Fast Algorithms for Estimating the Absorption Spectrum within the Linear Response Time-dependent Density Functional Theory

Jiri Brabec, Lin Lin, Meiyue Shao, Niranjan Govind, Chao Yang, Yousef Saad, Esmond Ng

July 17, 2015

Abstract

In this paper, we present two iterative algorithms for approximating the absorption spectrum of molecules within the linear response of time-dependent density functional theory (TDDFT) framework. These methods do not attempt to compute eigenvalues or eigenvectors of the linear response matrix. They are designed to approximate the absorption spectrum as a function directly. They take advantage of the special structure of the linear response matrix. Neither method requires the linear response matrix to be constructed explicitly. They only require a procedure that performs the multiplication of the linear response matrix with a vector. These methods can also be easily modified to efficiently estimate the density of states (DOS) of the linear response matrix without computing the eigenvalues of this matrix. We show by computational experiments that the methods proposed in this paper can be much more efficient than methods that are based on the exact diagonalization of the linear response matrix. We also show that they can also be more efficient than real-time TDDFT simulations. We compare the pros and cons of these methods in terms of their accuracy and computation and storage cost.

1 Introduction

Time-Dependent Density Functional Theory (TDDFT) [1] has emerged as an important tool for reliable excited-state calculations for a broad spectrum of applications from molecular to materials systems. The most common formulation of TDDFT in quantum chemistry is in the frequency domain via linear response theory or the Casida formulation[2, 3]. This approach is also known as linear-response (LR) TDDFT and is widely used to calculate absorption spectra. Computing the absorption spectrum with this approach involves solving a non-Hermitian eigenvalue problem. Formally, the numerical cost to diagonalize the full LR-TDDFT matrix equations scales as $\mathcal{O}(N^6)$ due to the tetradic nature of the excitation matrix. [4] As a result, for large systems, this approach becomes prohibitively expensive if a large number of excitations ($\sim 10^3 - 10^4$) are needed. We refer the reader to more comprehensive reviews of LR-TDDFT [2, 3, 5, 6, 7].

The absorption spectrum is an intrinsic property of a system. In most common cases, one is primarily interested in the lowest excited-states. However, in some cases, like large molecular complexes and high density of states (DOS) materials, excitations over a wider energy range may be required, resulting in a very demanding calculation[8, 9]. Therefore, it is appealing to look for approximate, yet accurate, computational approaches for large systems.

In this paper, we discuss computationally efficient ways to estimate the absorption spectrum of a finite system. Specifically, we are interested in methods that do not explicitly compute the eigenvalues and eigenvectors of the full LR-TDDFT matrix. Within the Tamm–Dancoff approximation (TDA) [10], the matrix to be diagonalized becomes Hermitian. In this case, the absorption spectrum can be approximated by the Kernel Polynomial method (KPM), originally proposed to estimate the density of states of a symmetric matrix [11]. This approach has been discussed in Ref. [12]. Here, we will show that KPM can be easily extended to the full LR-TDDFT equations.

A "two-sided" Lanczos procedure is proposed in Ref. [13] to approximate the absorption spectrum in the full LR-TDDFT framework. We will show that it is possible to use a more standard Lanczos algorithm with a properly chosen inner product to obtain an accurate approximation of the absorption spectrum.

The paper is organized as follows. For completeness, we first derive the expression to be computed in the absorption spectrum estimation using matrix notation. Although this result is well known e.g. Ref. [13], our

derivation highlights the relationship between several quantities associated with the LR-TDDFT eigenvalue problem. We then present the KPM and the Lanczos algorithms for estimating the absorption spectrum in section 3. We also discuss how to use the KPM and the Lanczos algorithm to compute the density of states (DOS) of the LR-TDDFT eigenvalue problem. Computational results that demonstrate the effectiveness of the Lanczos algorithm are presented in section 6 where we also compare the accuracy and cost of KPM and the Lanczos approach.

2 Linear Response and Absorption Spectrum

In the linear response regime of the TDDFT, the optical absorption spectrum of a finite system can be obtained from the trace of the 3×3 dynamic polarizability tensor $\alpha_{\mu,\nu}$ defined as

$$\alpha_{\mu,\nu}(\omega) = \langle \mu | -\frac{1}{\pi}\mathrm{Im}\chi(\omega)|\nu\rangle = -\frac{1}{\pi}\mathrm{Im}\langle \mu|\chi(\omega)|\nu\rangle, \tag{1}$$

where μ and ν are one of the coordinate variables x, y and z, and $\chi(\omega)$ characterizes the linearized charge density response $\Delta\rho$ to an external frequency dependent potential perturbation $v_{\mathrm{ext}}(r;\omega)$ of the ground state Kohn–Sham Hamiltonian in the frequency domain, i.e.

$$\Delta\rho(r) = \int \chi(r,r';\omega)v_{\mathrm{ext}}(r';\omega)dr'.$$

The symbol $\mathrm{Im}\chi$ denotes the imaginary part of χ, which is also called the spectral function of χ. The reason that Im appears in (1) will be explained in the Appendix.

It is well known [5] that χ can be expressed as

$$\chi(r,r';\omega) = \int \varepsilon^{-1}(r,r''';\omega)\chi_0(r''',r';\omega)dr''', \tag{2}$$

where ε is called the dielectric operator defined in Eq. (3), and ε^{-1} is the inverse of the dielectric operator in the operator sense.

$$\varepsilon(r,r''';\omega) = \delta(r,r''') - \int \chi_0(r,r'';\omega)f_{Hxc}(r'',r''')dr''. \tag{3}$$

In TDDFT, the Hartree and exchange–correlation kernel f_{Hxc} is frequency-independent, and is defined as

$$f_{Hxc}(r,r') = \frac{1}{|r-r'|} + \frac{\delta v_{xc}[\rho](r)}{\delta\rho(r')},$$

with v_{xc} being the exchange–correlation potential. The retarded irreducible independent-particle polarization function χ_0 is defined by

$$\chi_0(r,r';\omega) = \sum_j \sum_a \frac{\phi_j(r)\phi_a^*(r)\phi_j^*(r')\phi_a(r')}{\omega - \Delta\varepsilon_{a,j} + \mathrm{i}\eta} - \frac{\phi_j^*(r)\phi_a(r)\phi_j(r')\phi_a^*(r')}{\omega + \Delta\varepsilon_{a,j} + \mathrm{i}\eta}, \tag{4}$$

where (ε_j,ϕ_j) and (ε_a,ϕ_a) are eigenpairs of the ground state self-consistent Kohn–Sham Hamiltonian. Here j and a are indices for the occupied and virtual Kohn–Sham eigenfunctions (orbitals), $\Delta\varepsilon_{a,j} \equiv \varepsilon_a - \varepsilon_j$, and $\Delta\varepsilon_{a,j} > 0$ by definition. $\eta > 0$ is an infinitesimally small constant to keep Eq. (4) well defined for all ω.

Since we consider the excitation properties of finite systems, without loss of generality we assume that all Kohn–Sham eigenfunctions ϕ_j and ϕ_a are real. To simplify the notation, we will use $\Phi(r)$ to denote a matrix that contains all products of occupied and virtual wavefunction pairs, i.e., $\Phi(r) = \left[\phi_j(r)\phi_a^*(r),\dots\right]$ and D_0 to denote a diagonal matrix that contains $\Delta\varepsilon_{a,j}$ on its diagonal. Using this notation, we can rewrite χ_0 as

$$\chi_0(r,r';\omega) = [\Phi(r),\Phi(r)]\begin{bmatrix}(\omega+\mathrm{i}\eta)I - D_0 & 0\\ 0 & -(\omega+\mathrm{i}\eta)I - D_0\end{bmatrix}^{-1}\begin{bmatrix}\Phi(r')\\ \Phi(r')\end{bmatrix}$$

To simplify the notation further, we define

$$C = \begin{bmatrix} I & 0 \\ 0 & -I \end{bmatrix}, \quad D = \begin{bmatrix} D_0 & 0 \\ 0 & D_0 \end{bmatrix}, \quad \text{and} \quad \hat{\Phi}(r) = [\Phi(r), \Phi(r)],$$

which allows us to rewrite χ_0 succinctly as

$$\chi_0(r, r'; \omega) = \hat{\Phi}(r) \left[(\omega + \mathrm{i}\eta)C - D\right]^{-1} \hat{\Phi}(r'). \tag{5}$$

We denote by $\Phi, \hat{\Phi}$ the finite dimensional matrix obtained by discretizing $\Phi, \hat{\Phi}(r)$ on real space grids, respectively. Similarly, the discretized $\hat{\Phi}(r')$ can be viewed as the matrix transpose of $\hat{\Phi}$. Replacing all integrals with matrix–matrix multiplication notations and making use of the Sherman–Morrison–Woodbury formula for manipulating a matrix inverse, we can show that (see Appendix for a detailed derivation)

$$\chi(\omega) = \hat{\Phi} \left[(\omega + \mathrm{i}\eta)C - \Omega\right]^{-1} \hat{\Phi}^T \tag{6}$$

where

$$\Omega \equiv \begin{bmatrix} A & B \\ B & A \end{bmatrix} = \begin{bmatrix} D_0 + \Phi^T f_{Hxc} \Phi & \Phi^T f_{Hxc} \Phi \\ \Phi^T f_{Hxc} \Phi & D_0 + \Phi^T f_{Hxc} \Phi \end{bmatrix}. \tag{7}$$

Here $\Phi^T f_{Hxc} \Phi$ is an $n_o n_v \times n_o n_v$ matrix commonly known as the *coupling matrix*, where n_o and n_v are the number of occupied and virtual states respectively. The $(j, a; j', a')$th element of the matrix is evaluated as

$$\int dr dr' \phi_j(r) \phi_a(r) f_{Hxc}(r, r') \phi'_j(r') \phi'_a(r').$$

It follows from (6) that

$$\langle x|\chi(\omega)|x\rangle = \hat{x}^T \left[(\omega + \mathrm{i}\eta)C - \Omega\right]^{-1} \hat{x} \tag{8}$$

where $\hat{x} = \left[\hat{x}_1^T, \hat{x}_1^T\right]^T$, and $\hat{x}_1 = \Phi^T x$ is a column vector of size $n_o n_v$. The (j, a)th element of $\hat{x}_1$ can be evaluated as

$$\int x \phi_j(r) \phi_a(r) dr. \tag{9}$$

It can be easily verified that

$$\langle x|\chi(\omega)|x\rangle = \hat{x}^T \left[(\omega + \mathrm{i}\eta)I - H\right]^{-1} C\hat{x}, \tag{10}$$

where

$$H = \begin{bmatrix} A & B \\ -B & -A \end{bmatrix} \tag{11}$$

and is also referred to as the Casida or LR-TDDFT matrix equations. Although H is non-Hermitian, it has a special structure that has been examined in detail in previous work [6, 14, 15]. In particular, when both $K \equiv A - B$ and $M \equiv A + B$ are positive definite, it can be shown that eigenvalues of H come in positive and negative pairs $(-\lambda_i, \lambda_i)$, $\lambda_i > 0$, $i = 1, 2, \ldots, n_o n_v$. If $\left[u_i^T, v_i^T\right]^T$ is the right eigenvector associated with λ_i, the left eigenvector associated with the same eigenvalue is $\left[u_i^T, -v_i^T\right]^T$.

It follows from the eigen-decomposition of H (See the Appendix) and (10) that

$$\langle x|\chi(\omega)|x\rangle = \sum_{i=1}^{n_o n_v} \left(\frac{\left[\hat{x}_1^T (u_i + v_i)\right]^2}{\omega - \lambda_i + \mathrm{i}\eta} - \frac{\left[\hat{x}_1^T (u_i + v_i)\right]^2}{\omega + \lambda_i + \mathrm{i}\eta} \right), \tag{12}$$

where the eigenvectors satisfy the normalization condition $u_i^T u_i - v_i^T v_i = 1$. In the limit of $\eta \to 0^+$, $\alpha_{x,x} = -\frac{1}{\pi} \mathrm{Im} \langle x|\chi(\omega)|x\rangle$ becomes

$$\sum_{i=1}^{n_o n_v} \left[\hat{x}_1^T (u_i + v_i)\right]^2 \left[\delta(\omega - \lambda_i) - \delta(\omega + \lambda_i)\right] = \hat{x}^T \delta(\omega I - H) C \hat{x}.$$

Hence, the absorption spectrum has the form

$$\sigma(\omega) = -\frac{1}{3\pi}\lim_{\eta\to 0^+}\mathrm{Im}\Big(\langle x|\chi(\omega)|x\rangle + \langle y|\chi(\omega)|y\rangle + \langle z|\chi(\omega)|z\rangle\Big) = \frac{1}{3}\sum_i f_i^2\left[\delta(\omega-\lambda_i)-\delta(\omega+\lambda_i)\right], \quad (13)$$

where $f_i^2 = \left[\hat{x}_1^T(u_i+v_i)\right]^2 + \left[\hat{y}_1^T(u_i+v_i)\right]^2 + \left[\hat{z}_1^T(u_i+v_i)\right]^2$ is known as the *oscillator strength*.

It is not difficult to show that $w_i \equiv u_i + v_i$ is the ith left eigenvector of the $n_o n_v \times n_o n_v$ matrix MK, associated with the eigenvalue λ_i^2. The corresponding right eigenvector is in the direction of $u_i - v_i$. (See Appendix.) Therefore, the absorption spectrum can be obtained by computing eigenvalues of KM, which is of half of the size of H. It can be shown that $\alpha_{x,x}$ can be expressed as

$$\alpha_{x,x}(\omega) = 2\,\mathrm{sign}(\omega)\hat{x}_1^T K\delta(\omega^2 I - MK)\hat{x}_1, \quad (14)$$

where the matrix function $\delta(\omega^2 - MK)$ should be understood in terms of the eigendecomposition of MK, i.e., $\delta(\omega^2 - MK) = \sum_{i=1}^{n_o n_v}\delta(\omega^2-\lambda_i^2)(u_i-v_i)(u_i+v_i)^T$. The derivation can be found in the Appendix. In (13), the value of $\omega = \lambda_j > 0$ is often referred to as the jth excitation energy, whereas $\omega = -\lambda_j < 0$ is often referred to the jth deexcitation energy. Because the oscillator strength factors associated with these energy levels are identical in magnitude and opposite in signs, it is sufficient to focus on just one of them. In the following, we will only be concerned with excitation energies, i.e., we assume $\omega > 0$.

3 Algorithms for Approximating the Absorption Spectrum

If the eigenvalues and eigenvectors of H or MK are known, the absorption spectrum defined in (13) is easy to construct. However, because the dimensions of H and MK are $\mathcal{O}(n_o n_v)$, computing all eigenvalues and eigenvectors of these matrices is prohibitively expensive for large systems due to the $\mathcal{O}(n_o^3 n_v^3)$ complexity.

If we only need the excitation energy and oscillator strength of the first few low excited states, we may use an iterative eigensolver such as Davidson's method or a variant of the locally optimal block preconditioned gradient method (LOBPCG) to compute the lowest few eigenpairs of MK [6, 16, 17]. These methods only require a user to provide a procedure for multiplying A and B with a vector. The matrix A or B does not need to be explicitly constructed. However, these methods tend to become prohibitively expensive when the number of eigenvalues within the desired energy range becomes large.

In this section, we introduce two methods that do not require computing eigenvalues and eigenvectors of the Casida Hamiltonian explicitly. These methods also only require a procedure for multiplying A and B with a vector. They can provide satisfactory approximation to the overall shape of the absorption spectrum from which the position and the height of each major peak with a desire energy range can be easily identified.

3.1 The Kernel Polynomial Method

One approach follows from the Kernel Polynomial Method (KPM) reviewed in [11]. To use this method, it is more convenient to consider the case of $\omega > 0$ and view $\alpha_{x,x}(\omega)$ as a function of $\varpi := \omega^2$. The general idea is to express $\alpha_{x,x}$ (and similarly $\alpha_{y,y}$ as well as $\alpha_{z,z}$) formally by a polynomial expansion of the form

$$\hat{\alpha}_{x,x}(\varpi) = \sum_{k=0}^{\infty}\gamma_k\mathcal{T}_k(\varpi), \quad (15)$$

where

$$\hat{\alpha}_{x,x}(\varpi) = \frac{\sqrt{1-\varpi^2}}{2}\alpha_{x,x}(\omega) = \sqrt{1-\varpi^2}\,\hat{x}_1^T K\delta(\varpi I - MK)\hat{x}_1 = \sqrt{1-\varpi^2}\sum_i \lambda_i(\hat{x}_1^T w_i)^2\delta(\varpi-\lambda_i^2),$$

w_i is the ith left eigenvector of MK and $\{\mathcal{T}_k(\varpi)\}$ is a set of orthogonal polynomials of degree k.

For instance, we choose $\{\mathcal{T}_k(\varpi)\}$ to be Chebyshev polynomials. Using the identity $\int \mathcal{T}_k(\varpi)\delta(\varpi - \lambda_i^2) = \mathcal{T}_k(\lambda_i^2)$, we can compute the expansion coefficients γ_k's by

$$\begin{aligned}
\gamma_k &= \frac{2-\delta_{k0}}{\pi}\int_{-1}^{1}\frac{1}{\sqrt{1-\varpi^2}}\mathcal{T}_k(\varpi)\hat{\alpha}_{x,x}(\varpi)d\varpi \\
&= \frac{2-\delta_{k0}}{\pi}\hat{x}_1^T\left[\sum_{i=1}^{n_o n_v}\lambda_i w_i \mathcal{T}_k(\lambda_i^2) w_i^T\right]\hat{x}_1 \\
&= \frac{2-\delta_{k0}}{\pi}\hat{x}_1^T\left[\sum_{i=1}^{n_o n_v} K(u_i - v_i)\mathcal{T}_k(\lambda_i^2)(u_i+v_i)^T\right]\hat{x}_1 \\
&= \frac{2-\delta_{k0}}{\pi}\hat{x}_1^T K\mathcal{T}_k(MK)\hat{x}_1. \qquad (16)
\end{aligned}$$

Here δ_{ij} is the Kronecker δ symbol so that $2-\delta_{k0}$ is equal to 1 when $k=0$ and to 2 otherwise. In deriving (16) from the line before, we use the property that $w_i = u_i + v_i$ is an left eigenvector of MK and the corresponding right eigenvector is in the direction of $u_i - v_i = \lambda_i K^{-1}(u_i + v_i)$.

Because Chebyshev polynomials $\mathcal{T}_k(\varpi)$ can be generated recursively using the three-term recurrence

$$\mathcal{T}_0(\varpi) = 1, \quad \mathcal{T}_1(\varpi) = \varpi, \quad \mathcal{T}_{k+1}(\varpi) = 2\mathcal{T}_k(\varpi) - \mathcal{T}_{k-1}(\varpi),$$

for $\varpi \in [-1,1]$, we do not need to construct $\mathcal{T}_k(MK)$ explicitly. We can apply $\mathcal{T}_k(MK)\hat{x}_1$ recursively using a 3-term recurrence also. This three-term recurrence allows us to implement the KPM by storing 3 or 4 vectors depending on whether an intermediate vector is used to hold intermediate matrix vector products. A pseudocode for the KPM is given in Algorithm 1. The cost of the KPM is dominated by the multiplication of the matrix A and B with vectors, and proportional to the degree of the expansion polynomial or the number of expansion terms in (15).

The derivation above requires $\varpi \in [-1,1]$. To generate Chebyshev polynomials on an interval $[a,b]$ where a and b are the estimated lower and upper bounds of the eigenvalues of MK, a proper linear transformation should be used to map ϖ to $[-1,1]$ first.

In (15), we assume that the $\eta \to 0^+$ limit has been taken. Because the left hand side a sum of Dirac-δ distributions, which is discontinuous, a finite least squares polynomial expansion will produce the well known Gibbs oscillations as shown in [18], and lead to larger errors near the point of discontinuity. This effect is more pronounced for molecules of which excitation energies are well isolated, and is less severe for solids whose excitation energies are more closely spaced. To reduce the effects of the Gibbs oscillation, we can set η to a small positive constant (instead of taking the $\eta \to 0^+$ limit.) This is equivalent to replacing each Dirac-δ distribution with a Lorentzian of the form

$$g_\eta(\omega - \lambda_i) = \frac{1}{\pi}\frac{\eta}{(\omega-\lambda_i)^2 + \eta^2}.$$

In this case, the expansion coefficients γ_k's need to be computed in a different way. We may also replace the Dirac-δ distributions on the left hand side of (15) with Gaussians of the form

$$g_\sigma = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-(\omega-\lambda_i)^2/(2\sigma^2)}, \qquad (17)$$

where σ is a smoothing parameter that should be chosen according to the desired resolution of $\alpha_{x,x}$. We refer to the technique of replacing Dirac-δ distributions with Lorentzians or Gaussians as *regularization*. In [18], we show how the expansion coefficients γ_k's can be computed recursively when Dirac-δ distributions are replaced with Gaussians, and $\mathcal{T}_k(\varpi)$ is taken to be the kth degree Legendre polynomial. We referred to this particular expansion as the Delta–Gaussian–Legendre (DGL) expansion [18]. Another technique for reducing the Gibbs phenomenon is the use of *Jackson damping* [19]. However, it has been shown in [18], that Jackson damping can lead to over-regularized spectrum.

Algorithm 1: The kernel polynomial method for approximating the absorption spectrum of a finite system.

Input: Real symmetric matrices A and B defined in (7); A set of points $\{\omega_i\}$, $i = 1, 2, \ldots, m$ at which the absorption spectrum is to be evaluated, the dipole vector $\hat{x}_1$ defined in (9), the degree k of the expansion polynomial used in KPM.

Output: Approximate absorption spectrum $\sigma(\omega_i)$ for $i = 1, 2, \ldots, m$.

1: Estimate the upper and lower bounds (λ_{ub} and λ_{lb}) of the eigenvalues of MK by a few step of the Lanczos iteration.
2: Let $c = (\lambda_{lb} + \lambda_{ub})/2$; $d = (\lambda_{ub} - \lambda_{lb})/2$;
3: Set $\zeta_j = 0$ for $j = 0, \cdots, k$;
4: Let $q_0 \leftarrow \hat{x}_1$;
5: **for** $j = 0, 1, \ldots, k$ **do**
6: $\quad w_A \leftarrow Aq_j$;
7: $\quad w_B \leftarrow Bq_j$;
8: $\quad$ Compute $\zeta_j \leftarrow \zeta_j + q_0^T(w_A + w_B)$;
9: $\quad$ Compute q_{j+1} via the three-term recurrence

$$q_{j+1} = 2\left[(A - B)(w_A + w_B) - c(w_A + w_B)\right]/d - q_{j-1}$$

$\quad$ (for $j = 0$, $q_1 = \left[(A - B)(w_A + w_B) - c(w_A + w_B)\right]/d$);
10: **end for**
11: Set $\gamma_j \leftarrow \dfrac{2 - \delta_{j0}}{n_o n_v \pi}\zeta_j$ for $j = 0, 1, \ldots, k$;
12: Evaluate

$$\alpha_{x,x}(\omega_i) = 2\sum_{j=0}^{k} \gamma_j \mathcal{T}_j\Big(\frac{\omega_i^2 - c}{d}\Big)\Big/\sqrt{1 - \Big(\frac{\omega_i^2 - c}{d}\Big)^2}$$

$\quad$ for $i = 1, 2, \ldots, m$;
13: Set $q_0 = \hat{y}_1$ and $q_0 = \hat{z}_1$ and repeat Steps 5–10 to obtain $\alpha_{y,y}$ and $\alpha_{z,z}$;
14: $\sigma(\omega_i) = \big[\alpha_{x,x}(\omega_i) + \alpha_{y,y}(\omega_i) + \alpha_{z,z}(\omega_i)\big]/3$ for $i = 1, 2, \ldots, m$.

Algorithm 2: A k-step Lanczos method in the $(A-B)$-inner product.

Input: Real symmetric matrices A and B defined in (7); A starting vector v_0; The number of steps k.
Output: Matrices Q_k and T_k that satisfy (19) and (20)

1: $Q_k(:,1) = v_0/\sqrt{v_0^T(A-B)v_0}$;
2: **for** $j = 1, \ldots, k$ **do**
3: $w \leftarrow (A-B)Q_k(:,j)$;
4: $w \leftarrow (A+B)w$;
5: $h \leftarrow Q_k(:,1:j)^T(A-B)w$;
6: $H(1:j,j) = h$;
7: $w \leftarrow w - Q_k(:,1:j)h$;
8: **if** $j < k$ **then**
9: $Q_k(:,j+1) = w/\sqrt{w^T(A-B)w}$;
10: **end if**
11: **end for**

3.2 The Lanczos Method

Another way to estimate $\alpha_{x,x}$ is to use the Lanczos algorithm. Because MK is self-adjoint with respect to the inner produced induced by K, i.e.,

$$\langle v, MKv\rangle_K = v^T KMKv = \langle MKv, v\rangle_K, \tag{18}$$

we can use the K-inner product to generate a k-step Lanczos factorization of the form

$$MKQ_k = Q_k T_k + f_k e_k^T, \tag{19}$$

where

$$Q_k^T K Q_k = I \quad \text{and} \quad Q_k^T K f_k = 0, \tag{20}$$

and T_k is a tridiagonal matrix of size $k \times k$. The major steps of the K-inner product Lanczos procedure is shown in Algorithm 2. We use the MATLAB notation $Q(:,j)$ to denote the jth column of the matrix Q. In principle, because T_k is tridiagonal in exact arithmetic, columns of Q_k can be generated via a three-term recurrence. However, it is well known that as some of the eigenvalues of T_k converge to those of MK, and loss of orthogonality among columns of Q_k can occur due to potential instability in the numerical procedure. As a result, T_k may become singular with several spurious eigenvalues near 0 [20]. To avoid loss of orthogonality, we perform full reorthogonalization as shown in steps 5–7 of Algorithm 2. Unless k is extremely large, the cost of full reorthogonalization is relatively small compared to the cost of multiplying A and B with vectors. However, full reorthogonalization does require keeping all columns of Q_k in memory.

If we choose the starting vector for the Lanczos iteration to be $\hat{x}_1$, i.e., $Q_k e_1 = \hat{x}_1/\sqrt{\hat{x}_1^T K \hat{x}_1}$, then $e_1^T \delta(\omega^2 I - T_k) e_1^T$ serves as a good approximation of $\alpha_{x,x}(\omega)$.

To see why this is the case, let us first assume that $\delta(\omega^2 I - MK)$ can be formally approximated by a k-th degree polynomial of the form $p_k(MK;\omega^2)$, i.e.

$$\delta(\omega^2 I - MK) \approx p_k(MK;\omega^2) = \sum_{i=0}^{k} \gamma_i(\omega^2)(MK)^i.$$

It follows from (19) that

$$p_k(MK;\omega^2)Q_k = Q_k p_k(T_k;\omega^2) + R_k,$$

where R_k is a residual matrix that vanishes when $k = n_v n_o$.

Consequently, we can show that

$$\begin{aligned}
\lim_{\eta\to 0^+} -\frac{1}{\pi}\mathrm{Im}\langle x|\chi(\omega)|x\rangle &= \hat{x}_1^T K\delta(\omega^2 I - MK)\hat{x}_1 \\
&\approx e_1^T Q_k^T K p_k(MK;\omega^2) Q_k e_1 \\
&= e_1^T Q_k^T K\left[Q_k p_k(T_k;\omega^2) + R_k\right] e_1 \\
&\approx e_1^T p_k(T_k;\omega^2) e_1 \qquad (21)\\
&\approx e_1^T \delta(\omega^2 I - T_k) e_1 \\
&= \sum_{j=1}^{k} \tau_j^2 \delta(\omega^2 - \theta_j) \qquad (22)
\end{aligned}$$

where $\theta_j \geq 0$ is the jth eigenvalue of the $k \times k$ tridiagonal matrix T_k sorted in increasing order and τ_j is the first component of the jth eigenvector of T_k.

As shown in the Appendix, $\delta(\omega^2 - \theta_j)$ can be rewritten as

$$\frac{1}{2\sqrt{\theta_j}}\left[\delta(\omega + \sqrt{\theta_j}) + \delta(\omega - \sqrt{\theta_j})\right]. \qquad (23)$$

Therefore, for $\omega > 0$,

$$\alpha_{x,x}(\omega) \approx \sum_{j=1}^{k} \tau_j^2 \delta(\omega^2 - \theta_j) = \sum_{j=1}^{k} \frac{\tau_j^2}{2\sqrt{\theta_j}} \delta(\omega - \sqrt{\theta_j}) \qquad (24)$$

holds.

The cost of the Lanczos algorithm is proportional to the number of steps (k) to be taken in the Lanczos iteration. We would like to keep k as small as possible without losing desired information in Eq. (22). However, when k is small, (24) gives a few spikes at the square root of the eigenvalues of T_k, which are called *Ritz values*. No absorption intensity is given at other frequencies. To estimate the intensity of the absorption at these frequencies, we replace $\delta(\omega - \sqrt{\theta_j})$ in (22) by either a Gaussian or a Lorentzian. The use of these regularization functions allows us to interpolate the absorption intensity from the square root of Ritz values to any frequency. Replacing $\delta(\omega - \theta_j)$ with a Lorentzian is equivalent to not taking the $\eta \to 0^+$ limit in (22). It is also equivalent to computing the $(1,1)$ entry of the matrix inverse $[(\omega + i\eta)I - T_k]^{-1}$. This entry can be computed recursively by using a recursive expression that is related to continued fractions [21]. This is the method of Haydock [22]. However, since the cost of computing the inverse and the eigenvalue decomposition of a small tridiagonal matrix is negligibly small, we do not gain much by using the Haydock's recursion.

4 Estimating the Density of States of the LR-TDDFT Matrix Equations

In some cases, the oscillator strength associated with some of the excitation energies are negligibly small. These are sometimes referred to as *dark states* [5]. To reveal these states, it is sometimes useful to examine the density of states (DOS) associated with the LR-TDDFT matrix equations(11).

In the review paper [18], several numerical algorithms for estimating the DOS of symmetric matrices are presented and compared. To apply these techniques, we make use of the fact that eigenvalues of H can be obtained from that of MK or $K^{1/2}MK^{1/2}$ which is Hermitian. As a result, the DOS can be written as

$$\phi(\omega) = \mathrm{trace}(\omega^2 I - K^{1/2}MK^{1/2}). \qquad (25)$$

If we use the KPM to approximate (25), i.e.,

$$\phi(\omega) \approx \sum_{k=0}^{\infty} \gamma_k \mathcal{T}_k(\omega^2), \qquad (26)$$

the expansion coefficients γ_k can be computed as

$$\gamma_k = \frac{2-\delta_{k0}}{n\pi} \operatorname{trace}\left[\mathcal{T}_k(K^{1/2}MK^{1/2})\right] = \frac{2-\delta_{k0}}{n\pi} \operatorname{trace}\left[\mathcal{T}_k(MK)\right] \tag{27}$$

Thus, apart from the scaling factor $(2-\delta_{k0})/(n\pi)$, γ_k is the trace of $\mathcal{T}_k(MK)$, and such a trace can be estimated through a statistical sampling technique that involves computing

$$\zeta_k = \frac{1}{n_{\text{vec}}} \sum_{l=1}^{n_{\text{vec}}} \left(q_0^{(l)}\right)^T K\mathcal{T}_k(MK) q_0^{(l)} \tag{28}$$

for a set of randomly generated vectors $q_0^{(1)}$, $q_0^{(2)}$, ..., $q_0^{(n_{\text{vec}})}$. One possible choice of the random vectors is that each entry of $q_0^{(l)}$ follows an independent Gaussian distribution $\mathcal{N}(0,1)$ for all $l = 1, \ldots, n_{\text{vec}}$. The method only requires multiplying A and B with a number of vectors, and the multiplication of $\mathcal{T}_k(MK)$ with a vector can be implemented through a three-term recurrence.

To estimate the DOS by the Lanczos method, we can simply use the K-orthogonal Lanczos factorization shown in (19) to obtain a sequence of tridiagonal matrices $T_k^{(l)}$, $l = 1, 2, \ldots, n_{\text{vec}}$ using a set of randomly generated starting vectors $q_0^{(1)}$, $q_0^{(2)}$, ..., $q_0^{(n_{\text{vec}})}$ as mentioned above. The approximate DOS can be expressed as

$$\phi(\omega) \approx \frac{1}{n_{\text{vec}}} \sum_{l=1}^{n_{\text{vec}}} \sum_i (\tau_i^{(l)})^2 \delta(\omega^2 - \theta_i^{(l)}), \tag{29}$$

where $\theta_i^{(l)}$ is the ith eigenvalue of $T_k^{(l)}$, $\tau_i^{(l)}$ is the first component of the corresponding eigenvector. To regularize the approximate DOS, we rewrite $\delta(\omega^2 - \theta_i^{(l)})$ in the form of (23), and replace $\delta(\omega + \sqrt{\theta_i^{(l)}})$ and $\delta(\omega - \sqrt{\theta_i^{(l)}})$ with properly defined Gaussians $g_\sigma(\omega + \sqrt{\theta_i^{(l)}})$ and $g_\sigma(\omega - \sqrt{\theta_i^{(l)}})$ respectively.

5 Frozen Orbital and Tamm–Dancoff Approximation

When ω is relatively small compared with $\Delta\varepsilon_{a,j}$, the contribution of the corresponding term in (4) is relatively insignificant. Thus, it is reasonable to leave out these terms which constitute indices j's that correspond to the lowest occupied orbitals often known as the *core orbitals*. By using this *frozen core approximation* (FCA) [23], we effectively reduces the dimension of A and B matrices in (7) from $n_o n_v$ to $(n_o - n_f)n_v$, where n_f is the number of lowest occupied orbitals that are excluded (or "frozen") from (4). It is well known that the FCA preserves the low end of the absorption spectrum while removing poles of $\chi(\omega)$ at higher frequencies. Therefore, it is extremely useful to combine FCA with the Lanczos algorithm to obtain an accurate approximation to the absorption spectrum in the low excitation energy range at a reduced cost, as we will show in the next section.

Another commonly used technique to reduce the computational cost of estimating the absorption spectrum is to set the matrix B in (7) to zero. This approximation is often referred to as the Tamm–Dancoff approximation (TDA). It reduces the eigenvalue problem to a symmetric eigenvalue problem that involves the matrix A only. However, the TDA is known to work well for some problems, but fails in others [24]. Therefore, we do not consider the TDA in this paper.

6 Computational Results

We now present computational results to demonstrate the quality the approximate absorption spectrum obtained from the algorithms presented in section 3. In particular, we compare the absorption spectrum obtained from the new algorithms with that obtained from a traditional diagonalization based approach and that obtained from a real-time TDDFT (RT-TDDFT) simulation. For details of the RT-TDDFT approach, we refer the reader to Ref. [25]. We also compare the efficiency of the new algorithms with the aforementioned approaches and show that the algorithms presented in section 3 are much faster.

6.1 Test problems

The test examples used in this section are generated from the NWChem-6.5 program [26]. They include two small dyes: 2,3,5-trifluorobenzaldehyde (TFBA) and 4'-hydroxybenzylidene-2,3-dimethylimidazoline (HBDMI) as well as two relatively large molecules F2N12S and P3B2. The atomic configurations of these test problems are shown in Figure 1. All geometries were optimized in the ground state DFT calculation using the B3LYP functional. We use the 6-31G* basis set for all problems except F2N12S. The cc-pVDZ basis set is used for this molecule.

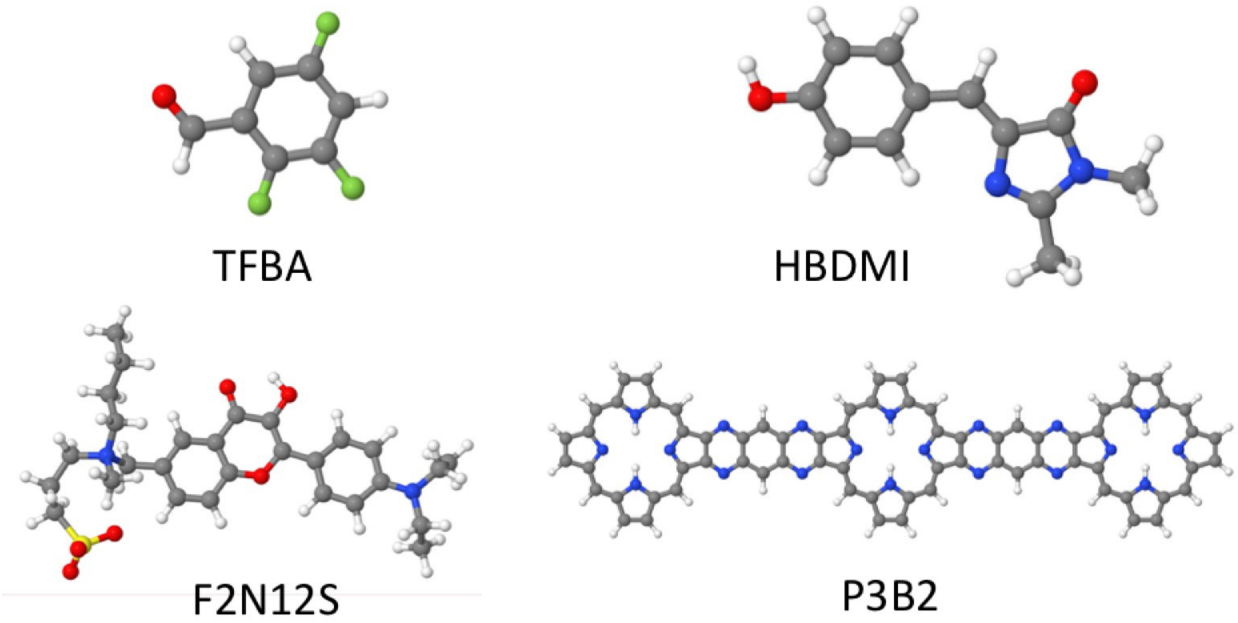


Figure 1: The molecules TFBA, HDMBI, F2N12S and P3B2.

In Table 1, we list n_o, n_v as well as the dimension of the excitation matrix ($n_o n_v$) for each problem. We also list the number of frozen orbitals (n_f) when FCA is used, and the corresponding reduced ($(n_o - n_f)n_v$.)

System	n_o	n_v	$n_o n_v$	n_f	$(n_o - n_f)n_v$
TFBA	40	120	4,800	11	3,480
HBDMI	57	191	10,887	16	7,831
F2N12S	142	456	64,752	41	46,056
P3B2	305	1,059	322,995	92	225,567

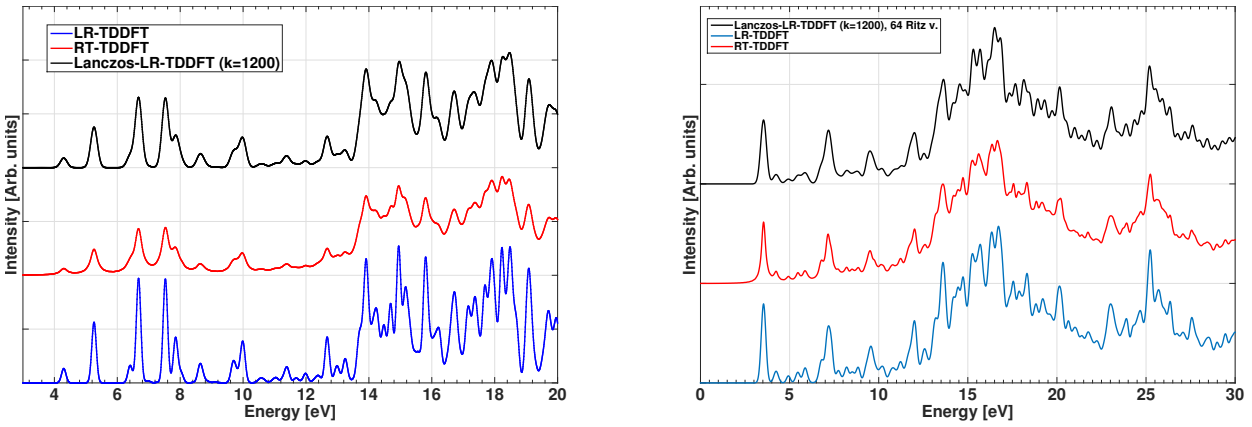
Table 1: The size of systems used in calculations, n_f denotes the number of frozen (core) orbitals. Geometries can be found in Supplementary information.

The TFBA and HBDMI problems are relatively small. Thus we can generate the matrices for these problems explicitly in NWChem and compute all eigenvalues and eigenvectors of these Hamiltonians. These eigenvalues and eigenvectors are used in the expression (13) to construct an "exact" linear response (LR) TDDFT absorption spectrum.

6.2 Approximation for small dyes

In Figure 2, we compare the approximate absorption spectrum obtained from the Lanczos algorithm with that obtained with traditional diagonalization for both the TFBA and HBDMI molecules. We only show results in the $[0, 20]$ eV energy range for TFBA and in the $[0, 30]$ eV energy range for HBDMI, since these are often the ranges of interest in practice.

We observe that the absorption spectrum obtained from $k = 1200$ steps of Lanczos iterations is nearly indistinguishable from the exact solution for TFBA. All major peaks are correctly captured. Similarly, for HBDMI, 1200 Lanczos iterations are required to achieve the same level of accuracy.



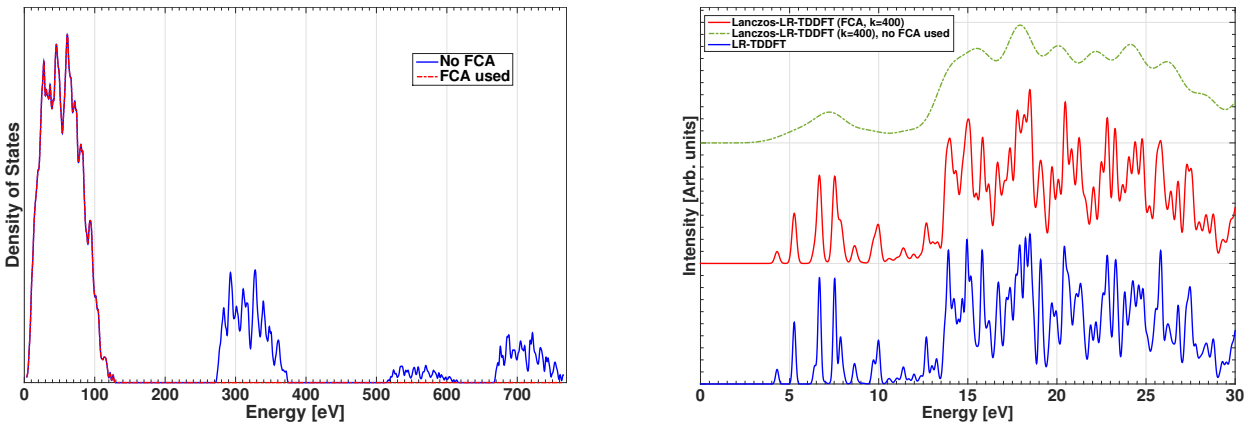
(a) (b)

Figure 2: Comparison of the approximate absorption spectrum obtained from the Lanczos algorithm with that obtained from an exact LR-TDDFT calculation and RT-TDDFT calculation for (a) TFBA (b) HBDMI.

We also plot, in Figure 2, the absorption spectrum obtained from a real-time (RT) TDDFT calculation reported in [9]. In this calculation, the time-dependent dynamic polarizability was calculated by propagating the solution to the time-dependent Kohn–Sham equation after an electric field pulse in the form of a small δ-kick and field strength 2×10^{-5} a.u. was applied at $t = 0$. The Fourier transform of the time-dependent polarizability yields the approximate absorption spectrum. Numerically, the RT-TDDFT calculations were carried out using a time step $\Delta t = 0.2$ a.u. (4.8 attoseconds), the size of the simulated trajectory is 1000 a.u. (24.2 fs, 5000 steps). We can see from this figure that the LR-TDDFT absorption spectrum matches well with that obtained from RT-TDDFT for both TFBA and HBDMI. The nearly perfect match indicates the validity of linear response approximation. Therefore, it seems reasonable to compare the absorption spectrum obtained from the Lanczos algorithm with that obtained from RT-TDDFT directly for larger problems where the traditional diagonalization approach is prohibitively expensive.

6.3 The effect of FCA

In Figure 3, we illustrate the effect of using the frozen core approximation (FCA) for TFBA by setting n_f to 11. That effectively reduces the dimension of the A and B matrices in (11) from $n_o n_v = 4800$ to $(n_o - n_f)n_v = 3480$. We can clearly see from the DOS shown in Figure 3(a) that FCA preserves the small eigenvalues of the original matrix, but the largest eigenvalues are absent. The absence of these eigenvalues allows the Lanczos algorithm to obtain more Ritz values in the low energy range (e.g. [0,20] eV) in fewer iterations, thereby producing an accurate absorption spectrum with lower cost. The cost reduction in FCA results from both the reduction in the dimension of the matrix and the reduction in the steps of Lanczos iterations required to achieve a desired resolution in the absorption spectrum.



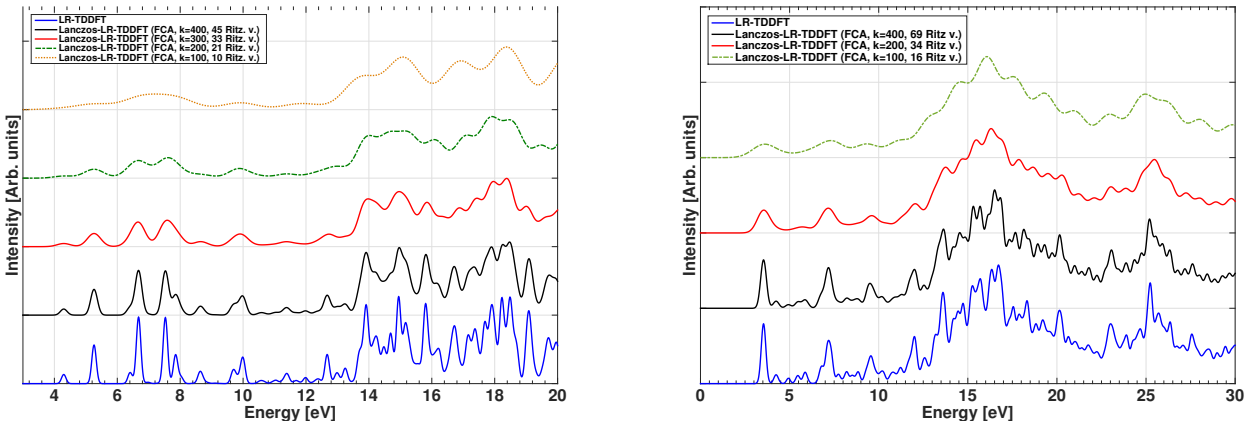
(a) (b)

Figure 3: (a) Density of states of H with and without the FCA (b) The absorption spectrum obtained from 400 Lanczos iterations with and without the FCA.

6.4 Resolution and the number of Lanczos steps

In Figure 4, we show that the resolution of the absorption spectrum obtained from the Lanczos algorithm clearly improves as we take more Lanczos iterations for TFBA and HBDMI. In this set of experiments, we use FCA for all test runs.

When $k = 400$ iterations are performed, the Lanczos approximation is nearly indistinguishable from traditional diagonalization. When k is as small as 100, the general features of the absorption spectrum are clearly captured.



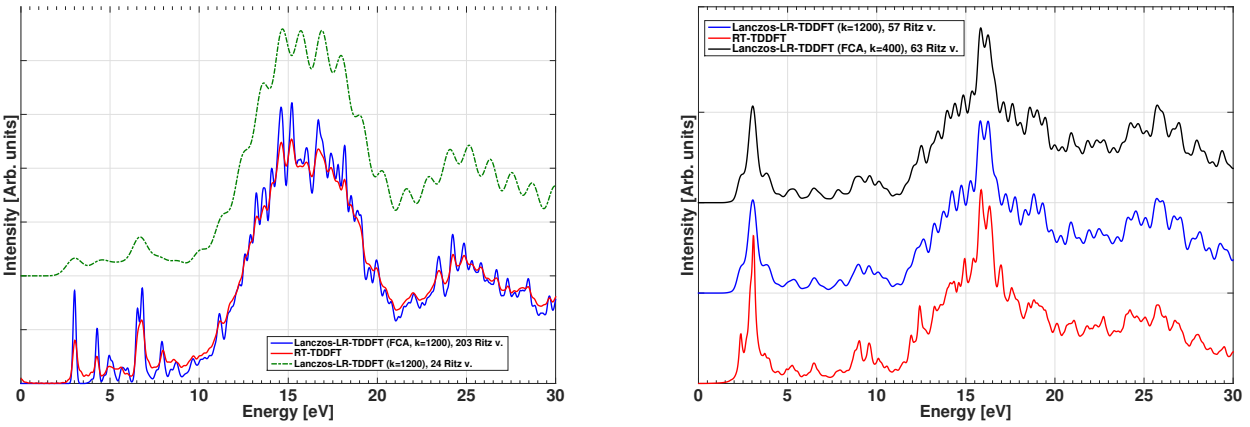
(a) (b)

Figure 4: Resolution improvement with respect to the number of Lanczos iterations k for (a) TFBA (b) HBDMI.

6.5 Approximation for larger molecules

In Figure 5, we compare the absorption spectra obtained from the Lanczos-based LR-TDDFT calculation with those obtained from RT-TDDFT simulations for both the F2N12S and P3B2 molecules. We can see that without the FCA, the Lanczos algorithm can capture the basic features of F2N12S's absorption spectrum exhibited by the RT-TDDFT simulation after $k = 1200$ iterations, but does not clearly reveal all the peaks. However, when the FCA is used, the result matches extremely well with that produced by RT-TDDFT. For the P3B2 molecule, we need to run 1200 Lanczos iterations to obtain the result that matches well with

that produced by RT-TDDFT. When the FCA is used, only 400 Lanczos iterations are needed to achieve essentially full resolution in the computed absorption spectrum.



(a) (b)

Figure 5: Comparison of simulated absorption spectra of F2N12S calculated by RT- and Lanczos-LR-TDDFT with and without FCA. Comparison of simulated absorption spectra of P3B2 calculated by RT and Lanczos-LR-TDDFT. The number of Ritz values in the displayed interval is 57 for $k = 1200$ or 63 for $k = 400$ while FCA is used.

6.6 Computational efficiency

To illustrate the computational efficiency of the Lanczos algorithm, we compare the wall clock time taken by 400 Lanczos steps with that used by the Davidson algorithm to compute the lowest 200 eigenvalues of the matrix KM, as well with that used by the RT-TDDFT to run a 24.2-femtosecond trajectory with a 4.8 attosecond time step (5000 time steps). We use the P3B2 molecule as the test example for this test. All these calculations were performed within the NWChem program on the Cascade system, which is equipped with many Xeon E5-2670 8C 2.6GHz 16-core CPUs and a Infiniband FDR network, and maintained at the Pacific Northwest National Laboratory's EMSL facility.

Note that the 200 eigenvalues obtained from the Davidson algorithm only allow us to reveal the first two major peaks of the absorption spectrum shown in Figure 5. Computing more eigenvalues of KM becomes prohibitively more expensive.

We report the wallclock time used by all three methods in Table 2. Each calculation is performed using 1536 cores. In both the Lanczos and RT-TDDFT methods, the xx, yy and zz components of the dynamic polarizability can be computed simultaneously. We use 512 cores for each component. However, when the Davidson algorithm is used to compute the lowest eigenpairs of KM, the xx, yy and zz components of the dynamic polarizability are obtained in a postprocessing step that takes a negligible amount of time. Most of the time is spent in the Davidson diagonalization process, which we ran on all 1536 cores.

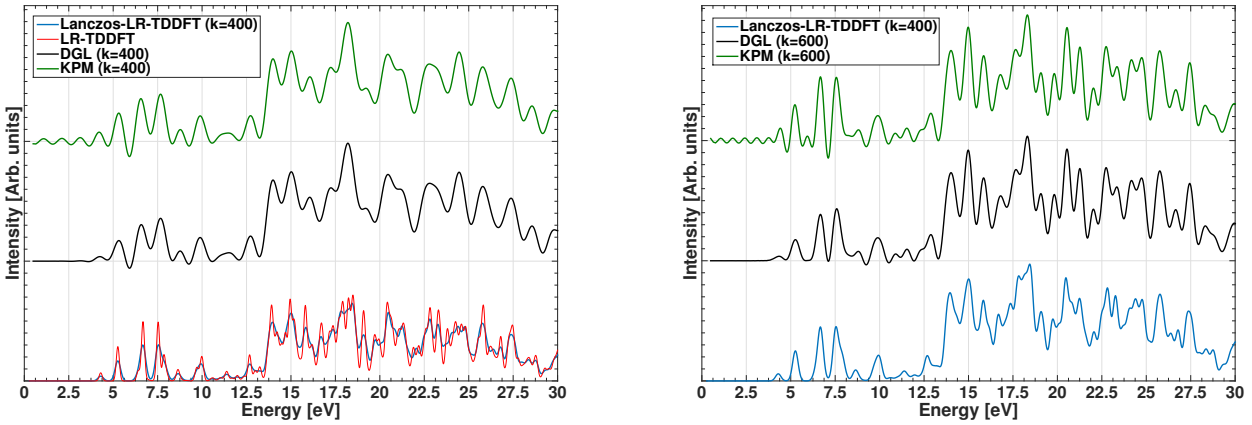
As we can see from this table that the Lanczos algorithm is much more efficient than both the Davidson diagonalization scheme and the RT-TDDFT calculation.

method	wall clock time (hours)
Lanczos	2.5
Davidson	4
RT-TDDFT	15

Table 2: A comparison of the wallclock time required to obtain the absorption spectrum of P3B2 by the Lanczos algorithm, Davidson diagonalization, and RT-TDDFT.

6.7 Lanczos vs. KPM and DGL

In Figure 6 (a), we compare the absorption spectrum obtained from the Lanczos algorithm, KPM and DGL for the TFBA dye. We ran 400 Lanczos steps, which produces a relatively high resolution approximation to the absorption spectrum produced by exact diagonalization. To make the computational cost of the KPM and DGL method comparable to that of the Lanczos algorithm, we set the degree of the polynomial approximation to 400 also in this test. We can see from Figure 6 (a) that the absorption spectra produced by the KPM and DGL algorithm match with "exact" solution reasonably well in the [0,10] eV energy range. However, they miss some of the peaks beyond 10 eV. We can see that the approximate absorption spectra produced by the KPM and the DGL algorithm are not strictly non-negative, which is an undesirable feature. Furthermore, KPM tends to produce some artificial oscillations and peaks in the approximate absorption spectrum, which may be misinterpreted as real excited states. These artificial oscillations and peaks are the result of the Gibbs oscillation that are present when a discontinuous function, such as the sum of a number of Dirac-δ distributions is approximated by a high degree polynomial. This observation is consistent with that reported in [18]. Because the absorption spectrum associated with molecules contain well isolated peaks, the KPM, which is based on polynomial approximation in a continuous measure, often exhibit Gibbs oscillation when degree of the polynomial is high. The Gibbs oscillations are clearly visible on Figure 6 between 0–5 eV and also create pseudopeak around 5.5 eV (Figure 6 (b)). This problem is reduced in the DGL algorithm.



(a) (b)

Figure 6: Comparison of simulated absorption spectra of TFBA calculated by LR-TDDFT, Lanczos-LR-TDDFT, KPM and DGL. The FCA was used. (a) $k = 400$ (b) $k = 400$ for the Lanczos-LR-TDDFT and $k = 600$ for KPM and DGL.

Figure 6 (b) shows that as the degree of the kernel polynomial increases, for example, to 600, all major peaks of the absorption spectrum in $[0, 30]$ eV are correctly resolved.

6.8 DOS by Lanczos

In Figure 7, we plot the DOS approximation for the FTBA dye obtained from the Lanczos algorithm and compare it with that obtained from a full diagonalization of the KM matrix. The FCA is used to show the DOS in the energy range $[0, 125]$ eV. We use $n_{\text{vec}} = 10$ randomly generated vectors to run the Lanczos algorithm. When 200 Lanczos steps are used in each run, the DOS approximation obtained from (29) is nearly indistinguishable from that constructed from the eigenvalues of the KM matrix. The total number of matrix vector multiplications used in this case is $10 \times 200 = 2000$. If fewer Lanczos steps are taken in each run, the resulting DOS approximation becomes less well resolved. However, the general features of the DOS can still be seen clearly when only 50 Lanczos steps are taken in each run.

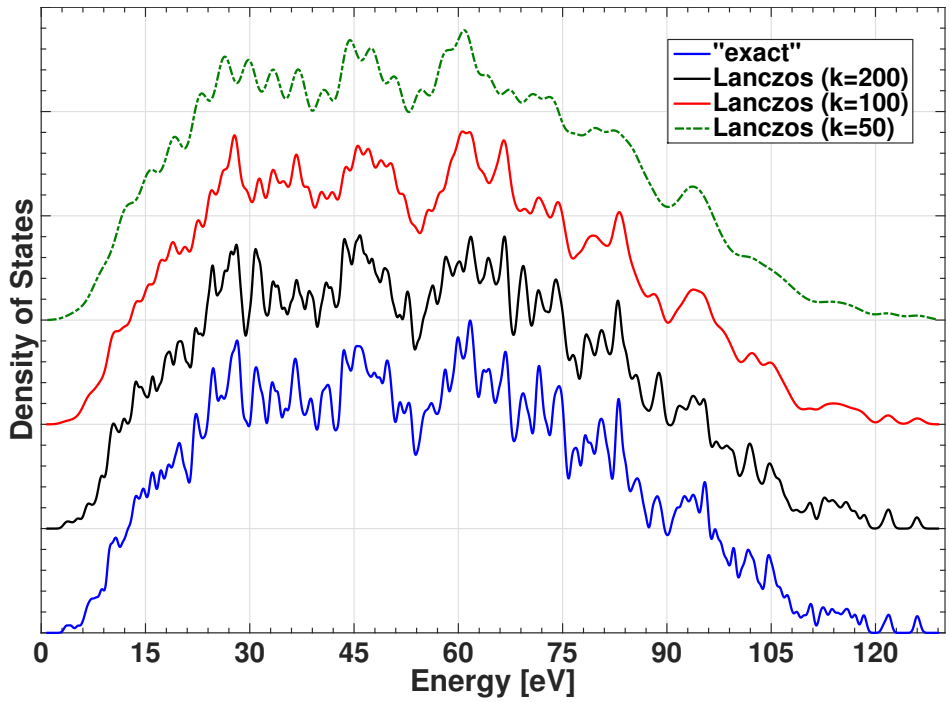


Figure 7: Resolution improvement of the density of states for FTBA, the FCA approximation is used. The number of random initial vectors for Lanczos iteration is 10.

7 Conclusion

We have described two iterative algorithms for approximating the absorption spectrum of finite systems within LR-TDDFT. These methods do not require the construction of the full matrix explicitly. Both methods exploit the special structure of the Casida matrix equations and reduce the dimension of the problem from $2n_on_v$ to n_on_v. Our computational examples show that these methods can be much more efficient than traditional methods that are based on exact diagonalization. They can also be more efficient than RT-TDDFT simulations. In addition, the Lanczos algorithm generally gives more accurate approximation to the absorption spectrum than that produced by the KPM or DGL, when the same number of matrix vector multiplications are used. In particular, the approximate absorption spectrum produced by the Lanczos algorithm is stictly non-negative. This is not necessarily the case for KPM or DGL. Furthermore, KPM tends to produce additional artificial oscillations, that may be misinterpreted as fictitious excited states. However, the Lanczos algorithm requires storing more vectors in order to overcome potential numerical instability. In contrast, both the KPM and DGL methods can be stably implemented using a three-term recurrence, leading to a minimal storage requirement.

Acknowledgment

Support for this work was provided through Scientific Discovery through Advanced Computing (SciDAC) program funded by U.S. Department of Energy, Office of Science, Advanced Scientific Computing Research and Basic Energy Sciences. A portion of the research was performed using EMSL, a DOE Office of Science User Facility sponsored by the Office of Biological and Environmental Research and located at the Pacific Northwest National Laboratory (PNNL). PNNL is operated by Battelle Memorial Institute for the United States Department of Energy under DOE contract number DE-AC05-76RL1830. The research also benefited from resources provided by the National Energy Research Scientific Computing Center (NERSC), a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231.

Appendix

In this section, we provide detailed derivation and explanation for some of the expressions presented in the previous sections.

A. Spectral function

The *spectral function* associated with a function of the form

$$f(\omega) = \sum_i \frac{a_i}{\omega - \lambda_i},$$

is defined to be

$$s(\omega) = \sum_i a_i \delta(\omega - \lambda_i).$$

It is an elegant way to describe the numerator (or weighting factor) associated with each pole of $f(\omega)$.

An alternative way to express $s(\omega)$ is through the expression

$$s(\omega) = -\frac{1}{\pi} \lim_{\eta \to 0^+} \mathrm{Im} f(\omega + i\eta) = \frac{1}{\pi} \lim_{\eta \to 0^+} \sum_i a_i \frac{\eta}{(\omega - \lambda_i)^2 + \eta^2}.$$

This is the expression we used in (1) to define the dynamic polarizability tensor α.

B. A derivation for χ

The expression of $\chi(\omega)$ given in (2) can be derived in a number of ways, (e.g. through the Liouville super-operator presented in [13].) We give a simple derivation of this expression here using basic linear algebra.

We start from a matrix representation of (2)

$$\chi(\omega) = \varepsilon^{-1}(\omega)\chi_0(\omega) = \left[I - \chi_0(\omega) f_{Hxc}\right]^{-1} \chi_0(\omega),$$

and substitute χ_0 with (5) to obtain

$$\chi(\omega) = \left[I - \hat{\Phi}\big((\omega + \mathrm{i}\eta)C - D\big)^{-1} \hat{\Phi}^T f_{Hxc}\right]^{-1} \hat{\Phi}\big((\omega + \mathrm{i}\eta)C - D\big)^{-1} \hat{\Phi}^T.$$

Using the Sherman–Morrison–Woodbury formula for the following matrix inverse

$$\begin{aligned}\left[I - \hat{\Phi}\big((\omega + \mathrm{i}\eta)C - D\big)^{-1} \hat{\Phi}^T f_{Hxc}\right]^{-1} &= I + \hat{\Phi}\left[\big((\omega + \mathrm{i}\eta)C - D\big) - \hat{\Phi}^T f_{Hxc} \hat{\Phi}\right]^{-1} \hat{\Phi}^T f_{Hxc} \\ &= I + \hat{\Phi}\big((\omega + \mathrm{i}\eta)C - \Omega\big)^{-1} \hat{\Phi}^T f_{Hxc},\end{aligned}$$

where Ω is defined by (7). We obtain

$$\begin{aligned}\chi(\omega) =& \left[I + \hat{\Phi}\big((\omega + \mathrm{i}\eta)C - \Omega\big)^{-1} \hat{\Phi}^T f_{Hxc}\right] \hat{\Phi}\big((\omega + \mathrm{i}\eta)C - D\big)^{-1} \hat{\Phi}^T \\ =& \hat{\Phi}\left[I + \big((\omega + \mathrm{i}\eta)C - \Omega\big)^{-1}(\Omega - D)\right] \big((\omega + \mathrm{i}\eta)C - D\big)^{-1} \hat{\Phi}^T \\ =& \hat{\Phi}\left[I + \big((\omega + \mathrm{i}\eta)C - \Omega\big)^{-1}\big(\Omega - (\omega + \mathrm{i}\eta)C + (\omega + \mathrm{i}\eta)C - D\big)\right] \big((\omega + \mathrm{i}\eta)C - D\big)^{-1} \hat{\Phi}^T \\ =& \hat{\Phi}\big((\omega + \mathrm{i}\eta)C - \Omega\big)^{-1} \hat{\Phi}^T.\end{aligned} \tag{30}$$

C. Eigen-decomposition of the Casida Matrix

It is known (see, e.g., [27, Theorem 3]) that when both $M \equiv A + B$ and $K \equiv A - B$ are positive definite, the Casida matrix, H, defined in (11) admits an eigen-decomposition of the form

$$\begin{bmatrix} A & B \\ -B & -A \end{bmatrix} = \begin{bmatrix} U & V \\ V & U \end{bmatrix} \begin{bmatrix} \Lambda & \\ & -\Lambda \end{bmatrix} \begin{bmatrix} U & V \\ V & U \end{bmatrix}^{-1} = \begin{bmatrix} U & V \\ V & U \end{bmatrix} \begin{bmatrix} \Lambda & \\ & -\Lambda \end{bmatrix} \begin{bmatrix} U & -V \\ -V & U \end{bmatrix}^{T},$$

where U, V are real $n_o n_v \times n_o n_v$ matrices satisfying $U^T U - V^T V = I$, $U^T V - V^T U = 0$, and $\Lambda = \mathrm{diag}\{\lambda_1, \ldots, \lambda_{n_o n_v}\}$. Let $U = [u_1, \ldots, u_{n_o n_v}]$, $V = [v_1, \ldots, v_{n_o n_v}]$. Then the right and left eigenvectors of H associated with λ_i are $\left[u_i^T, v_i^T\right]^T$ and $\left[u_i^T, -v_i^T\right]^T$, respectively. In addition, we have

$$\begin{aligned}
\langle x|\chi(\omega)|x\rangle &= \hat{x}^T \left[(\omega + \mathrm{i}\eta)I - H\right]^{-1} C\hat{x}^T \\
&= \left[\hat{x}_1^T, \hat{x}_1^T\right] \left((\omega + \mathrm{i}\eta)I - \begin{bmatrix} A & B \\ -B & -A \end{bmatrix}\right)^{-1} \begin{bmatrix} \hat{x}_1 \\ -\hat{x}_1 \end{bmatrix} \\
&= \left[\hat{x}_1^T, \hat{x}_1^T\right] \begin{bmatrix} U & V \\ V & U \end{bmatrix} \left((\omega + \mathrm{i}\eta)I - \begin{bmatrix} \Lambda & \\ & -\Lambda \end{bmatrix}\right)^{-1} \begin{bmatrix} U & -V \\ -V & U \end{bmatrix}^T \begin{bmatrix} \hat{x}_1 \\ -\hat{x}_1 \end{bmatrix} \\
&= \left[\hat{x}_1^T(U+V), \hat{x}_1^T(U+V)\right] \left((\omega + \mathrm{i}\eta)I - \begin{bmatrix} \Lambda & \\ & -\Lambda \end{bmatrix}\right)^{-1} \begin{bmatrix} (U+V)^T\hat{x}_1 \\ -(U+V)^T\hat{x}_1 \end{bmatrix} \\
&= \sum_{i=1}^{n_o n_v} \left(\frac{\left[\hat{x}_1^T(u_i + v_i)\right]^2}{\omega - \lambda_i + \mathrm{i}\eta} - \frac{\left[\hat{x}_1^T(u_i + v_i)\right]^2}{\omega + \lambda_i + \mathrm{i}\eta} \right). \qquad (31)
\end{aligned}$$

In the limit of $\eta \to 0^+$, the imaginary part of (31) becomes

$$\begin{aligned}
\lim_{\eta \to 0^+} \mathrm{Im}\langle x|\chi(\omega)|x\rangle &= -\pi \sum_{i=1}^{n_o n_v} \left(\left[\hat{x}_1^T(u_i + v_i)\right]^2 \delta(\omega - \lambda_i) - \left[\hat{x}_1^T(u_i + v_i)\right]^2 \delta(\omega + \lambda_i) \right). \\
&= -2\pi\, \mathrm{sign}(\omega) \sum_{i=1}^{n_o n_v} \lambda_i \left[\hat{x}_1^T(u_i + v_i)\right]^2 \delta(\omega^2 - \lambda_i^2).
\end{aligned}$$

Here we make use of the identity

$$\delta(\omega - \lambda_i) - \delta(\omega + \lambda_i) = 2\lambda_i\, \mathrm{sign}(\omega)\delta(\omega^2 - \lambda_i^2),$$

which can be proved by showing

$$\int_{-\infty}^{\infty} \left[\delta(\omega - \lambda_i) - \delta(\omega + \lambda_i)\right] \xi(\omega) d\omega = \int_{-\infty}^{\infty} 2\lambda_i\, \mathrm{sign}(\omega)\delta(\omega^2 - \lambda_i^2)\xi(\omega) d\omega.$$

The right hand side of the above equation can be evaluated by separating the integration from $-\infty$ to ∞ into integrations from $-\infty$ to 0 and from 0 to ∞, and making a change of variable $\omega' = \omega^2$, $\omega = \pm\sqrt{\omega'}$, and $d\omega = \pm \frac{d\omega'}{2\sqrt{\omega'}}$.

It can be verified that the normalization condition $U^T U - V^T V = I$ implies that $(U+V)^T(U-V) = I$ and $(U+V)^T M (U+V) = \Lambda$. Consequently, we have

$$M = (U-V)\Lambda(U-V)^T, \qquad K = (U+V)\Lambda(U+V)^T.$$

The eigen-decompositions of MK and KM are thus given by

$$\begin{aligned}
MK &= (U-V)\Lambda^2(U-V)^{-1} = (U-V)\Lambda^2(U+V)^T, \\
KM &= (U+V)\Lambda^2(U+V)^{-1} = (U+V)\Lambda^2(U-V)^T.
\end{aligned}$$

Therefore, the vectors $u_i + v_i$ $(i = 1, 2, \ldots, n_o n_v)$ in (31) are the right eigenvector of KM, as well as the left eigenvector of MK, both associated with the eigenvalues λ_i^2. These results lead to

$$\begin{aligned}
\alpha_{x,x}(\omega) &= 2\,\mathrm{sign}(\omega) \sum_{i=1}^{n_o n_v} \lambda_i \left[\hat{x}_1^T(u_i + v_i)\right]^2 \delta(\omega^2 - \lambda_i^2) \\
&= 2\,\mathrm{sign}(\omega)\hat{x}_1^T(U+V)\Lambda\delta(\omega^2 I - \Lambda^2)(U+V)^T\hat{x}_1 \\
&= 2\,\mathrm{sign}(\omega)\hat{x}_1^T(U+V)\Lambda(U+V)^T(U-V)\delta(\omega^2 I - \Lambda^2)(U+V)^T\hat{x}_1 \\
&= 2\,\mathrm{sign}(\omega)\hat{x}_1^T K\delta(\omega^2 I - MK)\hat{x}_1.
\end{aligned}$$

Thus we have proved (14), which is an expression for $\alpha_{x,x}(\omega)$ that involves matrices of dimension $n_o n_v \times n_o n_v$ (instead of $2n_o n_v \times 2n_o n_v$ in (8)).

References

[1] E. Runge and E. K. U. Gross, "Density-functional theory for time-dependent systems," *Phys. Rev. Lett.*, vol. 52, pp. 997–1000, Mar 1984.

[2] M. E. Casida, "Time-dependent density functional response theory for molecules," in *Recent Advances in Density Functional Methods* (D. E. Chong, ed.), pp. 155–192, World Scientific, Singapore, 1995.

[3] M. E. Casida, "Time-dependent density functional response theory of molecular systems: Theory, computational methods, and functionals," in *Recent Developments and Application of Modern Density Functional Theory* (J. M. Seminario, ed.), pp. 391–439, Elsevier, Amsterdam, 1996.

[4] S. Tretiak, C. M. Isborn, A. M. Niklasson, and M. Challacombe, "Representation independent algorithms for molecular response calculations in time-dependent self-consistent field theories," *J. Chem. Phys.*, vol. 130, no. 5, p. 054111, 2009.

[5] C. A. Ullrich, *Time-dependent Density Functional Theory.* Oxford University, 2012.

[6] J. Olsen, H. J. A. Jensen, and P. Jørgensen, "Solution of the large matrix equations which occur in response theory," *J. Comput. Phys.*, vol. 74, pp. 265–282, 1988.

[7] I. Vasiliev, S. Öğüt, and J. R. Chelikowsky, "Ab initio absorption spectra and optical gaps in nanocrystalline silicon," *Phys. Rev. Lett.*, vol. 86, p. 1813, 2001.

[8] Y. Wang, K. Lopata, S. A. Chambers, N. Govind, and P. V. Sushko, "Optical absorption and band gap reduction in (fe1–x cr x) 2o3 solid solutions: A first-principles study," *J. Phys. Chem. C*, vol. 117, no. 48, pp. 25504–25512, 2013.

[9] S. Tussupbayev, N. Govind, K. Lopata, and C. J. Cramer, "Comparison of real-time and linear-response time-dependent density functional theories for molecular chromophores ranging from sparse to high densities of states," *J. Chem. Theory Comput.*, vol. 11, no. 3, pp. 1102–1109, 2015.

[10] S. Hirata and M. Head-Gordon, "Time-dependent density functional theory within the Tamm–Dancoff approximation," *Chem. Phys. Lett.*, vol. 314, no. 3–4, pp. 291 – 299, 1999.

[11] R. N. Silver, H. Röder, A. F. Voter, and J. D. Kress, "Kernel polynomial approximations for densities of states and spectral functions," *J. Comput. Phys.*, vol. 124, pp. 115–130, 1996.

[12] A. B. Gordienko and S. I. Filippov, "Kernel polynomial method as an efficient $O(N^2)$ scheme for optical spectra calculations including electronhole interaction," *Phys. Status Solidi B*, vol. 251, no. 3, pp. 628–632, 2014.

[13] D. Rocca, R. Gebauer, Y. Saad, and S. Baroni, "Turbo charging time-dependent density-functional theory with Lanczos chains," *J. Chem. Phys.*, vol. 128, p. 154105, 2008.

[14] Z. Bai and R.-C. Li, "Minimization principles for the linear response eigenvalue problem I: Theory," *SIAM J. Matrix Anal. Appl.*, vol. 33, no. 4, pp. 1075–1100, 2012.

[15] P. Benner, V. Mehrmann, and H. Xu, "A numerically stable, structure preserving method for computing the eigenvalues of real Hamiltonian or symplectic pencils," *Numer. Math.*, vol. 78, pp. 329–357, 1998.

[16] Z. Bai and R.-C. Li, "Minimization principles for the linear response eigenvalue problem II: Computation," *SIAM J. Matrix Anal. Appl.*, vol. 34, no. 2, pp. 392–416, 2013.

[17] M. Challacombe, "Linear scaling solution of the time-dependent self-consistent-field equations," *Computation*, vol. 2, no. 1, pp. 1–11, 2014.

[18] L. Lin, Y. Saad, and C. Yang, "Approximating spectral densities of large matrices," *to appear in SIAM Rev.*, 2015.

[19] D. Jackson, *The theory of approximation.* American Mathematical Society, New York, 1930.

[20] C. C. Paige, *The computation of eigenvalues and eigenvectors of very large sparse matrices*. PhD thesis, University of London, London, England, 1971.

[21] G. Moro and J. H. Freed, "Classical time-correlation function and the Lanczos algorithm," *J. Chem. Phys.*, vol. 75, no. 6, p. 3157, 1975.

[22] R. Haydock, V. Heine, and M. J. Kelly, "Electronic structure based on the local atomic environment for tight-binding bands," *J. Phys. C: Solid State Phys.*, vol. 5, p. 2845, 1972.

[23] M. Cohen and P. S. Kelly, "Hartree–Fock wave functions for excited states: II. simplification of the orbital equations," *Can. J. Phys.*, vol. 44, no. 12, pp. 3227–3240, 1966.

[24] M. Grüning, A. Marini, and X. Gonze, "Exciton–plasmon states in nanoscale materials: Breakdown of the Tamm–Dancoff approximation," *Nano Lett.*, vol. 9, no. 8, pp. 2820–2824, 2009.

[25] K. Lopata and N. Govind, "Modeling fast electron dynamics with real-time time-dependent density functional theory: Application to small molecules and chromophores," *J. Chem. Theory Comput.*, vol. 7, no. 5, pp. 1344–1355, 2011.

[26] M. Valiev, E. Bylaska, N. Govind, K. Kowalski, T. Straatsma, H. V. Dam, D. Wang, J. Nieplocha, E. Apra, T. Windus, and W. de Jong, "NWChem: A comprehensive and scalable open-source solution for large scale molecular simulations," *Comput. Phys. Commun.*, vol. 181, no. 9, pp. 1477–1489, 2010.

[27] M. Shao, F. H. Jornada, C. Yang, J. Deslippe, and S. G. Louie, "Structure preserving parallel algorithms for solving the Bethe–Salpeter eigenvalue problem," *submitted to Linear Algebra Appl.*, 2015. arXiv:1501.03830.