues $\Omega_\mu\neq\Omega_\nu$ we have $\langle\xi_\mu,\xi_\nu\rangle=0$. L also has a property that if $L\xi_\nu=\Omega_\nu\xi_\nu$ then $L\xi_{-\nu}=-\Omega_\nu\xi_{-\nu}$ where $\xi_{-\nu}\equiv\xi_\nu^+$. This implies that once the lowest pair of eigenmodes ξ_1,ξ_1^+ with $\Omega_1>0$ is found, one can try to work in the subspace orthogonal to $\xi_{\pm1}$ and find the lowest eigenmode there. This procedure is, however, numerically unstable.

The deflection procedure described below is stable. When applied to a Hamiltonian matrix L it can be formulated as follows: Suppose we have found the j lowest eigenmodes $\xi_{\pm1},\xi_{\pm2},\ldots,\xi_{\pm j}$. We introduce the deflected operator L_d

$$L_d\xi \equiv L\xi+\sum_{\nu=1}^{j}\triangle\{\xi_\nu\langle\xi,\xi_\nu\rangle-\xi_\nu^+\langle\xi,\xi_\nu^+\rangle\}, \tag{2.2}$$

where the modes ξ_ν are normalized: $\langle\xi_\nu,\xi_\nu\rangle=1$ for $\nu>0$. The operator L_d has the same eigenmodes as L, however the eigenvalues of $\xi_{\pm\nu}$ for $\nu=1,\ldots,j$ are shifted: $\Omega_{\pm\nu}^{(d)}=\pm(\Omega_\nu+\triangle)$. The next pair of eigenmodes of L, $\xi_{\pm(j+1)}$, thus corresponds to the lowest pair of L_d, provided $\triangle$ is large enough. This pair can be found by applying either the OLA or IDSMA to L_d.

Both the deflection and the orthogonalization methods are based on the compatibility of the operator L with the symplectic structure.[5] The fact that the former is stable and the latter is not can be rationalized as follows. Consider first the orthogonalization method. Suppose we have found the lowest mode. To obtain the next mode, we try to work in the subspace orthogonal to the first mode. By looking for the lowest mode in this subspace (either IDSMA or Lanczos) we

start to approach the next mode. However, because of numerical errors, the vector has some residual component of the lowest mode and the iterative procedure eventually converges to the lowest mode. In practice we found a quasiconvergence: iterations lead to some neighborhood of the next mode, the system stays in this neighborhood for a while but finally converges to the lowest mode. The accuracy of the next mode is therefore determined by the size of the quasiconvergence region, and it decreases rapidly for higher modes. In the deflection method, on the other hand, we find the lowest mode of an operator which has been deflected from the ''correct'' one. There are no convergence problems, since we are always searching for the lowest mode of this deflected operator. The higher modes are therefore found with about the same accuracy as the deflection operator, in practice with machine precision. The superiority of the deflection over the orthogonalization method is clearly demonstrated in Sec. IV.

III. KRYLOV SPACE ALGORITHMS FOR THE TDHF EIGENVALUE PROBLEM

The TDHF equations of motion for the time-dependent single-electron density matrix[15,5] lead to an eigenvalue problem

$$\hat{L}\xi_\nu = \Omega_\nu \xi_\nu, \tag{3.1}$$

where ξ_ν are electronic mode density matrices and Ω_ν are optical transition frequencies. The Liouville operator $\hat{L}$ is non-Hermitian ($\hat{L} \neq \hat{L}^\dagger$) and is defined by[15,5]

$$\hat{L}\xi = [F,\xi] + [V(\xi),\overline{\rho}]. \tag{3.2}$$

Here $\overline{\rho}$ is the ground state density matrix, and F is a Fock matrix of size $N = N_p + N_h$. N is the total basis set size, N_p and N_h are the numbers of particle (occupied) and hole (virtual) orbitals. The particle and hole molecular orbitals are eigenvectors of the Fock matrix, which is diagonal in the basis of these orbitals and may be obtained using the Hartree–Fock iterative procedure.[15] The Coulomb operator V obeys

$$(\xi, V(\eta)) = (V(\xi), \eta), \tag{3.3}$$

where the scalar product of two arbitrary matrices η and ξ is defined as $(\xi,\eta) \equiv \mathrm{Tr}(\xi^\dagger \eta)$.

The dimensionality of the Liouville space $\mathcal{L}$ is $M = N^2$. A physical subspace $\mathcal{L}_{\mathrm{ph}}$ of all matrices limited to only particle–hole and hole–particle orbital pairs has dimensionality $M_2 = 2N_pN_h$ and is an invariant subspace of the operator $\hat{L}$.[5] The electronic oscillators algebra developed in Ref. 5 reduces the number of variables to the $\mathcal{L}_{\mathrm{ph}}$ subspace. For any matrix ξ, the commutator $[\xi,\overline{\rho}]$ belongs to $\mathcal{L}_{\mathrm{ph}}$, and the double commutator $[[\xi,\overline{\rho}],\overline{\rho}]$ is a projector onto $\mathcal{L}_{\mathrm{ph}}$. The following relation also holds in $\mathcal{L}_{\mathrm{ph}}$:

$$(\xi,\eta) \equiv ([\xi,\overline{\rho}],[\eta,\overline{\rho}]), \tag{3.4}$$

and the operator adjoint to $\hat{L}$ is

$$\hat{L}^\dagger \xi = [\xi,F] + V([\xi,\overline{\rho}]). \tag{3.5}$$

Using Eqs. (3.3) and (3.4) it is straightforward to prove Hermiticity with respect to our scalar product, i.e., $(L^\dagger \xi,\eta) \equiv (\xi,\hat{L}\eta)$.

The spectrum of $\hat{L}$ (and $\hat{L}^\dagger$) consists of $M_2/2$ pairs of conjugated eigenvectors with eigenfrequencies $\pm\Omega_\nu$. Indeed, Eq. (3.1) is equivalent to

$$\hat{L}\tilde{\xi}_\nu = -\Omega_\nu \tilde{\xi}_\nu, \tag{3.6}$$

$$\hat{L}^\dagger[\xi_\nu,\overline{\rho}] = \Omega_\nu[\xi,\overline{\rho}], \tag{3.7}$$

$$\hat{L}^\dagger[\tilde{\xi}_\nu,\overline{\rho}] = -\Omega_\nu[\tilde{\xi}_\nu,\overline{\rho}]. \tag{3.8}$$

Only the few lowest excited states are necessary to compute visible and UV electronic spectra.[14,16] The Lanczos algorithm is the method of choice for this purpose when the matrix is sparse. It is important to note that even though the matrix $\hat{L}$ is, in general, not sparse, it shares one useful property with sparse matrices: the evaluation of its action on a vector does not require storing all matrix elements in memory, and can be readily done using Eq. (3.2). We have employed the following novel method for computing the product $\hat{L}\xi$. The part of $\hat{L}$ which contains the Fock matrix is calculated in the Hartree–Fock molecular orbitals basis, in which the Fock matrix is diagonal. However, the evaluation of $V(\rho)$ is carried out by converting ρ into the atomic-orbital basis, in which V is sparse, and then transforming the results back to the molecular-orbital basis. We note that in the IDSMA, which is also a sparse matrix method, both F and V are evaluated in the atomic orbital basis.

IV. NUMERICAL RESULTS

We have tested the deflection algorithm by applying it to two molecules, the $NC_5H_6^+$ protonated Schiff base and poly (p-phenylene vinylene) (PPV) with five phenyl rings. We used the INDO/S Hamiltonian for our calculations, obtained from the ZINDO code.[28–30]

Figure 1 illustrates the near-exponential convergence of the oblique Lanczos algorithm with the number of iterations for PPV. We have plotted the difference between the energy at a given iteration step and the converged energy value, $(\Omega_\nu)_n - (\Omega_\nu)_\infty$, vs the number of iterations n. The two curves pertain to the lowest mode and a higher mode of excitation.

The top panel of Fig. 2 shows the oscillator strength f vs. the energy Ω, obtained for the $NC_5H_6^+$ protonated Schiff using OLA-D (20 modes, dashed), IDSMA-O (15 modes for each of the three polarization directions, dotted), and a single configuration interaction (SCI) calculation (using all 256 modes, solid line). The basis size for this molecule is 32. The three calculations virtually coincide. The middle panel shows the relative difference between the OLA-D and the SCI calculation, and the bottom panel depicts the relative difference between the IDSMA-O and the SCI. The largest relative differences occur in regions where the oscillator strength is close to zero. Both Krylov algorithms give equally good results in this test case, and compare well with the SCI.

The PPV oligomer with five phenyl rings (basis size 174) has been computed using OLA-D (50 modes) and

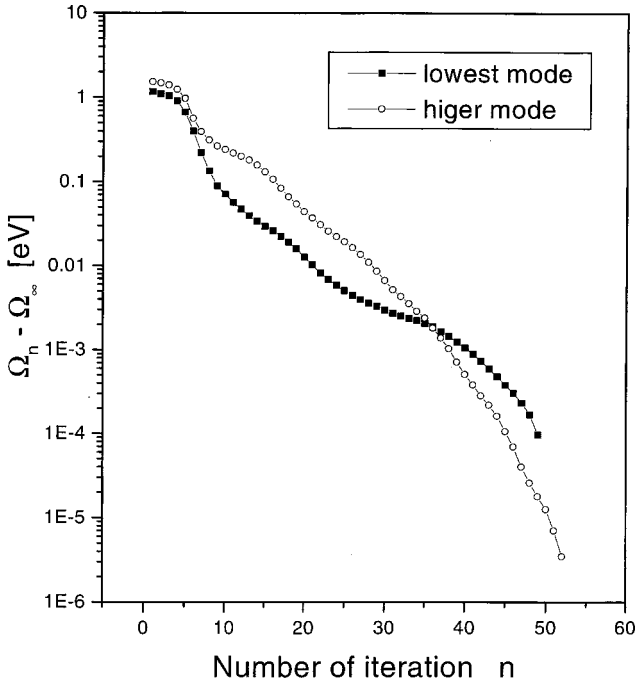

FIG. 1. Convergence of the oblique Lanczos algorithm for PPV.

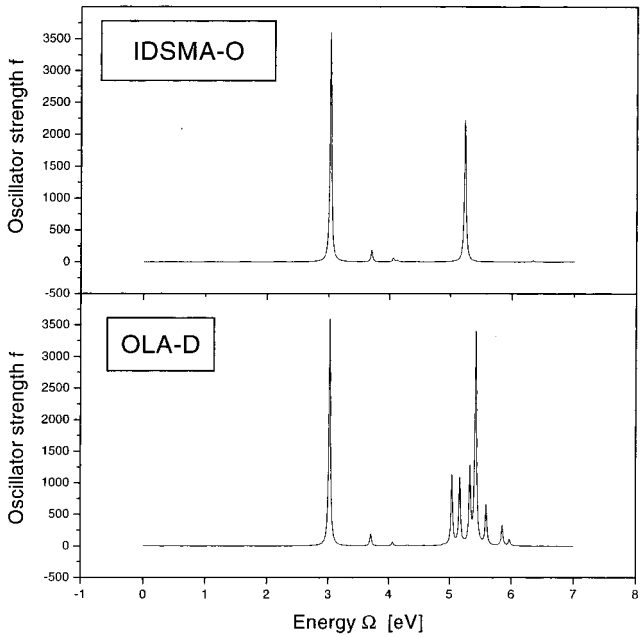

FIG. 3. Linear absorption spectrum for PPV, obtained using the IDSMA-O (top panel) and the OLA-D (bottom panel).

IDSMA-O (9 modes for each of the three polarization directions). The results shown in the top and bottom panel of Fig. 3, respectively, are very close for both algorithms for the modes below 4 eV. However, for higher energies the OLA-D code was able to resolve several closely-lying modes which were not resolved by the IDSMA-O. We attribute this primarily to the merits of the deflection compared with the orthogonalization method. The CPU time in this calculation for the IDSMA-O was approximately 1 h for each of the three runs, and 5.68 h for OLA-D at full accuracy (i.e., the algorithm runs until the energy has converged to machine precision, 14 digits). Here and below, CPU timing results are given using a single MIPS R10000 175 MHz processor on a 250 MB RAM SGI Octane workstation.

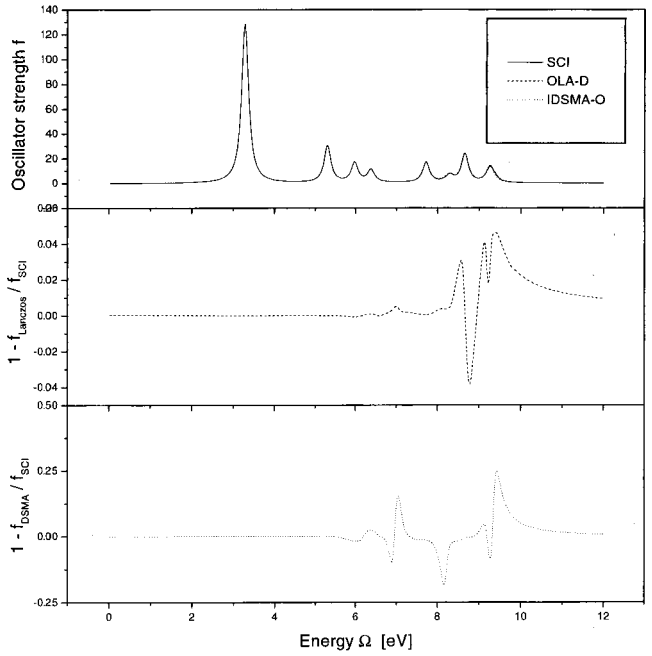

FIG. 2. Top panel: Linear absorption spectrum for the protonated Schiff base; the solid, dashed, and dotted lines represent single-CI,(SCI), OLA-D, and IDSMA-O results, respectively; the three curves are almost indistinguishable. Middle panel: Relative difference between the OLA-D and the SCI results. Bottom panel: Relative difference between the IDSMA-O and the SCI results.

The effect of reducing the accuracy of the oblique Lanczos computation is investigated for the same PPV oligomer in Fig. 4. The solid curve repeats the results obtained for full accuracy (machine precision), which is identical with the spectrum in the bottom panel of Fig. 3. The dashed line was obtained by terminating the calculation as soon as $\Omega_n - \Omega_{n-1} < 10^{-3}$ eV, where n denotes the iteration step. The reduced accuracy calculation took 1.58 h, i.e., only 28% of the more accurate calculation. However, the figure shows clear differences, particularly between 5–6 eV, where we have to resolve several closely-lying modes.

V. CONCLUSIONS

We have examined the oblique Lanczos[21] and Iterative DSMA algorithms[15,20] for the non-Hermitian eigenvalue problem that appears in the TDHF[5] and TDDFT[7] computations of electronic response functions. Although scaling of the overall memory and computational effort with system size is similar for these and analogous algorithms (e.g., Davidson's[23–25] and filter band[31,32]), the nature of targeted spectroscopic observables and available computational resources may make one of them preferable for a particular application. To illustrate this point, let us compare the resonant and static response in conjugated molecules.[19,20] The resonant response requires an accurate calculation of the excitation energies in the chosen frequency range. In this case

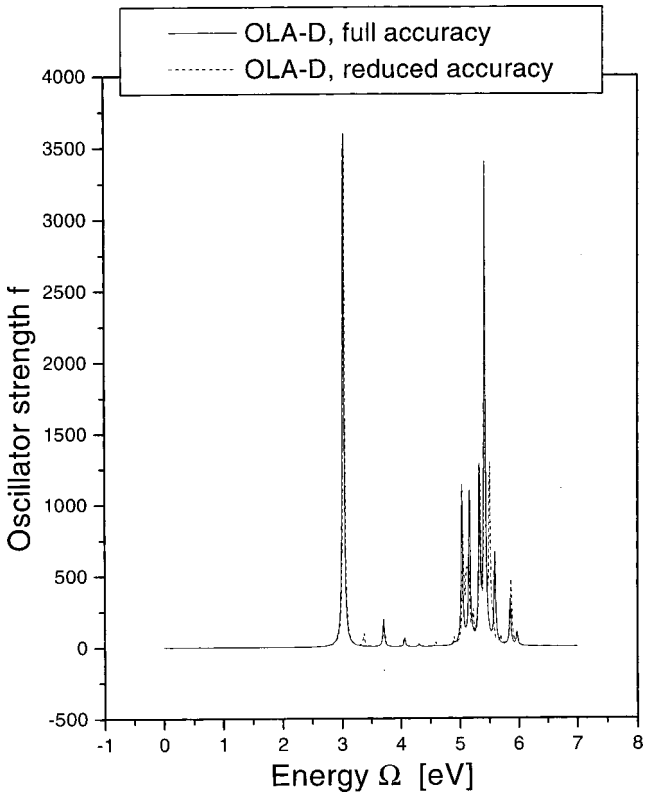

FIG. 4. OLA-D calculations for PPV, obtained with full accuracy (solid line) and reduced accuracy, as explained, in the text (dashed line).

the OLA-D approach, that provides the eigenvalues and eigenvectors of the Liouville operator with high precision, should be preferable. On the other hand, the static response does not require accurate calculations of the excitation frequencies, but rather the transitions with strong dipole moments that contribute significantly to the polarizability. OLA-D computed eigenstates will help to calculate only a part of the response coming from the covered frequency range. The DSMA algorithm that divides the response function into several dominant effective contributions from the whole spectrum should then be the method of choice.

Each diagonalization algorithm requires a separate procedure that excludes the already computed eigenmodes from the Krylov space when calculating the next eigenvector. We have demonstrated that the deflection procedure finds a set of lowest TDHF eigenmodes with more stability and accuracy compared to the orthogonalization method. The computations of electronic spectra focus on the low-frequency spectral range which can be effectively obtained using the deflection procedure. The OLA-D method allows the computation of a few hundreds of low-frequency electronic transitions in large conjugated molecules covering pretty much the entire visible-UV spectrum (1–8 eV).

Finally we note that the present article does not intend to demonstrate the largest system for which the oblique Lanczos procedure my be applied. Our PPV5 calculations simply demonstrate the comparison of OLA and DSMA. These calculations have only minor time (<2 min/state) and memory (<2 MB) requirements on an R10000 SGI workstation.

ACKNOWLEDGMENTS

The support of the National Science Foundation through Grants No. CHE-9526125 and No. PHY94-15583 is gratefully acknowledged.

APPENDIX A: LANCZOS ALGORITHM FOR HERMITIAN MATRICES

The Hermitian Lanczos algorithm finds a few lowest eigenvalues of a Hermitian matrix $\hat{H}$ by starting with an arbitrary vector v_0 and constructing linear combinations of vectors $v_m=\hat{H}^m v_0$, $m=0,1,\ldots$. The coefficients in the linear combination of v_m are found using the Ritz variational procedure[21,22] which guarantees to yield the best approximation to the lowest eigenvalue of $\hat{H}$ that belongs to a Krylov subspace. This subspace ($\mathcal{K}_m$) spanned by the vectors $v_0 \ldots v_m$, approximates an invariant subspace of $\hat{H}$ with increasing accuracy as the number of vectors is increased.

A simple recursive procedure allows building a set of orthogonal vectors w_m that span the same Krylov subspace. Finding each new vector w_{m+1} only requires the two previous vectors w_m and w_{m-1}:[22]

$$w_{m+1}=\beta_{m+1}^{-1}(\hat{H}w_m-\alpha_m w_m-\beta_m w_{m-1}). \tag{A1}$$

At each step m, the pair of coefficients α_m and β_m is chosen to preserve orthonormality of w_{m+1} with respect to w_m and w_{m-1}. The recursion Eq. (A1) ensures that w_m form an orthogonal set and that the Rayleigh–Ritz matrix $T_{ij}=(w_i,Hw_j)$ is symmetric tridiagonal, with the diagonal and subdiagonal given by the coefficients α_m and β_m, respectively.

The matrix T_{ij} can be viewed as the result of the orthogonal projection of the full matrix $\hat{H}$ onto the subspace $\mathcal{K}_m$. It can be written in matrix form

$$T_m=W_m^{\dagger}\hat{H}W_m, \tag{A2}$$

where W_m is the rectangular matrix whose columns are the vectors $w_1,\ldots,w_m$. The eigenvalues of T_m give approximations to the eigenvalues of $\hat{H}$ and the corresponding eigenvectors y give the coefficients of expansion of the eigenvector v of $\hat{H}$ in the basis of w_i, $v=Wy$. Indeed, if $T_m y=\lambda y$, then $(w_i,\hat{H}W_m y-\lambda W_m y)=0$, i.e., the residual vector is orthogonal to $\mathcal{K}_m$.

The recursive relation (A1) provides a great computational advantage to the Lanczos algorithm, making it possible to apply this technique to very large matrices, since the required memory does not grow with the number of iterations. The problem of loss of global orthogonality due to computer roundoff errors has been extensively studied[21] and is not addressed here.

APPENDIX B: LANCZOS ALGORITHM FOR NON-HERMITIAN MATRICES

Several existing variations of the Lanczos and Davidson methods are applicable to non-Hermitian eigenvalue problems.[21] The major difficulty with non-Hermitian matrices is that, in general, no variational principle exists for their

eigenvalues, and therefore the Ritz procedure is not applicable. In addition, the Lanczos recursion Eq. (A1), which is based on the Hermiticity of $\hat{H}$, does not yield an orthonormal set of vectors w_m when applied to a non-Hermitian matrix $\hat{L}$.

Lanczos has suggested a non-Hermitian algorithm that partially preserves the advantages of Hermitian Lanczos recursion Eq. (A1).[21] Instead of building an orthogonal basis of subspace $\mathcal{K}_m$, it constructs a pair of biorthogonal bases $\{v_m, w_m\}$ that span two Krylov subspaces $\mathcal{K}_m$ and $\mathcal{K}_m^\dagger$ of the operators $\hat{L}$ and $\hat{L}^\dagger$, respectively. The algorithm starts with two arbitrary vectors v_0 and w_0, normalized such that $(v_0, w_0)=1$, and uses the following recursion relations:

$$v_{m+1}=\delta_{m+1}^{*-1}(\hat{L}v_m-\alpha_m v_m-\beta_m v_{m-1}), \tag{B1}$$

$$w_{m+1}=\beta_{m+1}^{-1}(\hat{L}^\dagger w_m-\alpha_m^* w_m-\delta_m w_{m-1}), \tag{B2}$$

the coefficients α_m, β_m, and δ_m are chosen at each step m such that $(v_{m+1}, w_m)=(v_m, w_{m+1})=0$ and $(v_{m+1}, w_{m+1})=1$. As in the case of Hermitian Lanczos algorithm, this is sufficient to ensure the global biorthogonality

$$(v_m, w_n)=\delta_{mn}. \tag{B3}$$

The global biorthogonality becomes an advantage when combined with a special criterion for the ''best'' approximation to the eigenvector. Due to the lack of minimum principle, the Euclidean projection onto a subspace is not defined. With the biorthogonal basis Eq. (B3), it is natural to define an oblique projection into $\mathcal{K}_m$, orthogonal to $\mathcal{K}_m^\dagger$. This is why the algorithm is usually called an oblique Lanczos algorithm.[21]

The oblique Lanczos algorithm has a clear computational advantage since the approximations to the eigenvalues of $\hat{L}$ are given as the eigenvalues of the tridiagonal matrix

$$T=\begin{pmatrix} \alpha_1 & \beta_2 & & & \\ \delta_2 & \alpha_2 & \beta_3 & & \\ & \cdot & \cdot & \cdot & \\ & & \delta_{m-1} & \alpha_{m-1} & \beta_m \\ & & & \delta_m & \alpha_m \end{pmatrix}. \tag{B4}$$

Indeed, from Eq. (B2), the matrix elements T_{ij} are given by

$$T_{ij}=(w_i, \hat{L}v_j), \tag{B5}$$

which in matrix form reads

$$T=W_m^\dagger \hat{L} V_m. \tag{B6}$$

Here V_m and W_m are the rectangular matrices constructed of vectors v_i and w_i, respectively. In analogy with Eq. (A2), this equation can be interpreted as the oblique projection of $\hat{L}$ onto subspace $\mathcal{K}_m$, orthogonal to $\mathcal{K}_m^\dagger$.

The equation $T_m y=\lambda y$ leads to $(w_i, \hat{H}V_m y-\lambda V_m y)=0$, which means that the residual vector is orthogonal to all vectors w_i. Thus, an eigenvalue of T_m gives the approximation to the eigenvalue of $\hat{L}$, and the vector $v=Vy$ to the eigenvector of $\hat{L}$, subject to the condition that the residual vector $\hat{L}v-\lambda v$ is orthogonal to all vectors w_m. In order to construct the matrix T_m one needs to keep only two pairs of vectors in the course of the iteration. On the other hand, the oblique projection does not always exist if the subspace $\mathcal{K}_m$ has at least one vector orthogonal to $\mathcal{K}_m^\dagger$. The algorithm breaks down irreversibly in this case.

OLA has an essential advantage in memory requirements with respect to similar algorithms frequently used for non-Hermitian matrices, such as the Arnoldi or Davidson's methods. The Arnoldi algorithm deals with the non-Hermitian problem by building an orthonormal basis in $\mathcal{K}_m$ using the Gram–Schmidt orthogonalization procedure

$$v_{m+1}=h_{m+1,m}^{-1}\left(\hat{H}v_m-\sum_{i=1}^{m} h_{im}v_i\right). \tag{B7}$$

The coefficients h_{mi} form the Rayleigh matrix, i.e., the result of the projection of the matrix $\hat{H}$ onto the subspace $\mathcal{K}_m$. The eigenvalues of Rayleigh matrix provide approximations to the eigenvalues of $\hat{H}$ and the corresponding eigenvectors give the coefficients in the expansion of the corresponding eigenvectors of $\hat{H}$ in terms of v_m.

To find the coefficients h_{im} in Eq. (B7) one needs to know the scalar products of each new vector with all previously computed vectors. Therefore, the previously computed vectors must be stored in memory or hard disk, which limits the matrix size.

For a non-Hermitian case, the best approximation v in the subspace $\mathcal{K}_m$ is defined as the one that gives the smallest Euclidean norm of the residual vector $r=\hat{H}v-v|\hat{H}v|/|v|$. In other words, the best approximation is given as a Euclidean projection of the exact eigenvector onto the Krylov subspace $\mathcal{K}_m$.

The Hermitian Lanczos algorithm is mathematically the best method for approximating extreme eigenvalues, when no extra information about the matrix $\hat{H}$ is given besides the algorithm to compute the matrix-vector products. When some useful information about the internal structure of $\hat{H}$ is available, preconditioning techniques can speed up the convergence. One of the most widely used methods of this class is the Davidson algorithm[23] that utilizes the information about the diagonal elements of $\hat{H}$, and requires fewer iterations when the diagonal elements of $\hat{H}$ are dominant. The idea of the Davidson preconditioning is simple. As in the Lanczos algorithm, the solution of the eigenvalue problem is found by projecting the matrix onto a certain subspace $\mathcal{K}_m$ that expands with the number of iterations. In the Lanczos algorithm, the space $\mathcal{K}_m$ is augmented at each iteration step by the residual vector $r_m=\hat{H}w_m-\lambda w_m$. In contrast, the Davidson algorithm augments the subspace $\mathcal{K}_m$ by a different vector $\tilde{r}_m$, obtained from r_m by multiplying it by a diagonal matrix $(D-\lambda_m)^{-1}$, where D is the diagonal part of $\hat{H}$. The Davidson algorithm requires the knowledge of the entire basis of the subspace $\mathcal{K}_m$ which imposes heavy memory requirements, although it makes the calculations substantially less time-consuming.

The Davidson algorithm[27] and its variations[7] are variants of Arnoldi's method with diagonal preconditioning. Both Arnoldi's and Davidson's methods suffer from the same

memory constraints which make them hard to apply to very large matrices.

APPENDIX C: IMPLEMENTATION OF THE OBLIQUE METHOD FOR TDHF

The condition Eq. (B3) only determines the products $f_m=\beta_m\delta_m$. The absolute values of β_m and δ_m can be made equal by the proper choice of the norms of v_m and w_m. For a real matrix $\hat{L}$ the coefficients α_m, β_m, and δ_m are real. The matrix T is, in general, not symmetric, since some of the values f_m are negative. Let us consider the case when all products $f_m=\beta_m\delta_m>0$. The matrix T_m can then be made symmetric, and therefore its eigenvalues are real.

The symmetric $\pm\Omega$ form of the TDHF spectrum can be preserved in the following way: The TDHF operator $\hat{L}$ defined in Eq. (3.1) has the property that the result of its action on a symmetric matrix ξ is an antisymmetric matrix and vice versa. This means that if we choose, for example, two symmetric matrices as starting vectors v_0 and w_0, then all vectors v_{2j} and w_{2j} will be symmetric (as matrices) and all the vectors v_{2j+1} and w_{2j+1} will be antisymmetric (as matrices). The orthogonality of v_{m+1} to w_m and of w_{m+1} to v_m is thus guaranteed by symmetry, and therefore all $\alpha_m\equiv 0$.

If all f_m are positive, we obtain a real symmetric tridiagonal matrix T with zero diagonal. By rearranging the order of rows and columns of this matrix and writing all the even rows and columns first, we can transform T into the form

$$\begin{pmatrix} 0 & Z \\ Z^{\dagger} & 0 \end{pmatrix}, \tag{C1}$$

where

$$Z=\begin{pmatrix} \beta_2 & & & & & \\ \beta_3 & \beta_4 & & & & \\ & \beta_5 & \beta_6 & & & \\ & & . & . & & \\ & & & & \beta_{m-1} & \beta_m \end{pmatrix}, \tag{C2}$$

it then follows that the spectrum of T consists of pairs $\pm\Omega_\nu$. The method is found to converge exponentially fast to the lowest eigenvalue of $\hat{L}$.

Since we expect the approximations to the eigenvalues of $\hat{L}$ to be real, a negative f_m constitutes a breakdown. We found that when the starting vectors are symmetric and equal matrices ($v_0=w_0$) such breakdowns are rare. Using this initial condition, the oblique Lanczos recursion Eq. (B2) computes positive f_m step-by-step until a negative f_m is encountered. If this occurs before the accuracy reaches the desired tolerance, the program restarts with a new choice of initial vectors.

Two different choices of a starting vector were used. Each even attempt starts with a random v_0. Each odd attempt starts with the diagonal approximation, which in molecular orbital representation has one component of the vector v_0 set to 1 and all others set to 0. The component which is set to one is chosen based on the diagonal part of $\hat{L}$, that is assumed to be dominant.

The diagonal part of $\hat{L}$ in the molecular-orbital representation is dominant because the major contribution of the transition frequencies comes from the energy differences of filled and occupied molecular orbitals. Thus the diagonal approximation corresponds to the HOMO–LUMO approximation for the transition, and gives reasonable guess for the starting vector.

Implementation of the oblique Lanczos recursion Eq. (B2) requires evaluation of the vector products with both $\hat{L}$ and $\hat{L}^{\dagger}$ at each step. This task can be simplified, since $\hat{L}^{\dagger}$ can be expressed in terms of $\hat{L}$:

$$\hat{L}^{\dagger}\xi\equiv[\hat{L}[\xi,\bar{\rho}],\bar{\rho}]. \tag{C3}$$

Here Eqs. (3.2) and (3.5) and the Hartree–Fock equation $[F,\bar{\rho}]=0$ were used. If v_m and w_m are symmetric, then $[w_m,\bar{\rho}]$ is antisymmetric. Hence, one can act with $\hat{L}$ on the sum $v_m+[w_m,\bar{\rho}]$ and recover $\hat{L}v_m$ and $\hat{L}[w_m,\bar{\rho}]$ as the antisymmetric and the symmetric parts of the result, respectively, then using Eq. (C3) to obtain $\hat{L}^{\dagger}w_m$.

[1] J. Michl and V. Bonacic-Koutecky, *Electronic Aspects of Organic Photochemistry* (Willey, New York, 1990).
[2] M. Klessinger and J. Michl, *Excited States and Photochemistry of Organic Molecules* (VCH, New York, 1995).
[3] E. V. Tsiper, V. Chernyak, S. Tretiak, and S. Mukamel, Chem. Phys. Lett. **302**, 77 (1999).
[4] E. V. Tsiper, V. Chernyak, S. Tretiak, and S. Mukamel, J. Chem. Phys. **110**, 8328 (1999).
[5] V. Chernyak and S. Mukamel, J. Chem. Phys. **104**, 444 (1996).
[6] V. Chernyak and S. Mukamel, J. Chem. Phys. **111**, 4383 (1999).
[7] E. R. Stratmann, G. E. Scuseria, and M. J. Frisch, J. Chem. Phys. **109**, 8218 (1998).
[8] K. B. Wiberg, E. R. Stratmann, and M. J. Frisch, Chem. Phys. Lett. **297**, 60 (1998).
[9] C. Jamorski, M. E. Casida, and D. R. Salahub, J. Chem. Phys. **104**, 5134 (1996).
[10] M. E. Casida, in *Recent Advances in Density Functional Methods*, edited by D. P. Chong (World Scientific, Singapore, 1995), Vol. 1.
[11] M. E. Casida, in *Recent Developments and Applications of Modern Density Functional Theory, Theoretical and Computational Chemistry*, edited by J. M. Seminario (Elsevier, Amsterdam, 1996), Vol. 4.
[12] P. Hohenberg and W. Kohn, Phys. Rev. B **136**, 864 (1964).
[13] W. Kohn and L. J. Sham, Phys. Rev. A **140**, 1133 (1965).
[14] S. Mukamel, S. Tretiak, T. Wagersreiter, and V. Chernyak, Science **277**, 781 (1997).
[15] S. Tretiak, V. Chernyak, and S. Mukamel, J. Am. Chem Soc. **119**, 11408 (1997).
[16] S. Tretiak, V. Chernyak, and S. Mukamel, J. Phys. Chem. B **102**, 3310 (1998).
[17] R. McWeeny and B. T. Sutcliffe, *Methods of Molecular Quantum Mechanics* (Academic, New York, 1976).
[18] E. R. Davidson, *Reduced Density Matrices in Quantum Chemistry* (Academic, New York, 1976).
[19] S. Tretiak, V. Chernyak, and S. Mukamel, Phys. Rev. Lett. **77**, 4656 (1996).
[20] S. Tretiak, V. Chernyak, and S. Mukamel, J. Chem. Phys. **105**, 8914 (1996).
[21] Y. Saad, *Numerical Methods for Large Eigenvalue Problems* (Manchester University Press, Manchester, UK, 1992).
[22] B. N. Parlett, *Symmetric Eigenvalue Problem* (Prentice Hall, Englewood Cliffs, NJ, 1980).
[23] E. R. Davidson, J. Comput. Phys. **17**, 87 (1975).
[24] K. Hirao and H. Nakatsuji, J. Comput. Phys. **45**, 246 (1982).
[25] S. Rettrup, J. Comput. Phys. **45**, 100 (1982).
[26] GAUSSIAN 98, M. J. Frisch *et al.* (Gaussian Inc, Pittsburgh, PA, 1999).

[27] J. Olsen, H. J. A. Jensen, and P. Jørgensen, J. Comput. Phys. **74**, 265 (1988).; R. Bauernschmitt and R. Ahlrichs, Chem. Phys. Lett. **256**, 454 (1996).
[28] J. A. Pople and G. A. Segal, J. Chem. Phys. **43**, S136 (1965).; J. A. Pople, D. L. Beveridge, and P. Dobosh, *ibid.* **47**, 2026 (1967).
[29] J. Ridley and M. C. Zerner, Theor. Chim. Acta **32**, 111 (1973).
[30] M. C. Zerner, G. H. Loew, R. F. Kirchner, and U. T. Mueller-Westerhoff, J. Am. Chem. Soc. **102**, 589 (1980).
[31] V. A. Mandelshtam and H. S. Taylor, J. Chem. Phys. **106**, 5085 (1997).
[32] D. Neuhauser, J. Chem. Phys. **93**, 2611 (1990).