AIP | The Journal of Chemical Physics

Accelerating molecular property calculations with nonorthonormal Krylov space methods

Filipp Furche, Brandon T. Krull, Brian D. Nguyen, and Jake Kwon

Citation: The Journal of Chemical Physics **144**, 174105 (2016); doi: 10.1063/1.4947245
View online: http://dx.doi.org/10.1063/1.4947245
View Table of Contents: http://scitation.aip.org/content/aip/journal/jcp/144/17?ver=pdfcov
Published by the AIP Publishing

Accelerating molecular property calculations with nonorthonormal Krylov space methods

Filipp Furche,[a] Brandon T. Krull, Brian D. Nguyen, and Jake Kwon[b]
Department of Chemistry, University of California, Irvine, 1102 Natural Sciences II, Irvine, California 92697-2025, USA

(Received 19 January 2016; accepted 8 April 2016; published online 3 May 2016)

We formulate Krylov space methods for large eigenvalue problems and linear equation systems that take advantage of decreasing residual norms to reduce the cost of matrix-vector multiplication. The residuals are used as subspace basis without prior orthonormalization, which leads to generalized eigenvalue problems or linear equation systems on the Krylov space. These nonorthonormal Krylov space (nKs) algorithms are favorable for large matrices with irregular sparsity patterns whose elements are computed on the fly, because fewer operations are necessary as the residual norm decreases as compared to the conventional method, while errors in the desired eigenpairs and solution vectors remain small. We consider real symmetric and symplectic eigenvalue problems as well as linear equation systems and Sylvester equations as they appear in configuration interaction and response theory. The nKs method can be implemented in existing electronic structure codes with minor modifications and yields speed-ups of 1.2-1.8 in typical time-dependent Hartree-Fock and density functional applications without accuracy loss. The algorithm can compute entire linear subspaces simultaneously which benefits electronic spectra and force constant calculations requiring many eigenpairs or solution vectors. The nKs approach is related to difference density methods in electronic ground state calculations and particularly efficient for integral direct computations of exchange-type contractions. By combination with resolution-of-the-identity methods for Coulomb contractions, three- to fivefold speed-ups of hybrid time-dependent density functional excited state and response calculations are achieved. *Published by AIP Publishing.* [http://dx.doi.org/10.1063/1.4947245]

I. INTRODUCTION

Large eigenvalue problems and linear equation systems are the core of molecular electronic property calculations.[1] In response theory,[2–4] dispersive properties such as polarizabilities or optical rotations are obtained by solving linear equation systems, while resonant quantities such as excitation energies and transition moments require the solution of an eigenvalue problem. These problems are so large that direct solution, e.g., using elimination techniques, is prohibitive, and often the coefficient matrices cannot even be stored.

Iterative subspace or Krylov space methods[5] have revolutionized molecular property calculations because they only require the computation of matrix-vector products and thus eliminate the need for explicit calculation and storage of large matrices. Starting from an initial vector space $\mathbf{V}$, these methods generate a Krylov sequence $\{\mathbf{V}, \mathbf{AV}, \mathbf{A}^2\mathbf{V}, \ldots\}$ by successive multiplication with the coefficient matrix $\mathbf{A}$. This Krylov sequence is used as a basis for expanding the eigenvector or solution vector. Davidson's pioneering work in the 1970s[6] led to a highly efficient iterative eigenvalue solver that continues to be very widely used. Concurrently, integral direct configuration interaction methods which dramatically reduced the cost of matrix-vector multiplication by evaluating matrix elements in the basis of localized atomic orbitals (AOs) were introduced by Roos.[7] The 1980s saw the first efficient integral-direct implementations of response properties based on these techniques.[8,9] Olsen, Jensen, and Jørgensen presented an extension of Davidson's method to symplectic eigenvalue problems and linear problems in 1988.[10] It soon became clear that the computation of the matrix-vector products in time-dependent Hartree-Fock (TDHF) theory is equivalent to a Fock matrix build.[11] This led to further substantial speed-ups in response calculations, since Fock matrix build is the rate-determining step in most self-consistent field calculations of molecular ground states and has thus been at the center of method development efforts for many years.[12–15] The close analogy between matrix-vector product computation in response theory and ground-state calculations has been exploited for efficient implementations of adiabatic time-dependent density functional theory (TDDFT)[16–19] and coupled cluster (CC) response theory;[20] it has also inspired efficient approximation methods for molecular response calculations originally developed for ground states, including resolution-of-the-identity methods for Coulomb (RI-J)[21,22] and exchange contractions,[23] Laplace transform and linear scaling techniques,[24–29] as well as semi-numerical and pseudospectral approximations.[30–32]

However, an important technique for accelerating ground-state calculations has been elusive for molecular property calculations: Recursive Fock matrix construction, also known as the difference density method, uses the linear dependence

[a]Electronic mail: filipp.furche@uci.edu
[b]Present address: Department of Chemical Engineering, Columbia University, New York, NY.

of the Fock matrix on the density matrix to reduce the cost of Fock matrix construction after the first iteration.[12,13,33] Thus, subsequent self-consistent field iterations only require a Fock matrix build using the difference density matrix. Since the norm of the difference density matrix is smaller than the norm of the full density matrix, fewer integrals need to be computed, leading to speed-ups of 1.25-1.75 compared to the conventional method. This method is particularly effective in combination with integral-direct fast exchange algorithms[14,34] which critically depend on a small density matrix norm.[35]

Molecular property calculations compute changes of the ground state wavefunction and density under small external perturbations. Thus, it might appear that these methods already build on difference densities and that no further acceleration is possible. Here we show that this assumption is incorrect, and that molecular property calculations can be considerably accelerated by techniques akin to recursive Fock matrix build. We introduce a modification of iterative subspace algorithms that takes advantage of decreasing residual norms to reduce the cost of the matrix-vector products. The key idea is to use a nonorthonormal basis of residual vectors to construct the Krylov space. This leads to generalized subspace eigenvalue problems or linear equation systems, which can be solved by standard methods. In 1989, Davidson mentioned this possibility as an "alternative which has not been tried."[36] We show that this alternative is numerically stable and preferable to the conventional method for integral direct calculations.

We illustrate the use and the performance of these nonorthonormal Krylov space (nKs) algorithms for TDHF and hybrid TDDFT calculations of molecular response properties. Hybrid exchange is essential for accurate response properties[37–40] and non-adiabatic molecular dynamics simulations,[41] yet its computation remains a key bottleneck in most applications. Similar to ground-state difference density methods, the nKs algorithms are particularly effective for exchange-type contractions.

Sec. II outlines the nKs algorithms for solving eigenvalue problems and linear problems. We start with the conceptually simple real symmetric eigenvalue and linear problems for which iterative subspace methods were originally developed. These types of problems occur in configuration interaction (CI) theory and static response calculations; the Tamm-Dancoff approximation (TDA) to excitation energies also falls into this category.[11,42] However, frequency dependent response properties with correct pole structure require the solution of symplectic eigenvalue problems and linear problems, and thus we extend our methods to the symplectic case in Secs. II D and II E. We show that the Olsen, Jensen, and Jørgensen method[10] can be obtained by minimizing a quadratic functional. Since more than one eigenpair or solution vector is needed in most applications, we consider blocked nKs algorithms from the outset, which expand all desired eigenvectors or solution vectors into one single subspace. Sec. II F presents a theoretical error analysis along with a numerical example. The nKs algorithms are optimal in the sense that they minimize the norm of the new basis vectors, as explained in Sec. II G. We also discuss relationships of the nKs method to other iterative subspace algorithms and give recommendations for best implementation practices. Our method explicitly constructs the Krylov subspace basis and thus does not exhibit the instabilities observed in recursive Krylov space methods. We illustrate the implementation and performance of the nKs algorithms for hybrid time-dependent density functional calculations of excitation energies and response properties in Sec. III. Conclusions are presented in Sec. IV.

II. ITERATIVE SUBSPACE METHODS

A. Krylov space methods

The eigenvalue problems and linear equation systems considered here can be recast into optimization problems for a quadratic functional on the underlying vector spaces. This functional is then extremized on a Krylov space which is iteratively expanded, see Figure 1. Starting from an initial vector space which must not be orthogonal to the subspace of desired eigenvectors or solution vectors, the Rayleigh and overlap matrices are computed which are the subspace projections of the coefficient matrix and the metric. The resulting small generalized eigenvalue problems or linear systems are easily solved and yield best approximations to the desired eigenvectors or solution vectors for the given subspace (Ritz step). The residual or error vector of the approximate eigenvectors or solution vectors, which is identical to the gradient of the functional, is computed. If the residual norm is below a specified tolerance, the iteration is stopped. If not, the residual is preconditioned and added to the subspace basis. For further details on the theory and implementation of Krylov space methods, the reader is referred to the monograph by Saad.[5]

B. Real symmetric eigenvalue problems

Computation of the p lowest eigenpairs of the real symmetric $n \times n$ matrix $\mathbf{A}$ amounts to solving

$$\mathbf{AX} = \mathbf{X\Omega}, \tag{1}$$

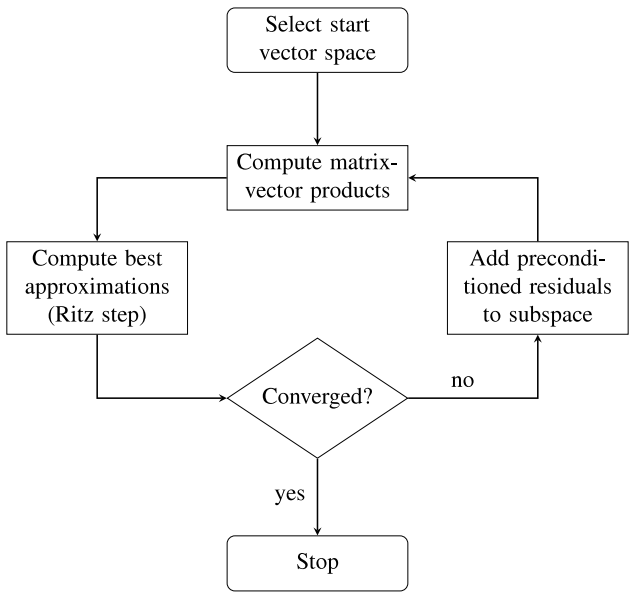

FIG. 1. Flow-chart of Krylov space methods.

$\mathbf{\Omega} = \mathrm{diag}(\Omega_1, \ldots, \Omega_p)$ is a $p \times p$ diagonal matrix containing the lowest p eigenvalues, and the $n \times p$ matrix $\mathbf{X}$ contains the corresponding eigenvectors satisfying the orthonormality condition

$$\mathbf{X}^T\mathbf{X} = \mathbf{1}. \tag{2}$$

Examples of symmetric eigenvalue problems arising in molecular property calculations are excitation energies within configuration interaction singles (CIS)[43] theory or non-hybrid TDDFT.[16,44] For notational clarity, the rank of the identity matrix is implied throughout this paper, e.g., it is p in Eq. (2). We further assume that the spectrum of $\mathbf{A}$ is non-degenerate; nevertheless, the present method still applies with minor modifications even in the presence of degeneracies as long as the p-th and $(p+1)$-th eigenvalues of $\mathbf{A}$ are non-degenerate. It is convenient to consider the problem of minimizing the functional

$$F[\mathbf{X}, \mathbf{\Omega}] = \langle \mathbf{X}^T\mathbf{A}\mathbf{X} - \mathbf{\Omega}(\mathbf{X}^T\mathbf{X} - \mathbf{1}) \rangle \tag{3}$$

instead of Eq. (1). Brackets $\langle \cdot \rangle$ are used to denote the trace operation. Here $\mathbf{\Omega}$ appears as a Lagrange multiplier matrix enforcing orthonormality of the eigenvectors; $\mathbf{\Omega}$ is symmetric but not necessarily diagonal. F is invariant under unitary transformations of the p eigenvectors,

$$F[\mathbf{X}, \mathbf{\Omega}] = F[\mathbf{X}\mathbf{U}, \mathbf{U}^T\mathbf{\Omega}\mathbf{U}], \tag{4}$$

and thus minimization of F only yields the desired invariant subspace of eigenvectors.

In iterative subspace methods, the eigenvectors $\mathbf{X}$ are expanded into a Krylov subspace $\mathcal{K}^{(k)}$ of dimension $q_k \geq p$. q_k increases with the iteration number $k = 1, \ldots$. The k-th iterate $\mathbf{X}^{(k)}$ is thus a linear combination,

$$\mathbf{X}^{(k)} = \mathbf{V}^{(k)}\mathbf{x}^{(k)}, \tag{5}$$

where $\mathbf{V}^{(k)}$ is $n \times q_k$ and contains the subspace basis vectors, and the $q_k \times p$ matrix $\mathbf{x}^{(k)}$ is the projection of the eigenvectors $\mathbf{X}$ onto $\mathcal{K}^{(k)}$. The projection of $\mathbf{A}$ onto $\mathcal{K}^{(k)}$ is

$$\mathbf{a}^{(k)} = \mathbf{V}^{(k)T}\mathbf{A}\mathbf{V}^{(k)}. \tag{6}$$

In the original Davidson method, orthonormality of the subspace vectors $\mathbf{V}^{(k)}$ is enforced explicitly in each iteration by modified Gram-Schmidt (MGS) orthogonalization.[6,45] The gist of the nKs method is to relax this constraint. Thus, the k-th subspace overlap matrix

$$\mathbf{s}^{(k)} = \mathbf{V}^{(k)T}\mathbf{V}^{(k)} \tag{7}$$

is generally not equal to $\mathbf{1}$. As in the original Davidson method, computation of the $q_k \times q_k$ matrices $\mathbf{a}^{(k)}$ and $\mathbf{s}^{(k)}$ requires only matrix-vector and vector-vector operations.

In the following Ritz step, the best approximations to the eigenpairs $(\mathbf{X}^{(k)}, \mathbf{\Omega}^{(k)})$ are obtained by minimizing F for $\mathbf{X} \in \mathcal{K}^{(k)}$. This leads to the reduced generalized eigenvalue problem

$$\mathbf{a}^{(k)}\mathbf{x}^{(k)} = \mathbf{s}^{(k)}\mathbf{x}^{(k)}\mathbf{\Omega}^{(k)}, \tag{8}$$

subject to the orthonormality constraint

$$\mathbf{x}^{(k)T}\mathbf{s}^{(k)}\mathbf{x}^{(k)} = \mathbf{1}. \tag{9}$$

Solution of Eq. (8) is inexpensive since $q_k \ll n$ and possible by standard eigenvalue solvers after factorization of $\mathbf{s}^{(k)}$, see Subsection 1 of the Appendix. Typically, $\mathbf{\Omega}^{(k)}$ is chosen to be diagonal at this point. Subspace eigenvalue problem (8) has q_k solutions but only the lowest p eigenpairs minimize F.

The gradient of F with respect to $\mathbf{X}$ is the $n \times p$ residual matrix

$$\mathbf{R}^{(k)} = \mathbf{A}\mathbf{X}^{(k)} - \mathbf{X}^{(k)}\mathbf{\Omega}^{(k)}. \tag{10}$$

The columns of $\mathbf{R}^{(k)}$ contain the residual vectors of the approximate eigenvectors $\mathbf{X}^{(k)}$. Conventionally, the convergence of the ith eigenpair, $i = 1, \ldots, p$, is checked by computing the Euclidean residual norm $\|\mathbf{R}_i^{(k)}\|_2$, and residual vectors whose norms are below the convergence threshold are eliminated. An alternative that does not require diagonal $\mathbf{\Omega}^{(k)}$ is the Frobenius norm of the residual matrix,

$$\|\mathbf{R}^{(k)}\|_F = \sqrt{\langle \mathbf{R}^{(k)T}\mathbf{R}^{(k)} \rangle}. \tag{11}$$

Since $\|\mathbf{R}^{(k)}\|_F \geq \|\mathbf{R}_i^{(k)}\|_2$, this convergence criterion is somewhat more stringent. The rank of $\mathbf{R}^{(k)}$ may subsequently be reduced by singular value decomposition (SVD).

If the residual norm is above the convergence threshold, the preconditioned residual $\tilde{\mathbf{R}}^{(k)}$ is computed from the linear equation

$$\mathbf{D}\tilde{\mathbf{R}}^{(k)} - \tilde{\mathbf{R}}^{(k)}\mathbf{\Omega}^{(k)} = \mathbf{R}^{(k)}. \tag{12}$$

The preconditioner $\mathbf{D}$ is a diagonal matrix whose spectrum approximates the spectrum of $\mathbf{A}$.[6,45] $\mathbf{D}$ is typically chosen on physical grounds, e.g., by considering non-interacting particles. Finally, the subspace basis is extended by adding the preconditioned residuals,

$$\mathbf{V}^{(k+1)} = \begin{bmatrix} \mathbf{V}^{(k)} & \tilde{\mathbf{R}}^{(k)} \end{bmatrix}, \tag{13}$$

and the $(k+1)$th iteration starts. The iteration terminates when the residual norm is below the specified convergence threshold.

The main advantage of the nKs variant of Davidson's algorithm is the possibility to reduce the cost of matrix-vector multiplication by exploiting the decreasing residual norm after the first iteration(s). The nKs method also does not require MGS orthonormalization of the subspace basis, but only formation of the subspace overlap matrix $\mathbf{s}^{(k)}$ and subsequent Cholesky decomposition, which may be somewhat preferable for very large vector space dimensions n.

C. Linear equations with a symmetric coefficient matrix

While resonant properties such as excitation energies and transition strengths are related to the solutions of eigenvalue problem (1), the Sylvester equation

$$\mathbf{A}\mathbf{X} - \mathbf{X}\boldsymbol{\omega} = \mathbf{P} \tag{14}$$

determines dispersive properties such as linear or non-linear polarizabilities. Here, $\mathbf{A}$ is real and symmetric $n \times n$ as before, $\mathbf{X}$ is $n \times p$ and contains the p solution vectors, the $p \times p$ diagonal matrix $\boldsymbol{\omega} = \mathrm{diag}(\omega_1, \ldots, \omega_p)$ contains the frequencies, and the $n \times p$ right-hand side (RHS) matrix

P contains the vectors corresponding to each perturbation. Examples of linear equations with a symmetric coefficient matrix are the coupled-perturbed Hartree-Fock[46] and Kohn-Sham (KS) equations,[47] which typically need to be solved for p perturbations simultaneously, e.g., for computing the nuclear force constant matrix.[48] Non-zero frequency matrices $\boldsymbol{\omega}$ arise in dynamic polarizability calculations using non-hybrid density functionals.[16,44]

Instead of solving linear problem (14) directly, which is prohibitive for large n, we attempt to make the functional

$$F[\mathbf{X}|\boldsymbol{\omega},\mathbf{P}] = \langle \mathbf{X}^T\mathbf{A}\mathbf{X} - \boldsymbol{\omega}\mathbf{X}^T\mathbf{X} - \mathbf{X}^T\mathbf{P} - \mathbf{P}^T\mathbf{X}\rangle \quad (15)$$

stationary with respect to **X**; the dependence on $\boldsymbol{\omega}$ and **P** is parametric. F is invariant under $p \times p$ unitary transformations **U**,

$$F[\mathbf{X}|\boldsymbol{\omega},\mathbf{P}] = F[\mathbf{X}\mathbf{U}|\mathbf{U}^T\boldsymbol{\omega}\mathbf{U},\mathbf{P}\mathbf{U}], \quad (16)$$

but for a given $\boldsymbol{\omega}$ and **P**, the solution is unique.

As before, the approximate solution in the k-th iteration is iteratively expanded into a linear combination of q_k subspace basis vectors, see Eq. (5). The Rayleigh matrix $\mathbf{a}^{(k)}$ and the subspace overlap matrix $\mathbf{s}^{(k)}$ are computed according to Eqs. (6) and (7). Similarly, the projection of the RHS vectors on the Krylov space $\mathcal{K}^{(k)}$ is obtained as

$$\mathbf{p}^{(k)} = \mathbf{V}^{(k)T}\mathbf{P}. \quad (17)$$

The Ritz step now requires solution of the reduced "generalized" Sylvester equation

$$\mathbf{a}^{(k)}\mathbf{x}^{(k)} - \mathbf{s}^{(k)}\mathbf{x}^{(k)}\boldsymbol{\omega} = \mathbf{p}^{(k)}. \quad (18)$$

The reduced problem is typically much smaller than the original one and may be solved by the techniques outlined in Subsection 2 of the Appendix. The norm of the residual matrix

$$\mathbf{R}^{(k)} = \mathbf{A}\mathbf{X}^{(k)} - \mathbf{X}^{(k)}\boldsymbol{\omega} - \mathbf{P} \quad (19)$$

may be used to test convergence, and the rank of $\mathbf{R}^{(k)}$ may be reduced either vector by vector or by SVD. In analogy to Eq. (12), the preconditioned residual $\tilde{\mathbf{R}}^{(k)}$ is determined from

$$\mathbf{D}\tilde{\mathbf{R}}^{(k)} - \tilde{\mathbf{R}}^{(k)}\boldsymbol{\omega} = \mathbf{R}^{(k)}. \quad (20)$$

D. Symplectic eigenvalue problems

Molecular TDHF and TDDFT excitation energies are eigenvalues of a symplectic rather than symmetric eigenvalue problem,[44,49]

$$\begin{pmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{B} & \mathbf{A}\end{pmatrix}\begin{pmatrix}\mathbf{X} & \mathbf{Y}\\ \mathbf{Y} & \mathbf{X}\end{pmatrix} = \begin{pmatrix}\mathbf{1} & \mathbf{0}\\ \mathbf{0} & -\mathbf{1}\end{pmatrix}\begin{pmatrix}\mathbf{X} & \mathbf{Y}\\ \mathbf{Y} & \mathbf{X}\end{pmatrix}\begin{pmatrix}\boldsymbol{\Omega} & \mathbf{0}\\ \mathbf{0} & -\boldsymbol{\Omega}\end{pmatrix}. \quad (21)$$

A and **B** are real symmetric $n \times n$ matrices, **X** and **Y** are $n \times p$, and $\boldsymbol{\Omega} = \mathrm{diag}(\Omega_1,\ldots,\Omega_p)$ is $p \times p$ diagonal as in Section II B. We assume that all eigenvalues are real and non-degenerate. However, eigenvector orthonormality condition (2) is now replaced by the symplectic orthonormality constraint

$$\begin{pmatrix}\mathbf{X} & \mathbf{Y}\\ \mathbf{Y} & \mathbf{X}\end{pmatrix}^T\begin{pmatrix}\mathbf{1} & \mathbf{0}\\ \mathbf{0} & -\mathbf{1}\end{pmatrix}\begin{pmatrix}\mathbf{X} & \mathbf{Y}\\ \mathbf{Y} & \mathbf{X}\end{pmatrix} = \begin{pmatrix}\mathbf{1} & \mathbf{0}\\ \mathbf{0} & -\mathbf{1}\end{pmatrix}. \quad (22)$$

The symplectic structure of eigenvalue problem (21) ensures that eigenvalues are ± pairs, which is required by time-reversal symmetry. Multi-configurational self-consistent field,[50] higher-order propagator,[51] and orbital-optimized CC[52] methods also require the solution of symplectic eigenvalue problems to determine excitation energies and transition moments.

In the symplectic case, the optimization problem corresponding to eigenvalue problem (21) requires minimization of

$$F\left[\begin{pmatrix}\mathbf{X} & \mathbf{Y}\\ \mathbf{Y} & \mathbf{X}\end{pmatrix},\begin{pmatrix}\boldsymbol{\Omega} & \boldsymbol{\Lambda}\\ \boldsymbol{\Lambda} & \boldsymbol{\Omega}\end{pmatrix}\right] = \left|\begin{pmatrix}\mathbf{X} & \mathbf{Y}\\ \mathbf{Y} & \mathbf{X}\end{pmatrix}^T\begin{pmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{B} & \mathbf{A}\end{pmatrix}\begin{pmatrix}\mathbf{X} & \mathbf{Y}\\ \mathbf{Y} & \mathbf{X}\end{pmatrix} - \begin{pmatrix}\boldsymbol{\Omega} & \boldsymbol{\Lambda}\\ -\boldsymbol{\Lambda} & -\boldsymbol{\Omega}\end{pmatrix} \times\left[\begin{pmatrix}\mathbf{X} & \mathbf{Y}\\ \mathbf{Y} & \mathbf{X}\end{pmatrix}^T\begin{pmatrix}\mathbf{1} & \mathbf{0}\\ \mathbf{0} & -\mathbf{1}\end{pmatrix}\begin{pmatrix}\mathbf{X} & \mathbf{Y}\\ \mathbf{Y} & \mathbf{X}\end{pmatrix} - \begin{pmatrix}\mathbf{1} & \mathbf{0}\\ \mathbf{0} & -\mathbf{1}\end{pmatrix}\right]\right|. \quad (23)$$

The symmetric $p \times p$ Lagrange multipliers $\boldsymbol{\Omega}$ and $\boldsymbol{\Lambda}$ enforce the symplectic orthonormality condition (22). The need for the off-diagonal Lagrange multiplier matrix $\boldsymbol{\Lambda}$ to enforce $\mathbf{X}^T\mathbf{Y} - \mathbf{Y}^T\mathbf{X} = \mathbf{0}$ has been overlooked previously.[53,54] F is invariant under unitary transformations satisfying

$$\begin{pmatrix}\mathbf{U} & \mathbf{T}\\ \mathbf{T} & \mathbf{U}\end{pmatrix}^T\begin{pmatrix}\mathbf{1} & \mathbf{0}\\ \mathbf{0} & \mathbf{1}\end{pmatrix}\begin{pmatrix}\mathbf{U} & \mathbf{T}\\ \mathbf{T} & \mathbf{U}\end{pmatrix} = \begin{pmatrix}\mathbf{1} & \mathbf{0}\\ \mathbf{0} & \mathbf{1}\end{pmatrix}, \quad (24)$$

where **U** and **T** are $p \times p$, i.e.,

$$F\left[\begin{pmatrix}\mathbf{X} & \mathbf{Y}\\ \mathbf{Y} & \mathbf{X}\end{pmatrix},\begin{pmatrix}\boldsymbol{\Omega} & \boldsymbol{\Lambda}\\ \boldsymbol{\Lambda} & \boldsymbol{\Omega}\end{pmatrix}\right] = F\left[\begin{pmatrix}\mathbf{X} & \mathbf{Y}\\ \mathbf{Y} & \mathbf{X}\end{pmatrix}\begin{pmatrix}\mathbf{U} & \mathbf{T}\\ \mathbf{T} & \mathbf{U}\end{pmatrix},\begin{pmatrix}\mathbf{U} & \mathbf{T}\\ \mathbf{T} & \mathbf{U}\end{pmatrix}^T\begin{pmatrix}\boldsymbol{\Omega} & \boldsymbol{\Lambda}\\ \boldsymbol{\Lambda} & \boldsymbol{\Omega}\end{pmatrix}\begin{pmatrix}\mathbf{U} & \mathbf{T}\\ \mathbf{T} & \mathbf{U}\end{pmatrix}\right]. \quad (25)$$

Thus, minimization of F only yields the lowest invariant subspace up to a unitary transformation. However, **U** and **T** are uniquely determined by requiring the Lagrange multiplier matrix $\left(\begin{smallmatrix}\boldsymbol{\Omega} & \boldsymbol{\Lambda}\\ \boldsymbol{\Lambda} & \boldsymbol{\Omega}\end{smallmatrix}\right)$ to be diagonal, as shown below.

The eigenvectors are iteratively expanded into a subspace $\mathcal{K}^{(k)}$ of dimension $4q_k$ according to

$$\begin{pmatrix}\mathbf{X}^{(k)} & \mathbf{Y}^{(k)}\\ \mathbf{Y}^{(k)} & \mathbf{X}^{(k)}\end{pmatrix} = \begin{pmatrix}\mathbf{V}^{(k)} & \mathbf{W}^{(k)}\\ \mathbf{W}^{(k)} & \mathbf{V}^{(k)}\end{pmatrix}\begin{pmatrix}\mathbf{x}^{(k)} & \mathbf{y}^{(k)}\\ \mathbf{y}^{(k)} & \mathbf{x}^{(k)}\end{pmatrix}. \quad (26)$$

Since the present nKs algorithm preserves the symplectic structure of the problem, only $2q_k$ vectors need to be stored and handled. The overlap matrix of the subspace basis vectors is

$$\begin{pmatrix}\boldsymbol{\sigma}^{(k)} & \boldsymbol{\pi}^{(k)}\\ -\boldsymbol{\pi}^{(k)} & -\boldsymbol{\sigma}^{(k)}\end{pmatrix} = \begin{pmatrix}\mathbf{V}^{(k)} & \mathbf{W}^{(k)}\\ \mathbf{W}^{(k)} & \mathbf{V}^{(k)}\end{pmatrix}^T\begin{pmatrix}\mathbf{1} & \mathbf{0}\\ \mathbf{0} & -\mathbf{1}\end{pmatrix}\begin{pmatrix}\mathbf{V}^{(k)} & \mathbf{W}^{(k)}\\ \mathbf{W}^{(k)} & \mathbf{V}^{(k)}\end{pmatrix}. \quad (27)$$

Similarly, the projection of $\left(\begin{smallmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{B} & \mathbf{A}\end{smallmatrix}\right)$ onto $\mathcal{K}^{(k)}$ is

$$\begin{pmatrix}\mathbf{a}^{(k)} & \mathbf{b}^{(k)}\\ \mathbf{b}^{(k)} & \mathbf{a}^{(k)}\end{pmatrix} = \begin{pmatrix}\mathbf{V}^{(k)} & \mathbf{W}^{(k)}\\ \mathbf{W}^{(k)} & \mathbf{V}^{(k)}\end{pmatrix}^T\begin{pmatrix}\mathbf{A} & \mathbf{B}\\ \mathbf{B} & \mathbf{A}\end{pmatrix}\begin{pmatrix}\mathbf{V}^{(k)} & \mathbf{W}^{(k)}\\ \mathbf{W}^{(k)} & \mathbf{V}^{(k)}\end{pmatrix}. \quad (28)$$

Minimization of F on $\mathcal{K}^{(k)}$ leads to the reduced generalized symplectic eigenvalue problem

$$\begin{pmatrix} \mathbf{a}^{(k)} & \mathbf{b}^{(k)} \\ \mathbf{b}^{(k)} & \mathbf{a}^{(k)} \end{pmatrix} \begin{pmatrix} \mathbf{x}^{(k)} & \mathbf{y}^{(k)} \\ \mathbf{y}^{(k)} & \mathbf{x}^{(k)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\sigma}^{(k)} & \boldsymbol{\pi}^{(k)} \\ -\boldsymbol{\pi}^{(k)} & -\boldsymbol{\sigma}^{(k)} \end{pmatrix} \begin{pmatrix} \mathbf{x}^{(k)} & \mathbf{y}^{(k)} \\ \mathbf{y}^{(k)} & \mathbf{x}^{(k)} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Omega}^{(k)} & \boldsymbol{\Delta}^{(k)} \\ -\boldsymbol{\Delta}^{(k)} & -\boldsymbol{\Omega}^{(k)} \end{pmatrix} \tag{29}$$

subject to the generalized symplectic orthonormality condition

$$\begin{pmatrix} \mathbf{x}^{(k)} & \mathbf{y}^{(k)} \\ \mathbf{y}^{(k)} & \mathbf{x}^{(k)} \end{pmatrix}^T \begin{pmatrix} \boldsymbol{\sigma}^{(k)} & \boldsymbol{\pi}^{(k)} \\ -\boldsymbol{\pi}^{(k)} & -\boldsymbol{\sigma}^{(k)} \end{pmatrix} \begin{pmatrix} \mathbf{x}^{(k)} & \mathbf{y}^{(k)} \\ \mathbf{y}^{(k)} & \mathbf{x}^{(k)} \end{pmatrix} = \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & -\mathbf{1} \end{pmatrix}. \tag{30}$$

Subsection 3 of the Appendix explains how this problem may be reduced to a standard symmetric eigenvalue problem by a lower upper (LU) and a Cholesky decomposition. As before, the lowest p eigenvectors minimize F on $\mathcal{K}^{(k)}$ and are thus variational best approximations.

The residual is now

$$\begin{pmatrix} \mathbf{R}^{(k)} & \mathbf{S}^{(k)} \\ \mathbf{S}^{(k)} & \mathbf{R}^{(k)} \end{pmatrix} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B} & \mathbf{A} \end{pmatrix} \begin{pmatrix} \mathbf{X}^{(k)} & \mathbf{Y}^{(k)} \\ \mathbf{Y}^{(k)} & \mathbf{X}^{(k)} \end{pmatrix} - \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & -\mathbf{1} \end{pmatrix} \begin{pmatrix} \mathbf{X}^{(k)} & \mathbf{Y}^{(k)} \\ \mathbf{Y}^{(k)} & \mathbf{X}^{(k)} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Omega}^{(k)} & \boldsymbol{\Delta}^{(k)} \\ -\boldsymbol{\Delta}^{(k)} & -\boldsymbol{\Omega}^{(k)} \end{pmatrix}, \tag{31}$$

where $\mathbf{R}^{(k)}$ and $\mathbf{S}^{(k)}$ are $n \times p$. Convergence may be tested by computing the Euclidean norm of individual residual vectors, or by computing the Frobenius norm of the residual matrix. After a possible rank reduction, the preconditioned residual matrix is obtained by solving the linear equations[10]

$$\begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{R}}^{(k)} & \tilde{\mathbf{S}}^{(k)} \\ \tilde{\mathbf{S}}^{(k)} & \tilde{\mathbf{R}}^{(k)} \end{pmatrix} - \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & -\mathbf{1} \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{R}}^{(k)} & \tilde{\mathbf{S}}^{(k)} \\ \tilde{\mathbf{S}}^{(k)} & \tilde{\mathbf{R}}^{(k)} \end{pmatrix} \times \begin{pmatrix} \boldsymbol{\Omega}^{(k)} & \boldsymbol{\Delta}^{(k)} \\ -\boldsymbol{\Delta}^{(k)} & -\boldsymbol{\Omega}^{(k)} \end{pmatrix} = \begin{pmatrix} \mathbf{R}^{(k)} & \mathbf{S}^{(k)} \\ \mathbf{S}^{(k)} & \mathbf{R}^{(k)} \end{pmatrix}. \tag{32}$$

Finally, the subspace basis is augmented by the preconditioned residual according to

$$\mathbf{V}^{(k+1)} = \begin{bmatrix} \mathbf{V}^{(k)} & \tilde{\mathbf{R}}^{(k)} \end{bmatrix}, \quad \mathbf{W}^{(k+1)} = \begin{bmatrix} \mathbf{W}^{(k)} & \tilde{\mathbf{S}}^{(k)} \end{bmatrix}. \tag{33}$$

E. Symplectic linear problems

Non-resonant frequency-dependent response properties in the TDDFT and TDHF framework are solutions of a symplectic linear problem,[44,49]

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B} & \mathbf{A} \end{pmatrix} \begin{pmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Y} & \mathbf{X} \end{pmatrix} - \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & -\mathbf{1} \end{pmatrix} \begin{pmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Y} & \mathbf{X} \end{pmatrix} \begin{pmatrix} \boldsymbol{\omega} & \mathbf{0} \\ \mathbf{0} & -\boldsymbol{\omega} \end{pmatrix} = \begin{pmatrix} \mathbf{P} & \mathbf{Q} \\ \mathbf{Q} & \mathbf{P} \end{pmatrix}. \tag{34}$$

As before, $\mathbf{A}$ and $\mathbf{B}$ are real symmetric $n \times n$ matrices, $\mathbf{X}$ and $\mathbf{Y}$ are $n \times p$, and $\boldsymbol{\omega} = \mathrm{diag}(\omega_1, \ldots, \omega_p)$ contains the frequencies of each perturbation represented by the $n \times p$ RHS matrices $\mathbf{P}$ and $\mathbf{Q}$. Equation (34) may be transformed into a standard Sylvester equation by left-multiplication with $\left(\begin{smallmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & -\mathbf{1} \end{smallmatrix}\right)$.

Instead of attempting to solve Eq. (34) directly, we consider the problem of making the functional

$$F\left[\begin{pmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Y} & \mathbf{X} \end{pmatrix} \middle| \begin{pmatrix} \boldsymbol{\omega} & \boldsymbol{\delta} \\ \boldsymbol{\delta} & \boldsymbol{\omega} \end{pmatrix}, \begin{pmatrix} \mathbf{P} & \mathbf{Q} \\ \mathbf{Q} & \mathbf{P} \end{pmatrix}\right] = \left\langle \begin{pmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Y} & \mathbf{X} \end{pmatrix}^T \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B} & \mathbf{A} \end{pmatrix} \begin{pmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Y} & \mathbf{X} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\omega} & \boldsymbol{\delta} \\ -\boldsymbol{\delta} & -\boldsymbol{\omega} \end{pmatrix} \begin{pmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Y} & \mathbf{X} \end{pmatrix}^T \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & -\mathbf{1} \end{pmatrix} \begin{pmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Y} & \mathbf{X} \end{pmatrix} - \begin{pmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Y} & \mathbf{X} \end{pmatrix}^T \begin{pmatrix} \mathbf{P} & \mathbf{Q} \\ \mathbf{Q} & \mathbf{P} \end{pmatrix} - \begin{pmatrix} \mathbf{P} & \mathbf{Q} \\ \mathbf{Q} & \mathbf{P} \end{pmatrix}^T \begin{pmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Y} & \mathbf{X} \end{pmatrix} \right\rangle \tag{35}$$

stationary. The symmetric $p \times p$ matrices $\boldsymbol{\omega}$ and $\boldsymbol{\delta}$ are not necessarily diagonal. F is invariant under unitary transformations (24),

$$F\left[\begin{pmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Y} & \mathbf{X} \end{pmatrix} \middle| \begin{pmatrix} \boldsymbol{\omega} & \boldsymbol{\delta} \\ \boldsymbol{\delta} & \boldsymbol{\omega} \end{pmatrix}, \begin{pmatrix} \mathbf{P} & \mathbf{Q} \\ \mathbf{Q} & \mathbf{P} \end{pmatrix}\right] = F\left[\begin{pmatrix} \mathbf{X} & \mathbf{Y} \\ \mathbf{Y} & \mathbf{X} \end{pmatrix} \begin{pmatrix} \mathbf{U} & \mathbf{T} \\ \mathbf{T} & \mathbf{U} \end{pmatrix} \middle| \begin{pmatrix} \mathbf{U} & \mathbf{T} \\ \mathbf{T} & \mathbf{U} \end{pmatrix}^T \begin{pmatrix} \boldsymbol{\omega} & \boldsymbol{\delta} \\ \boldsymbol{\delta} & \boldsymbol{\omega} \end{pmatrix} \begin{pmatrix} \mathbf{U} & \mathbf{T} \\ \mathbf{T} & \mathbf{U} \end{pmatrix}, \begin{pmatrix} \mathbf{P} & \mathbf{Q} \\ \mathbf{Q} & \mathbf{P} \end{pmatrix} \begin{pmatrix} \mathbf{U} & \mathbf{T} \\ \mathbf{T} & \mathbf{U} \end{pmatrix}\right]. \tag{36}$$

The approximate solution in the kth iteration is expanded into the $4q_k$-dimensional subspace $\mathcal{K}^{(k)}$ according to Eq. (26). The subspace overlap matrices $\boldsymbol{\sigma}^{(k)}$ and $\boldsymbol{\pi}^{(k)}$ and the Rayleigh matrices $\mathbf{a}^{(k)}$ and $\mathbf{b}^{(k)}$ are obtained from Eqs. (27) and (28), respectively. Similarly, the projections of the RHS matrices onto $\mathcal{K}^{(k)}$ are

$$\begin{pmatrix} \mathbf{p}^{(k)} & \mathbf{q}^{(k)} \\ \mathbf{q}^{(k)} & \mathbf{p}^{(k)} \end{pmatrix} = \begin{pmatrix} \mathbf{V}^{(k)} & \mathbf{W}^{(k)} \\ \mathbf{W}^{(k)} & \mathbf{V}^{(k)} \end{pmatrix}^T \begin{pmatrix} \mathbf{P} & \mathbf{Q} \\ \mathbf{Q} & \mathbf{P} \end{pmatrix}. \tag{37}$$

In each iteration, the best approximations to the solution on $\mathcal{K}^{(k)}$ are computed by solving the reduced symplectic linear problem

$$\begin{pmatrix} \mathbf{a}^{(k)} & \mathbf{b}^{(k)} \\ \mathbf{b}^{(k)} & \mathbf{a}^{(k)} \end{pmatrix} \begin{pmatrix} \mathbf{x}^{(k)} & \mathbf{y}^{(k)} \\ \mathbf{y}^{(k)} & \mathbf{x}^{(k)} \end{pmatrix} - \begin{pmatrix} \boldsymbol{\sigma}^{(k)} & \boldsymbol{\pi}^{(k)} \\ -\boldsymbol{\pi}^{(k)} & -\boldsymbol{\sigma}^{(k)} \end{pmatrix} \begin{pmatrix} \mathbf{x}^{(k)} & \mathbf{y}^{(k)} \\ \mathbf{y}^{(k)} & \mathbf{x}^{(k)} \end{pmatrix} \begin{pmatrix} \boldsymbol{\omega} & \boldsymbol{\delta} \\ -\boldsymbol{\delta} & -\boldsymbol{\omega} \end{pmatrix} = \begin{pmatrix} \mathbf{p}^{(k)} & \mathbf{q}^{(k)} \\ \mathbf{q}^{(k)} & \mathbf{p}^{(k)} \end{pmatrix}. \tag{38}$$

Subsection 4 of the Appendix shows how this equation may be reduced to a symmetric linear problem by an LU and a Cholesky decomposition.

The residual matrix is now

$$\begin{pmatrix} \mathbf{R}^{(k)} & \mathbf{S}^{(k)} \\ \mathbf{S}^{(k)} & \mathbf{R}^{(k)} \end{pmatrix} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B} & \mathbf{A} \end{pmatrix} \begin{pmatrix} \mathbf{X}^{(k)} & \mathbf{Y}^{(k)} \\ \mathbf{Y}^{(k)} & \mathbf{X}^{(k)} \end{pmatrix} - \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & -\mathbf{1} \end{pmatrix} \times \begin{pmatrix} \mathbf{X}^{(k)} & \mathbf{Y}^{(k)} \\ \mathbf{Y}^{(k)} & \mathbf{X}^{(k)} \end{pmatrix} \begin{pmatrix} \omega & \delta \\ -\delta & -\omega \end{pmatrix} - \begin{pmatrix} \mathbf{P} & \mathbf{Q} \\ \mathbf{Q} & \mathbf{P} \end{pmatrix}. \tag{39}$$

The convergence may be tested as described in Sec. II D. Finally, the preconditioned residuals are determined by solving

$$\begin{pmatrix} \mathbf{D} & \mathbf{0} \\ \mathbf{0} & \mathbf{D} \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{R}}^{(k)} & \tilde{\mathbf{S}}^{(k)} \\ \tilde{\mathbf{S}}^{(k)} & \tilde{\mathbf{R}}^{(k)} \end{pmatrix} - \begin{pmatrix} \mathbf{1} & \mathbf{0} \\ \mathbf{0} & -\mathbf{1} \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{R}}^{(k)} & \tilde{\mathbf{S}}^{(k)} \\ \tilde{\mathbf{S}}^{(k)} & \tilde{\mathbf{R}}^{(k)} \end{pmatrix} \begin{pmatrix} \omega & \delta \\ -\delta & -\omega \end{pmatrix} = \begin{pmatrix} \mathbf{R}^{(k)} & \mathbf{S}^{(k)} \\ \mathbf{S}^{(k)} & \mathbf{R}^{(k)} \end{pmatrix}. \tag{40}$$

F. Error analysis

For theoretical error analysis, we only consider errors in the matrix-vector products, since other errors such as roundoff are orders of magnitude smaller in typical applications. Rather than attempting an exhaustive analysis, we consider a model system to illustrate key characteristics of the nKs algorithm. This model consists of real symmetric eigenvalue problem (1) for a single eigenvector and without preconditioning.

We start with an approximate eigenvector $\mathbf{X}$ normalized to 1. $\mathbf{X}$ does not need to be a unit vector and could be the result of previous iterations. In integral-direct algorithms using screening, the matrix vector products $\mathbf{AX}$ are computed approximately, with a predefined tolerance δ,

$$[\mathbf{AX}]_\delta = \mathbf{AX} + \delta\mathbf{e}. \tag{41}$$

Here, $\mathbf{e}$ is a unit vector whose direction is determined by the details of the system at hand and the integral screening algorithm. According to Eq. (41), the matrix-vector product computed with a finite tolerance δ and all subsequent quantities become functions of δ, which is indicated by subscript.

The approximate eigenvalue belonging to $\mathbf{X}$ is

$$a_{11\delta} = \mathbf{X}^T[\mathbf{AX}]_\delta = \mathbf{X}^T\mathbf{AX} + \delta\mathbf{X}^T\mathbf{e} = a_{11} + \delta\mathbf{X}^T\mathbf{e}. \tag{42}$$

Thus, a finite tolerance produces an error linear in δ in the approximate eigenvalue $a_{11\delta}$.

By Eqs. (41) and (42), the first residual is

$$\mathbf{R}_\delta = [\mathbf{AX}]_\delta - a_{11\delta}\mathbf{X} = \mathbf{AX} + \delta\mathbf{e} - a_{11\delta}\mathbf{X} = \mathbf{R} + \delta\mathbf{e}_\perp, \tag{43}$$

where the vector

$$\mathbf{e}_\perp = (\mathbf{1} - \mathbf{X}\mathbf{X}^T)\mathbf{e} \tag{44}$$

has been introduced for notational convenience. Since $\|\mathbf{e}_\perp\| \le 1$, the error in $\mathbf{R}_\delta$ is also $O(\delta)$. Even for finite δ, $\mathbf{R}_\delta$ is orthogonal to $\mathbf{X}$,

$$\mathbf{X}^T\mathbf{R}_\delta = 0, \tag{45}$$

since $\mathbf{X}^T\mathbf{e}_\perp = 0$. As a result,

$$a_{12\delta} = [\mathbf{X}^T\mathbf{A}]_\delta\mathbf{R}_\delta = \mathbf{R}_\delta^T\mathbf{R}_\delta = \mathbf{R}^T\mathbf{R} + 2\delta\mathbf{R}^T\mathbf{e} + O(\delta^2). \tag{46}$$

The Euclidean residual norm is thus

$$r_\delta = \sqrt{\mathbf{R}_\delta^T\mathbf{R}_\delta} = r + \delta\frac{\mathbf{R}^T\mathbf{e}}{r} + O(\delta^2). \tag{47}$$

Since $\|\mathbf{R}^T\mathbf{e}/r\|$ is $O(1)$, the error in r_δ is linear in δ, independent of the magnitude of r. Equation (47) suggests that δ should be chosen well below the desired convergence criterion for the residual norm.

The only missing matrix element of the Rayleigh matrix is

$$a_{22\delta} = \mathbf{R}_\delta^T[\mathbf{AR}_\delta]_\delta = \mathbf{R}^T\mathbf{AR} + \delta\mathbf{R}^T\mathbf{A}(2\mathbf{e}_\perp + \mathbf{e}') + O(\delta^2), \tag{48}$$

where $\mathbf{e}'$ is another unit vector. By Eqs. (6) and (7), the matrices $\mathbf{a}$ and $\mathbf{s}$ are now functions of δ,

$$\mathbf{a}_\delta = \begin{pmatrix} a_{11\delta} & r_\delta^2 \\ r_\delta^2 & a_{22\delta} \end{pmatrix}, \quad \mathbf{s}_\delta = \begin{pmatrix} 1 & 0 \\ 0 & r_\delta^2 \end{pmatrix}. \tag{49}$$

Since $\mathbf{s}_\delta$ is diagonal, the generalized symmetric subspace eigenvalue problem is easily transformed to a standard symmetric eigenvalue problem for matrix

$$\tilde{\mathbf{a}}_\delta = [\mathbf{s}_\delta]^{-1/2}\mathbf{a}_\delta[\mathbf{s}_\delta]^{-1/2} = \begin{pmatrix} a_{11\delta} & r_\delta \\ r_\delta & \dfrac{a_{22\delta}}{r_\delta^2} \end{pmatrix}. \tag{50}$$

Taylor expansion of $\tilde{\mathbf{a}}_\delta$ around $\delta = 0$ leads to

$$\tilde{\mathbf{a}}_\delta = \tilde{\mathbf{a}} + \delta\tilde{\mathbf{a}}^{(1)} + O(\delta^2), \tag{51}$$

where

$$\tilde{\mathbf{a}} = \begin{pmatrix} a_{11} & r \\ r & \dfrac{\mathbf{R}^T\mathbf{AR}}{r^2} \end{pmatrix}, \qquad \tilde{\mathbf{a}}^{(1)} = \begin{pmatrix} \mathbf{X}^T\mathbf{e} & \dfrac{\mathbf{R}^T\mathbf{e}}{r} \\ \dfrac{\mathbf{R}^T\mathbf{e}}{r} & \dfrac{\mathbf{R}^T\mathbf{A}(2\mathbf{e}_\perp + \mathbf{e}')}{r^2} - 2\dfrac{\mathbf{R}^T\mathbf{AR}}{r^4}\mathbf{R}^T\mathbf{e} \end{pmatrix}. \tag{52}$$

Depending on the magnitude of the residual norm r, two cases may be distinguished: (i) Early in the iteration, $r \simeq 1$, and thus $\|\tilde{\mathbf{a}}^{(1)}\|$ is comparable to $\|\tilde{\mathbf{a}}\|$. In this case, $\delta\tilde{\mathbf{a}}^{(1)}$ is a small perturbation producing errors $O(\delta)$ in the eigenvectors and eigenvalues of $\|\tilde{\mathbf{a}}\|$. Since other errors are assumed to be much smaller than δ, the method becomes numerically equivalent to the conventional Davidson algorithm, which is stable.[45] (ii) Close to convergence, r may become small enough such that $r \simeq \delta \ll 1$, especially if the threshold δ is chosen close to the residual norm required for convergence. In this case, the condition number of $\mathbf{s}_\delta$ grows as $1/r^2$, and $\|\tilde{\mathbf{a}}^{(1)}\|$ grows as $1/r$, so $\delta\tilde{\mathbf{a}}^{(1)}$ is no longer small. However, since $r_\delta \simeq r$ according to Eq. (47), the off-diagonal element $a_{12\delta}$ is now small, leading to the expansion

$$\Omega_\delta = a_{11\delta} + \delta\mathbf{X}^T\mathbf{e} + \frac{r_\delta^2}{\tilde{a}_{\delta 11} - \tilde{a}_{\delta 22}} + O(r_\delta^4) \tag{53}$$

for the lowest eigenvalue of $\tilde{\mathbf{a}}$. The corresponding Ritz vector is

$$\mathbf{X}_\delta = \mathbf{X} + \frac{1}{\tilde{a}_{\delta 11} - \tilde{a}_{\delta 22}}\mathbf{R}_\delta + O(r_\delta^2). \tag{54}$$

Thus, the leading error in the first eigenpair remains $O(r_\delta) \sim O(\delta)$, even as the residual becomes small.

The second eigenpair of the Rayleigh matrix is discarded, because it does not minimize the functional F in Sec. II B. It is nevertheless instructive to compare the effect of finite δ on the second eigenpair to that on the first one: The error in the second eigenpair is $O(\delta)$ in case (i), but $O(1)$ in case (ii). This illustrates a key characteristic of the nKs method compared to the conventional algorithm: As the residual norm decreases, the weight of the corrections to the eigenvectors decreases and thus the increasing relative errors in the matrix-vector products $\mathbf{AX}$ do not show up in the desired lowest p eigenpairs. Instead, the errors accumulate in the orthogonal complement that will be discarded after the iteration. The conventional algorithm computes all $q_k \geq p$ eigenpairs in each iteration, even the ones that are not of interest, with the same precision, which requires constant relative error in the matrix-vector products.

A numerical experiment comparing the stability of the conventional and the nKs algorithm is shown in Table I. To test the limits of the nKs algorithm, the convergence threshold for the maximum Euclidean residual norm was chosen to be 1×10^{-10}. This value is much smaller than the typical convergence criterion for applications of 1×10^{-5}. The nKs algorithm converges monotonically up to the 18th iteration. The two methods agree to within 14 figures which is close to machine precision. The deviations in the last iteration are 10^{-15} H which is comparable to the integral neglect threshold 10^{-16} used here, in agreement with the result that the eigenvalue error is linear in δ.

G. Optimality

In the ground state difference density algorithm, linear combinations of density matrices are commonly chosen to minimize the Frobenius norm of the difference density in each iteration.[13] This is equivalent to orthogonalizing the difference densities with respect to the Frobenius inner product. While this choice is sensible from the viewpoint of screening, the norm of the thus determined difference density does not decrease monotonically in conventional self-consistent field algorithms which minimize the energy. This reduces efficiency and can result in error accumulation.[13]

For the nKs methods proposed here, the unconditioned residuals of the kth iterates are orthogonal in the Frobenius inner product, since their projection on $\mathcal{K}^{(k)}$ vanishes by definition. In this sense, they are already optimal and additional orthogonalization is unnecessary. Preconditioning can lead to a slight departure from orthogonality, but the effect on the efficiency of screening appears to be small.

An additional advantage of the nKs approach for screening and numerical stability is the monotonic decrease of the residual norm for linear systems (14) and (34). This follows because residuals (19) and (39) are linear in the Ritz vector, and thus the squared Frobenius residual norm is a quadratic positive semidefinite form which is minimized by the respective Ritz vectors. For eigenvalue problems (1) and (21), the residuals (10) and (31) are non-linear in the Ritz vector due to the dependence on the approximate eigenvalues. This is consistent with the observation that the residual norm sometimes increases in the first iterations for these eigenvalue problems. Nevertheless, as convergence is reached, the eigenvalues converge rapidly, and thus the residual norm becomes quadratic and decreases monotonically.

H. Relation to other Krylov space methods

For $p = 1$ and in the absence of preconditioning, the subspace matrices $\mathbf{s}^{(k)}$ and $\mathbf{a}^{(k)}$ in Secs. II B and II C become diagonal and tridiagonal, respectively. In this case, the approximate eigenpairs or solution vectors can be constructed recursively using the Lanczos method.[59] The conjugate gradient method[60] and the modified Davidson algorithm of van Lenthe and Pulay[61] are efficient and stable implementations of Lanczos' idea and use a fixed preconditioner that does not change from iteration to iteration. The space-saving deflection algorithm by Chernyak and co-workers[62] also falls in this category.

An important distinction between the Davidson and Lanczos-type schemes is the explicit construction of the Krylov space basis in the former, while orthonormality of the Krylov space basis is implicit in Lanczos-type recursion. Thus, Lanczos-type methods require less vector storage, but are susceptible to loss of subspace orthonormality due to numerical imprecision in matrix-vector products. This well-known limitation of Lanczos-type iteration can lead to divergence and spurious eigenvalues and numerous approaches exist which detect a breakdown of orthonormality and "restart" the iteration[63,64] or selectively orthonormalize.[65] The Davidson method, on the other hand, does not rely on recursion and completely orthonormalizes the Krylov subspace basis in each iteration using MGS. This requires additional operations and storage compared to Lanczos-type methods, but substantially improves numerical stability. The nKs algorithm is a generalization of the Davidson method that replaces the MGS orthonormalization with a Cholesky decomposition of the subspace metric. Thus, although nKs operates with a nonorthonormal subspace basis, it does not exhibit the instability of Lanczos recursion.

Lanczos-type iterations are most useful for extremely large banded coefficient matrices with small bandwidth, because the number of vectors that need to be stored is small and does not increase with the number of iterations, and the number of vector-vector operations is small compared to the present methods. However, the matrices occurring in electronic structure calculations using atom-centered basis sets are typically much less sparse and less structured. In such applications, the computation of matrix-vector products is often rate-determining and several orders of magnitude more expensive than vector-vector operations, whereas vector storage during the iterations is not critical, with the exception of very large configuration interaction calculations. Davidson-type methods, including the nKs variant, minimize the number of matrix-vector products, because they can take advantage of

TABLE I. Convergence of the conventional and the nKs algorithm for the 1^1B_2 excitation energy of methylene blue cation ($C_{16}H_{18}N_3S^+$). k is the iteration number, Ω denotes the excitation energy in mH, and N_{int} is the number of integral batches computed in each iteration. To test the stability of the algorithm, a very tight convergence criterion for the maximum residual norm of 1×10^{-10} was chosen. The PBE0 hybrid functional,[55] def2-SVPD[56,57] basis sets ($n = 10\,734$), and very fine grids of size 5[58] were employed. The machine precision was 2×10^{-16}.

	Conventional		nKs	
k	Ω	N_{int}	Ω	N_{int}
1	118.125 831 402 822 8	63 784 001	118.125 831 402 822 8	63 784 001
2	94.136 286 336 343 47	89 560 923	94.136 286 336 342 88	88 470 417
3	92.834 913 540 604 32	90 213 163	92.834 913 540 603 94	86 190 854
4	92.646 546 761 230 99	89 742 681	92.646 546 761 230 39	84 179 040
5	92.619 616 461 941 21	90 496 883	92.619 616 461 940 60	82 628 501
6	92.613 454 499 008 19	90 312 932	92.613 454 499 007 29	81 382 610
7	92.612 846 774 049 25	90 476 762	92.612 846 774 049 55	79 206 221
8	92.612 756 580 345 40	90 923 424	92.612 756 580 344 19	77 603 072
9	92.612 749 214 326 90	90 752 643	92.612 749 214 326 59	75 112 115
10	92.612 748 023 698 09	90 721 179	92.612 748 023 698 69	72 403 118
11	92.612 747 920 457 97	91 343 895	92.612 747 920 457 37	70 563 627
12	92.612 747 910 978 01	90 677 578	92.612 747 910 976 22	67 721 512
13	92.612 747 909 094 73	91 175 482	92.612 747 909 094 13	65 712 534
14	92.612 747 908 989 24	91 382 113	92.612 747 908 987 44	62 184 742
15	92.612 747 908 979 05	90 918 893	92.612 747 908 980 84	58 627 779
16	92.612 747 908 982 65	91 322 006	92.612 747 908 980 25	55 967 402
17	92.612 747 908 985 05	91 457 651	92.612 747 908 977 85	51 759 535
18	92.612 747 908 980 25	91 231 997	92.612 747 908 980 25	47 962 378
19	92.612 747 908 977 85	91 311 227	92.612 747 908 977 85	44 447 439
20	92.612 747 908 977 85	91 570 576	92.612 747 908 975 45	40 384 598
Sum		1 789 376 009		1 356 291 495

variable preconditioners and construct a single subspace for all iterated eigenpairs or solution vectors with different RHSs or frequencies. It may be possible to reformulate Lanczos-type methods to take advantage of the decreasing residual norm in each iteration. However, there may be little benefit for large banded matrices, whereas for less sparse systems the present methods are preferable.

For a single eigenpair, the nKs algorithm recovers the Davidson algorithm[6,45] for symmetric eigenvalue problems and the algorithm by Olsen, Jensen, and Jørgensen[10] for symplectic eigenvalue problems if orthonormality of the subspace basis is enforced. Further analysis of the relation between the various methods discussed here may be found in Ref. 66.

I. Best practices

Iterative methods are very powerful, but their success depends considerably on details of their implementation. For example, Lanczos iteration was considered unstable for two decades until Paige presented a stable implementation.[67] The practices described below are essential for fast and reliable convergence of nKs methods.

1. Start vectors

For eigenvalue problems, p start vectors are chosen as unit vectors corresponding to the lowest eigenvalues of the diagonal matrix $\mathbf{D}$. Additional unit vectors corresponding to higher eigenvectors of $\mathbf{D}$ should be added if the eigenvalue difference is small. For linear problems, the p start vectors are obtained by applying a single preconditioning step to the RHS vectors.

2. Rayleigh matrices

For finite δ, the Rayleigh matrices $\mathbf{a}^{(k)}$ and $\mathbf{b}^{(k)}$ are slightly unsymmetric due to errors in the matrix vector products. If $\mathbf{a}^{(k)} = \mathbf{V}^{(k)T}[\mathbf{A}\mathbf{V}^{(k)}]$, then the lower triangle has higher precision than the upper one and should be used instead of symmetrization; analogous recommendations apply to $\mathbf{b}^{(k)}$.

3. Subspace metric

With decreasing residual norm, the subspace overlap matrix $\mathbf{s}^{(k)}$ quickly becomes ill-conditioned. However, the condition of $\mathbf{s}^{(k)}$ can be dramatically improved by a simple diagonal preconditioner, as described in Subsections 1-4 of the Appendix. Without preconditioning, the eigenpairs and residuals may exhibit large errors, and the iteration may become unstable.

4. Restart

If the condition of the subspace metric is poor despite preconditioning, the iteration may be restarted using the previously computed approximate eigenpairs or solution vectors as initial basis. Such restarts can be beneficial for

Lanczos-type iterations with extremely large banded matrices, where the cost for a single matrix-vector product is relatively low. For electronic structure calculations where the matrix-vector products are rate-determining, restarts should be used very sparingly. We find that restarts are hardly ever necessary when good preconditioners are used.

5. Convergence

Convergence should be tested using the residual norm, which provides *a posteriori* error bounds,[5] rather than or in addition to changes in eigenvalues or other properties. The tolerance δ should be chosen at least 3 orders of magnitude below the maximum residual norm desired at convergence.

III. APPLICATION TO HYBRID TDDFT

A. Computation of matrix-vector products

In hybrid time-dependent density functional theory, a symplectic eigenvalue problem of the form (21) determines the electronic excitation energies and transition densities of a many-electron ground state, whereas linear response properties such as frequency-dependent polarizabilities are obtained by solving the symplectic linear problem (34).[16,17,44] Within the spin-unrestricted approximation, the underlying vector space is spanned by all direct products of occupied and virtual static KS molecular orbitals (MOs) of identical spin σ ("particle-hole space"), and its dimension scales with the square of the system size. We assume that the spin unrestricted KS MOs are given by real expansion coefficients $C_{\mu p\sigma}$ with canonical orbital energies $\epsilon_{p\sigma}$. Indices $j,k,\ldots$ label occupied, $a,b,\ldots$ virtual, and $p,q,\ldots$ general KS MOs, and Greek indices label atomic orbitals (AOs).

The matrix vector products $\mathbf{X} = \mathbf{AV}$ and $\mathbf{Y} = \mathbf{BW}$ are computed by first evaluating[19]

$$\mathbf{H}^{+}_{\mu\nu\sigma} = \sum_{\kappa\lambda\sigma'} \Big\{ 2(\mu\nu|\kappa\lambda) + 2f^{\mathrm{XC}+}_{\mu\nu\sigma\kappa\lambda\sigma'} - \delta_{\sigma\sigma'} c^{\mathrm{X}} \times [(\mu\kappa|\nu\lambda) + (\mu\lambda|\nu\kappa)] \Big\} (\mathbf{V}+\mathbf{W})_{\kappa\lambda\sigma'}, \quad (55a)$$

$$\mathbf{H}^{-}_{\mu\nu\sigma} = \sum_{\kappa\lambda\sigma'} \Big\{ 2f^{\mathrm{XC}-}_{\mu\nu\sigma\kappa\lambda\sigma'} + \delta_{\sigma\sigma'} c^{\mathrm{X}} [(\mu\kappa|\nu\lambda) - (\mu\lambda|\nu\kappa)] \Big\} (\mathbf{V}-\mathbf{W})_{\kappa\lambda\sigma'}. \quad (55b)$$

Here, $(\mu\nu|\kappa\lambda)$ is a two-electron repulsion integral (ERI) in Mulliken notation, $f^{\mathrm{XC}\pm}$ are the symmetric and antisymmetric exchange-correlation (XC) kernels within the adiabatic approximation,[68] respectively, and c^{X} is Becke's global hybrid exchange scaling parameter.[69] The TDHF method corresponds to $c^{\mathrm{X}} = 1$ and zero XC kernels. The TDA or CIS theory is recovered for $\mathbf{Y} \equiv \mathbf{0}$. The bold quantities in Eq. (55) are $q^{(k)}$-dimensional vectors in the Krylov space $\mathcal{K}^{(k)}$, but we do not explicitly resolve this additional dimension for notational clarity. The AO-transformed basis vectors

$$(\mathbf{V}\pm\mathbf{W})_{\mu\nu\sigma} = \sum_{ia} C_{\mu i\sigma} C_{\nu a\sigma} (\mathbf{V}\pm\mathbf{W})_{ia\sigma} \quad (56)$$

are symmetric and antisymmetric under exchange of μ and ν, respectively,[11,19,66] and take the place of the real and imaginary parts of the density matrix in the ground state formalism. The final matrix-vector products are obtained by transforming back to the MO basis and adding the diagonal part of $\mathbf{A} \pm \mathbf{B}$,

$$(\mathbf{X}\pm\mathbf{Y})_{ia\sigma} = \sum_{\mu\nu} C_{\mu i\sigma} C_{\nu a\sigma} \mathbf{H}^{\pm}_{\mu\nu\sigma} + (\epsilon_{a\sigma} - \epsilon_{i\sigma})(\mathbf{V}\pm\mathbf{W})_{ia\sigma}. \quad (57)$$

The above method avoids explicit calculation of the elements of $\mathbf{A}$ and $\mathbf{B}$ in the MO basis, which is prohibitive for larger systems, since these matrices are not sparse. The required matrices of the ERIs and the XC kernels in the AO basis are sparse due to locality of the AO basis and the short range of semi-local XC kernels. The observation that $\mathbf{H}^{\pm}$ equals the real and imaginary parts of the first-order Fock matrix, respectively,

$$\mathbf{H}^{\pm} = \frac{d}{d\lambda} \mathbf{F}^{\pm}[\mathbf{D}^{(0)} + \lambda \mathbf{D}^{(1)}]\Big|_{\lambda=0}, \quad (58)$$

where $\mathbf{D}^{(0)}$ is the ground state density matrix and that, for $q_k = 1$,

$$\mathbf{D}^{(1)} = \begin{pmatrix} \mathbf{0} & \mathbf{V} \\ \mathbf{W}^T & \mathbf{0} \end{pmatrix} \quad (59)$$

can be identified with the first-order density matrix expressed in the MO basis, reveals the intimate connection between the matrix-vector products and Fock matrix construction.[11] This is fundamental for efficient molecular property calculations and follows from the derivative formulation of response theory.[3,53] Thus, efficient prescreening methods for ground-state Fock matrix construction[13,58] carry over directly to Eq. (55): Integrals $(\mu\nu|\kappa\lambda)$ are skipped if

$$Q_{\mu\nu} Q_{\kappa\lambda} \max_{\sigma} \{ |(\mathbf{V}+\mathbf{W})_{\kappa\lambda\sigma}|, |(\mathbf{V}+\mathbf{W})_{\mu\nu\sigma}| \} \le \delta_{\mathrm{coul}} \quad (60)$$

for Coulomb contractions and

$$Q_{\mu\nu} Q_{\kappa\lambda} \max_{\sigma} \{ |(\mathbf{V}\pm\mathbf{W})_{\mu\kappa\sigma}|, |(\mathbf{V}\pm\mathbf{W})_{\nu\lambda\sigma}|, |(\mathbf{V}\pm\mathbf{W})_{\mu\lambda\sigma}|, |(\mathbf{V}\pm\mathbf{W})_{\nu\kappa\sigma}|, \} \le \delta_{\mathrm{ex}} \quad (61)$$

for exchange contractions. $|(\mu\nu|\kappa\lambda)| \le Q_{\mu\nu} Q_{\kappa\lambda}$ by the Cauchy-Schwarz inequality,[13] where $Q_{\mu\nu} = \sqrt{(\mu\nu|\mu\nu)}$ is pre-computed; similar estimates are used to estimate contributions to the XC kernel-vector products.[21] By taking the maximum of the estimates over shell pairs, entire shell quadruples (batches) of integrals may be neglected at once. The thresholds δ_{coul} and δ_{ex} are chosen such that the maximum error in each matrix-vector product equals a pre-defined constant independent of the system size.

A key feature of the nKs method is the decrease of the norm of the basis vectors $\mathbf{V}$ and $\mathbf{W}$ with the residual norm. Thus, nKs methods have a distinct advantage when the matrix-vector products are computed in an integral direct fashion, because fewer and fewer integrals need to be calculated as the iteration progresses. Conventional iterative subspace methods, on the other hand, use orthonormal bases, and thus there is no efficiency gain as convergence is reached, even though the corrections to the eigenpairs or solution vectors become smaller and smaller. This is illustrated for the first singlet excitation of methylene blue in Table I: In the first iteration, the number of integral batches computed is

identical for both methods and fairly small, because the initial vector space includes only (localized) low-energy states. After the first iteration, the number of integral batches remains approximately constant in the conventional case, while it drops at a nearly constant rate for the nKs algorithm. Even for this small system and with an exceedingly tight convergence criterion, the total number of integral batches is reduced by 24% overall in the nKs algorithm, without any appreciable loss of precision.

The prefactor of the Coulomb contractions in Eq. (55) can be substantially reduced by a factor of 30–50 using RI-J methods.[21,22] In the conventional method, this affords substantial speed-ups for non-hybrid functionals, but little improvement for hybrid functionals, where four-center ERIs are still necessary for the hybrid exchange contractions. However, as demonstrated below, the nKs method can be synergistically combined with tighter bounds and optimized loop structures for integral-direct exchange contractions developed for ground states,[14,34,35,70] since these methods benefit disproportionately from smaller density matrix norms.

B. Performance

Table II compares the performance of the nKs algorithm with and without the RI-J approximation to the conventional Davidson method using Turbomole[82] revision number 19015. The selection of compounds covers organic chromophores, two-photon dyes, organometallic complexes, a 282 atom cadmium selenide cluster, and a 296 atom polythiophene dendrimer.[83] Static and dynamic polarizability calculations as well as excitation energy calculations for $p = 1$ to 100 lowest states were performed. Methods include TDHF and hybrid TDDFT using the PBE0[55] functional as well as Becke's half

TABLE II. CPU timings for response calculations using the conventional algorithm (Conv) and relative speed-ups of the nKs method without and with the RI-J approximation. N_{at} is the number of atoms; if point group symmetry higher than C_1 was used, the Schoenflies symbol is given in parentheses. p denotes the number of lowest eigenpairs computed; static polarizability calculations are denoted "Stat," while dynamic polarizability calculations at 800 nm frequency are denoted "Dyn" in this column.

Compound	N_{at}	Method/basis sets	p	Conv	nKs	nKs+RI
Coumarin[71]	17 (C_s)	PBE0/TZVPD	Dyn	0.85	1.00	1.17
YP Chromophore[31]	20	PBE0/TZVP	20	1.57	1.25	1.58
DPA[39]	24 (C_2)	PBE0/QZVP	20	22.90	1.17	1.65
D-Luciferin[31]	26	PBE0/TZVPD	Dyn	4.89	1.17	1.49
		PBE0/TZVPD	5	6.90	1.32	1.70
cis-Thioindigo[72]	28 (C_2)	B3LYP/SVPD	Stat	0.34	1.28	1.63
		TDA-B3LYP/SVPD	10	0.77	1.27	1.25
trans-Thioindigo[72]	28	TDA-B3LYP/SVPD	5	1.09	1.24	1.30
2,6-APD[73]	39	BHLYP/SVPD	Stat	1.17	1.26	1.63
		TDHF/SVPD	10	4.14	1.27	1.54
2,6-ACH[73]	65	PBE0/SVPD	Stat	4.79	1.33	2.02
		BH-LYP/SV(P)	10	1.38	1.52	1.89
		TDA-BHLYP/SV(P)	10	1.23	1.46	1.82
		BH-LYP/SV(P)	10	1.47	1.43	1.76
Tris-Ytterbium(III)[74]	67	PBE0/SV(P)	100	27.4	1.26	4.88
AF-270[75]	79	PBE0/SVPD	Stat	9.08	1.29	1.98
		PBE0/TZVP	1	11.30	1.44	2.89
		TDA-PBE0/TZVP	1	9.29	1.46	2.58
		TDHF/TZVP	1	16.75	1.83	3.85
Astaxanthin[76]	96	PBE0/SVPD	Stat	7.79	1.27	1.98
		PBE0/TZVP	1	7.64	1.40	2.79
AF-295[75]	124	PBE0/TZVP	1	34.24	1.42	3.81
Ru(II)polypyridyl[77]	132 (C_{2h})	PBE0/SVP	Dyn	3.05	1.38	2.28
		PBE0/SVP	10	6.67	1.68	3.02
		TDA-PBE0/SVP	10	5.96	1.69	3.12
		PBE0/SVP	10	11.72	1.33	2.01
Dicyanocobinamide	144	PBE0/SV(P)	50	103.3	1.25	2.05
Chlorophyll A dimer[78]	191	PBE0/SV(P)	50	211.1	1.62	3.21
C60Im-ZnP-BDP[79]	202	PBE0/SV(P)	Dyn	39.66	1.44	2.72
		PBE0/SV(P)	5	17.41	1.80	3.32
		TDA-PBE0/SV(P)	5	14.01	1.69	3.04
CdSe cluster[80]	282 (T)	PBE0/SV(P)	5	7.82	1.53	2.91
Thiophene dendrimer[81]	296	PBE0/SV(P)	Stat	15.80	1.49	4.36
		PBE0/SV(P)	1	19.31	1.48	4.48
		TDA-PBE0/SV(P)	1	16.67	1.48	4.00
		TDHF/SV(P)	1	20.27	1.40	3.95

and half (BHLYP)[84] and three-parameter hybrid (B3LYP)[69] functionals without and with the TDA. Segmented contracted Gaussian basis sets of polarized split valence quality without (SV(P)) and with (SVP) polarization functions on hydrogen atoms were employed along with polarized triple-ζ (TZVP) and polarized quadruple-ζ quality basis sets.[56,85] Property optimized diffuse augmentation,[57] indicated by the letter D, was added in some calculations to explore the effect on timings. Optimized auxiliary basis sets were used for the RI-J approximation.[86] Eigenvectors or solution vectors were converged to a Euclidean residual norm $<1 \times 10^{-5}$; quadrature grids were of size 3.[58] All calculations were performed on a single Intel Nehalem 2.8 GHz X5560 core using up to 12 GB of memory for vector storage.

The nKs algorithm produces speedups of 1.2-1.8 without any further approximations; the one exception is coumarin in C_s symmetry, which is too small for significant integral screening with TZVPD basis sets. Larger speed-ups are seen for smaller basis sets and larger systems, consistent with increased screening efficiency. The speed-up also depends somewhat on the total number of iterations, which tends to be larger for larger systems and smaller values of p. Combining the nKs algorithm with the RI-J approximation for Coulomb contractions and integral direct exchange methods yields two- to fivefold speedups in most cases. The speed-ups are smaller than for non-hybrid functionals, but considering the importance of hybrid exchange for molecular response properties this is a significant result. For example, the time for computing the first singlet excitation energy of the dendrimer from Ref. 81 containing 42 thiophene units ($n = 2\,116\,551$) is reduced from ~19 h to ~4 h.

Table III compares computed isotropic polarizabilities using the conventional and the nKs methods. The results obtained with the nKs algorithm agree perfectly with those obtained conventionally. This parallels the results for excitation energies in Table I and illustrates that the nKs algorithm does not lead to any appreciable loss of precision. Errors introduced by the RI-J approximation are somewhat larger but still an order of magnitude below typical method errors, in agreement with previous observations for non-hybrid functionals.[21,22]

The nKs algorithms expand entire p-dimensional invariant subspaces of eigenvectors or blocks of solution vectors into one single Krylov space. This considerably reduces the cost per eigenvector or solution vector for $p > 1$, as illustrated in Fig. 2 for the 132 atom ruthenium(II)polypyridyl luminescent DNA probe from Ref. 77. The sharp drop for small p reflects the fact that integrals and many other intermediates need to be computed only once for each block of matrix-vector products. Moreover, the number of iterations decreases for larger p because the Krylov space dimension q is larger. This effect is also important for analytical Hessian calculations, where p scales with the number of nuclear degrees of freedom.[48] For large p, the cost per vector is dominated by the actual multiplication step and becomes constant.

Our implementation of nKs methods for TDHF and TDDFT response methods will be made available in Turbomole[82] 7.1.

TABLE III. Comparison of isotropic polarizabilities (a.u.) computed with the conventional algorithm (Conv) and the nKs method without and with the RI-J approximation. ω denotes the frequency in nm. Details as in Table II.

Compound	Method	ω	Conv	nKs	nKs+RI
Astaxanthin	PBE0/SVPD	0	1131.8	1131.8	1132.3
Coumarin	PBE0/TZVPD	800	117.6	117.6	117.7
C60Im-ZnP-BDP	PBE0/SV(P)	800	1696.9	1696.9	1697.1
Thiophene dendrimer	PBE0/SV(P)	0	3102.7	3102.7	3104.2
D-Luciferin	PBE0/TZVPD	800	218.9	218.9	219.0
Ruthenium(II)polypyridyl	PBE0/SVP	800	1148.1	1148.1	1148.4
cis-Thioindigo	B3LYP/SVPD	0	267.2	267.2	267.4
2,6-APD	PBE0/SVPD	0	272.3	272.3	272.5
2,6-ACH	BHLYP/SVPD	0	477.5	477.5	477.8
AF-270	PBE0/SVPD	0	640.7	640.7	641.1

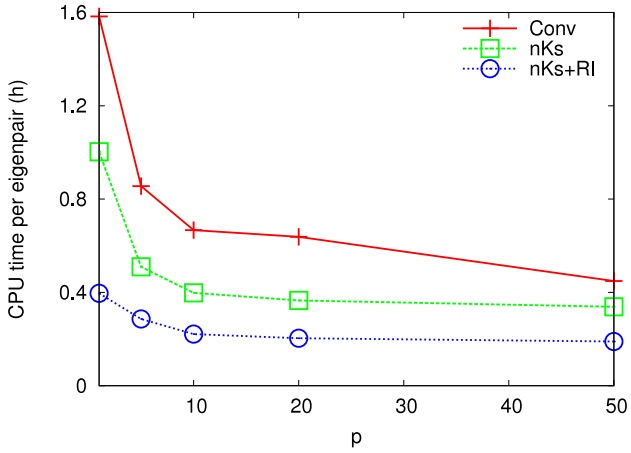

FIG. 2. Comparison of CPU time per eigenpair for the conventional (Conv) and nKs methods without and with the RI-J approximation as a function of the number of eigenpairs p for the 132 atom ruthenium(II)polypyridyl complex.[77] The PBE0 functional and def2-SVP basis sets along with size 3 grids were used. Calculations were performed on a single Intel Nehalem 2.8 GHz X5560 core using up to 8 GB of memory for vector storage.

IV. CONCLUSIONS

In all iterative methods, the norms of the residual vectors become small as convergence is reached. This implies that the magnitude of the corrections to the iterated eigenpairs or solution vectors decreases. Nevertheless, in conventional Krylov space methods, the computational cost is approximately the same for every iteration; as illustrated in Sec. II F, these methods spend increasing effort on accurately computing a part of the subspace which is orthogonal to the desired eigenvectors or solution vectors and later discarded. The nKs methods introduced here address this shortcoming by taking advantage of the decreasing residual norm to reduce the cost of the rate-determining matrix-vector multiplication. Thus, the time spent per iteration progressively decreases as convergence is reached. This introduces additional errors in the Rayleigh matrix, but the error in the desired eigenvectors or solution vectors remains small.

The nKs algorithm achieves computational equivalence of a ground-state complex Fock matrix build with a single

matrix-vector multiplication in response theory. In typical hybrid TDDFT calculations, ∼10 iterations are needed for convergence to residual norms $<10^{-5}$. The number of complex Fock matrix constructions required for similar accuracy in real-time propagation approaches is orders of magnitude higher.[87,88]

The nKs algorithms are derived by transformation of the original eigenvalue problems or linear equation systems to optimization problems. This has apparently not been achieved in the symplectic case before, and it provides additional justification for the algorithm by Olsen, Jensen, and Jørgensen.[10]

Recent work on Krylov space methods has focused on the development of improved preconditioners to reduce the *number* of iterations.[5,66,89] The nKs approach reduces the *cost* per iteration, and it is conceivable that combination with advanced preconditioning will result in further efficiency gains. Our method can be combined with interior eigenvalue solvers[90,91] or methods for complex polarization propagator calculations.[92,93] Extension to unsymmetric eigenvalue problems occurring, e.g., in CC response theory,[94,95] is also possible.

Our method is most effective for coefficient matrices with large bandwidth and irregular sparsity which need to be computed on the fly during matrix-vector operations. Thus, the present results broadly apply to integral-direct electronic structure methods using atom-centered basis sets, including ground and excited state calculations and molecular response calculations. Plane-wave or numerical basis sets, on the other hand, lead to much larger vector space dimensions with different sparsity patterns and require different techniques.[96,97] Since many electronic structure codes already use density matrix screening to reduce integral computation and contraction costs, the nKs method can be adopted with minor code modifications. The observed speedups of 1.2-1.8 in total computation times may not be spectacular, but they are consistent, do not affect accuracy, and can be harnessed for a wide range of electronic structure methods and molecular properties.

Up to fivefold speedups in TDHF and hybrid TDDFT calculations using state-of-the-art functionals and basis sets are realized by combining the nKs method with linear-scaling exchange methods that benefit strongly from sparsity, and treating Coulomb contractions by RI-J methods. This is similar to typical speedups observed in ground-state hybrid TDDFT calculations using difference density methods.[70] Especially for larger systems and moderate-size basis sets, the nKs approach with RI-J is comparable in efficiency with semi-numerical and pseudospectral approaches for exchange contractions[30,31,98] without additional approximations. Since hybrid functionals are crucial for controlling self-interaction error in TDDFT applications, this widens the scope of TDDFT response methods in applications to spectroscopy and non-adiabatic molecular dynamics simulations.

ACKNOWLEDGMENTS

This material is based on work supported by the Department of Energy under Award No. DE-SC0008694.

APPENDIX: STABLE SOLUTION OF REDUCED GENERALIZED EIGENVALUE AND LINEAR PROBLEMS

1. Real symmetric generalized eigenvalue problems

We consider direct methods for solving real symmetric generalized eigenvalue problems,

$$\mathbf{ax} = \mathbf{sx\Omega}, \tag{A1}$$

where $\mathbf{a}$ is symmetric $q \times q$, $\mathbf{s}$ is symmetric $q \times q$ and positive definite, and $\mathbf{\Omega}$ is $q \times q$ and diagonal. The non-singular $q \times q$ eigenvector matrix is $\mathbf{s}$-orthogonal,

$$\mathbf{x}^T\mathbf{sx} = \mathbf{1}. \tag{A2}$$

Among the many well-established direct methods for solving Eq. (A1),[99] a particularly efficient and convenient one is to first compute the Cholesky decomposition,

$$\mathbf{LL}^T = \mathbf{d}^{-1/2}\mathbf{sd}^{-1/2}. \tag{A3}$$

The scaling by $\mathbf{d} = \mathrm{diag}(s_{11}, \ldots, s_{qq})$ is crucial to reduce the condition number of $\mathbf{s}$, which increases rapidly once the residual norm begins to decrease. Transforming $\mathbf{a}$ according to

$$\tilde{\mathbf{a}} = \mathbf{L}^{-1}\mathbf{d}^{-1/2}\mathbf{ad}^{-1/2}\mathbf{L}^{-T} \tag{A4}$$

leads to the symmetric eigenvalue problem

$$\tilde{\mathbf{a}}\tilde{\mathbf{x}} = \tilde{\mathbf{x}}\mathbf{\Omega}, \tag{A5}$$

where $\tilde{\mathbf{x}}$ is $q \times q$ orthogonal. The $\mathbf{s}$-orthogonal eigenvectors are obtained from

$$\mathbf{x} = \mathbf{d}^{-1/2}\mathbf{L}^{-T}\tilde{\mathbf{x}}. \tag{A6}$$

2. Generalized linear problems with a symmetric coefficient matrix

Here we consider the generalized Sylvester equation

$$\mathbf{ax} - \mathbf{sx}\omega = \mathbf{p}, \tag{A7}$$

where $\mathbf{a}$ and $\mathbf{s}$ are defined as in Subsection 1 of the Appendix, the RHS $\mathbf{p}$ and the solution $\mathbf{x}$ are $q \times p$, and ω is symmetric $p \times p$ but not necessarily diagonal.

In the first step, $\mathbf{s}$ is factorized according to Eq. (A3). The transformed coefficient matrix $\tilde{\mathbf{a}}$ is then computed using Eq. (A4), and unless it is already diagonal, the frequency matrix ω is diagonalized according to

$$\omega = \mathbf{t}\tilde{\omega}\mathbf{t}^T, \tag{A8}$$

i.e., $\tilde{\omega}$ is $p \times p$ diagonal and $\mathbf{t}$ is $p \times p$ orthogonal. The transformed RHS is obtained from

$$\tilde{\mathbf{p}} = \mathbf{L}^{-1}\mathbf{d}^{-1/2}\mathbf{pt}. \tag{A9}$$

It is not necessary to decompose $\mathbf{s}$ and transform to an orthogonal basis to solve Eq. (A7), but doing so improves the condition of the problem. Equation (A7) now takes the form of a standard Sylvester equation,

$$\tilde{\mathbf{a}}\tilde{\mathbf{x}} - \tilde{\mathbf{x}}\tilde{\omega} = \tilde{\mathbf{p}}. \tag{A10}$$

Methods for solving this type of linear problem are well established. Since $\tilde{\omega}$ is diagonal, a straightforward approach

is to consider Eq. (A10) as p linear equations for each of the column vectors $\tilde{\mathbf{x}}_i$ and $\tilde{\mathbf{p}}_i$ with a symmetric coefficient matrix $\tilde{\mathbf{a}} - \tilde{\omega}_i$, which may be solved by standard elimination methods such as LU factorization. If p is large and all $\tilde{\omega}_i$ are different, it is more efficient to diagonalize $\tilde{\mathbf{a}}$ once for all p vectors. This method is also preferable close to resonances, where $\tilde{\mathbf{a}} - \tilde{\omega}_i$ is nearly singular. In the static limit, $\boldsymbol{\omega} = \mathbf{0}$, and thus Cholesky decomposition of $\tilde{\mathbf{a}}$ may be used for positive definite coefficient matrices $\mathbf{a}$. This covers the important special case of static response of a stable ground state. The solution of the original problem (A7) is obtained by transforming the solution $\tilde{\mathbf{x}}$ of (A10) according to

$$\mathbf{x} = \mathbf{d}^{-1/2}\mathbf{L}^{-T}\tilde{\mathbf{x}}\mathbf{t}^T. \tag{A11}$$

3. Real symplectic generalized eigenvalue problems

The generalized symplectic eigenvalue problem occurring in Sec. II E has the form

$$\begin{pmatrix}\mathbf{a} & \mathbf{b}\\ \mathbf{b} & \mathbf{a}\end{pmatrix}\begin{pmatrix}\mathbf{x} & \mathbf{y}\\ \mathbf{y} & \mathbf{x}\end{pmatrix} = \begin{pmatrix}\boldsymbol{\sigma} & \boldsymbol{\pi}\\ -\boldsymbol{\pi} & -\boldsymbol{\sigma}\end{pmatrix}\begin{pmatrix}\mathbf{x} & \mathbf{y}\\ \mathbf{y} & \mathbf{x}\end{pmatrix}\begin{pmatrix}\boldsymbol{\Omega} & \mathbf{0}\\ \mathbf{0} & -\boldsymbol{\Omega}\end{pmatrix}, \tag{A12}$$

where $\mathbf{a}$, $\mathbf{b}$, and $\boldsymbol{\sigma}$ are $q \times q$ symmetric, $\boldsymbol{\pi}$ is $q \times q$ antisymmetric, $\boldsymbol{\Omega}$ is $p \times p$ diagonal, and the $q \times p$ matrices $\mathbf{x}$ and $\mathbf{y}$ satisfy the symplectic orthonormality condition

$$\begin{pmatrix}\mathbf{x} & \mathbf{y}\\ \mathbf{y} & \mathbf{x}\end{pmatrix}^T\begin{pmatrix}\boldsymbol{\sigma} & \boldsymbol{\pi}\\ -\boldsymbol{\pi} & -\boldsymbol{\sigma}\end{pmatrix}\begin{pmatrix}\mathbf{x} & \mathbf{y}\\ \mathbf{y} & \mathbf{x}\end{pmatrix} = \begin{pmatrix}\mathbf{1} & \mathbf{0}\\ \mathbf{0} & -\mathbf{1}\end{pmatrix}. \tag{A13}$$

It is further assumed that $\mathbf{a} \pm \mathbf{b}$ is positive definite, which corresponds to the ground-state stability condition.[100]

As in the symmetric case, we first attempt to decompose the metric on the RHS of Eq. (A12). Cholesky decomposition is not possible here because the metric is not positive definite. However, by performing the LU decomposition

$$\mathbf{LU} = \mathbf{d}^{-1/2}(\boldsymbol{\sigma} - \boldsymbol{\pi})\mathbf{d}^{-1/2}, \tag{A14}$$

where $\mathbf{d} = \mathrm{diag}(|\sigma_{11}|, \ldots, |\sigma_{qq}|)$, we obtain the upper triangular matrices

$$\mathbf{K}_\pm = \frac{1}{2}(\mathbf{L}^T \pm \mathbf{U}). \tag{A15}$$

$\mathbf{K}_\pm$ affords the decomposition

$$\begin{pmatrix}\boldsymbol{\sigma} & \boldsymbol{\pi}\\ -\boldsymbol{\pi} & -\boldsymbol{\sigma}\end{pmatrix} = \begin{pmatrix}\mathbf{K}_+ & \mathbf{K}_-\\ \mathbf{K}_- & \mathbf{K}_+\end{pmatrix}^T\begin{pmatrix}\mathbf{1} & \mathbf{0}\\ \mathbf{0} & -\mathbf{1}\end{pmatrix}\begin{pmatrix}\mathbf{K}_+ & \mathbf{K}_-\\ \mathbf{K}_- & \mathbf{K}_+\end{pmatrix}, \tag{A16}$$

as may be verified by insertion of Eqs. (A15) and (A14) into Eq. (A16). Defining

$$\begin{aligned}\tilde{\mathbf{a}} + \tilde{\mathbf{b}} &= \mathbf{L}^{-1}\mathbf{d}^{-1/2}(\mathbf{a} + \mathbf{b})\mathbf{d}^{-1/2}\mathbf{L}^{-T},\\ \tilde{\mathbf{a}} - \tilde{\mathbf{b}} &= \mathbf{U}^{-T}\mathbf{d}^{-1/2}(\mathbf{a} - \mathbf{b})\mathbf{d}^{-1/2}\mathbf{U}^{-1},\end{aligned} \tag{A17}$$

the original problem (A1) is transformed to the standard symplectic eigenvalue problem

$$\begin{pmatrix}\tilde{\mathbf{a}} & \tilde{\mathbf{b}}\\ \tilde{\mathbf{b}} & \tilde{\mathbf{a}}\end{pmatrix}\begin{pmatrix}\tilde{\mathbf{x}} & \tilde{\mathbf{y}}\\ \tilde{\mathbf{y}} & \tilde{\mathbf{x}}\end{pmatrix} = \begin{pmatrix}\mathbf{1} & \mathbf{0}\\ \mathbf{0} & -\mathbf{1}\end{pmatrix}\begin{pmatrix}\tilde{\mathbf{x}} & \tilde{\mathbf{y}}\\ \tilde{\mathbf{y}} & \tilde{\mathbf{x}}\end{pmatrix}\begin{pmatrix}\boldsymbol{\Omega} & \mathbf{0}\\ \mathbf{0} & -\boldsymbol{\Omega}\end{pmatrix} \tag{A18}$$

with the symplectic orthonormality constraint for the $q \times q$ vectors $\tilde{\mathbf{x}}$, $\tilde{\mathbf{y}}$,

$$\begin{pmatrix}\tilde{\mathbf{x}} & \tilde{\mathbf{y}}\\ \tilde{\mathbf{y}} & \tilde{\mathbf{x}}\end{pmatrix}^T\begin{pmatrix}\mathbf{1} & \mathbf{0}\\ \mathbf{0} & -\mathbf{1}\end{pmatrix}\begin{pmatrix}\tilde{\mathbf{x}} & \tilde{\mathbf{y}}\\ \tilde{\mathbf{y}} & \tilde{\mathbf{x}}\end{pmatrix} = \begin{pmatrix}\mathbf{1} & \mathbf{0}\\ \mathbf{0} & -\mathbf{1}\end{pmatrix}. \tag{A19}$$

Cholesky decomposition of $\tilde{\mathbf{a}} - \tilde{\mathbf{b}}$,

$$\mathbf{GG}^T = \tilde{\mathbf{a}} - \tilde{\mathbf{b}} \tag{A20}$$

leads to the standard eigenvalue problem for the squared eigenvalues

$$\mathbf{MZ} = \mathbf{Z}\boldsymbol{\Omega}^2, \tag{A21}$$

where

$$\mathbf{M} = \mathbf{G}^T(\tilde{\mathbf{a}} + \tilde{\mathbf{b}})\mathbf{G} \tag{A22}$$

is $q \times q$ symmetric and positive definite, and $\mathbf{Z}$ is $q \times q$ orthogonal. The eigenvectors in Eq. (A18) are related to $\mathbf{Z}$ by

$$\begin{aligned}\tilde{\mathbf{x}} + \tilde{\mathbf{y}} &= \mathbf{GZ}\boldsymbol{\Omega}^{-1/2},\\ \tilde{\mathbf{x}} - \tilde{\mathbf{y}} &= \mathbf{G}^{-T}\mathbf{Z}\boldsymbol{\Omega}^{1/2},\end{aligned} \tag{A23}$$

and thus the eigenvectors of the original problem (A12) are

$$\begin{aligned}\mathbf{x} + \mathbf{y} &= \mathbf{d}^{-1/2}\mathbf{L}^{-T}(\tilde{\mathbf{x}} + \tilde{\mathbf{y}}) = \mathbf{d}^{-1/2}\mathbf{L}^{-T}\mathbf{GZ}\boldsymbol{\Omega}^{-1/2},\\ \mathbf{x} - \mathbf{y} &= \mathbf{d}^{-1/2}\mathbf{U}^{-1}(\tilde{\mathbf{x}} + \tilde{\mathbf{y}}) = \mathbf{d}^{-1/2}\mathbf{U}^{-1}\mathbf{G}^{-T}\mathbf{Z}\boldsymbol{\Omega}^{1/2}.\end{aligned} \tag{A24}$$

4. Generalized linear problems with a symplectic coefficient matrix

We aim to solve the generalized symplectic linear problem

$$\begin{pmatrix}\mathbf{a} & \mathbf{b}\\ \mathbf{b} & \mathbf{a}\end{pmatrix}\begin{pmatrix}\mathbf{x} & \mathbf{y}\\ \mathbf{y} & \mathbf{x}\end{pmatrix} - \begin{pmatrix}\boldsymbol{\sigma} & \boldsymbol{\pi}\\ -\boldsymbol{\pi} & -\boldsymbol{\sigma}\end{pmatrix}\begin{pmatrix}\mathbf{x} & \mathbf{y}\\ \mathbf{y} & \mathbf{x}\end{pmatrix}\begin{pmatrix}\boldsymbol{\omega} & \boldsymbol{\delta}\\ -\boldsymbol{\delta} & -\boldsymbol{\omega}\end{pmatrix} = \begin{pmatrix}\mathbf{p} & \mathbf{q}\\ \mathbf{q} & \mathbf{p}\end{pmatrix}, \tag{A25}$$

by direct methods. $\mathbf{a}$, $\mathbf{b}$, $\boldsymbol{\sigma}$, and $\boldsymbol{\pi}$ are defined as in Subsection 3 of the Appendix, and the RHS and solution vectors $\mathbf{p}$, $\mathbf{q}$ and $\mathbf{x}$, $\mathbf{y}$ are $q \times p$. $\boldsymbol{\omega}$ and $\boldsymbol{\delta}$ are $p \times p$ but not necessarily diagonal.

As in Subsection 3 of the Appendix, the symplectic metric may be factorized using Eqs. (A14)-(A16). Further, the frequency matrix is diagonalized by solving the symplectic eigenvalue problem

$$\begin{pmatrix}\boldsymbol{\omega} & \boldsymbol{\delta}\\ \boldsymbol{\delta} & \boldsymbol{\omega}\end{pmatrix}\begin{pmatrix}\mathbf{t} & \mathbf{u}\\ \mathbf{u} & \mathbf{t}\end{pmatrix} = \begin{pmatrix}\mathbf{1} & \mathbf{0}\\ \mathbf{0} & -\mathbf{1}\end{pmatrix}\begin{pmatrix}\mathbf{t} & \mathbf{u}\\ \mathbf{u} & \mathbf{t}\end{pmatrix}\begin{pmatrix}\tilde{\boldsymbol{\omega}} & \mathbf{0}\\ \mathbf{0} & -\tilde{\boldsymbol{\omega}}\end{pmatrix} \tag{A26}$$

under the symplectic orthonormality constraint for the $p \times p$ vectors $\mathbf{t}$, $\mathbf{u}$,

$$\begin{pmatrix}\mathbf{t} & \mathbf{u}\\ \mathbf{u} & \mathbf{t}\end{pmatrix}^T\begin{pmatrix}\mathbf{1} & \mathbf{0}\\ \mathbf{0} & -\mathbf{1}\end{pmatrix}\begin{pmatrix}\mathbf{t} & \mathbf{u}\\ \mathbf{u} & \mathbf{t}\end{pmatrix} = \begin{pmatrix}\mathbf{1} & \mathbf{0}\\ \mathbf{0} & -\mathbf{1}\end{pmatrix}, \tag{A27}$$

using the techniques of Subsection 3 of the Appendix. Here it has been assumed that $\boldsymbol{\omega} \pm \boldsymbol{\delta}$ is positive definite, which corresponds to the convention that frequencies of perturbations be chosen positive. If the frequency matrix is already diagonal, $\mathbf{t} = \mathbf{1}$ and $\mathbf{u} = \mathbf{0}$. It follows from Eqs. (A26) and (A27) that

$$\begin{pmatrix}\boldsymbol{\omega} & \boldsymbol{\delta}\\ -\boldsymbol{\delta} & -\boldsymbol{\omega}\end{pmatrix} = \begin{pmatrix}\mathbf{t} & \mathbf{u}\\ \mathbf{u} & \mathbf{t}\end{pmatrix}\begin{pmatrix}\tilde{\boldsymbol{\omega}} & \mathbf{0}\\ \mathbf{0} & \tilde{\boldsymbol{\omega}}\end{pmatrix}\begin{pmatrix}\mathbf{t} & \mathbf{u}\\ \mathbf{u} & \mathbf{t}\end{pmatrix}^T\begin{pmatrix}\mathbf{1} & \mathbf{0}\\ \mathbf{0} & -\mathbf{1}\end{pmatrix}. \tag{A28}$$

After decomposing the metric according to Eqs. (A14)–(A16), the transformed matrices $\tilde{\mathbf{a}}$ and $\tilde{\mathbf{b}}$ are obtained from Eq. (A17); the transformed RHS is

$$\begin{aligned}\tilde{\mathbf{p}} + \tilde{\mathbf{q}} &= \mathbf{L}^{-1}\mathbf{d}^{-1/2}(\mathbf{p} + \mathbf{q})(\mathbf{t} + \mathbf{u}),\\ \tilde{\mathbf{p}} - \tilde{\mathbf{q}} &= \mathbf{U}^{-T}\mathbf{d}^{-1/2}(\mathbf{p} - \mathbf{q})(\mathbf{t} - \mathbf{u}).\end{aligned} \tag{A29}$$

Eq. (A25) thus becomes

$$\begin{pmatrix}\tilde{\mathbf{a}} & \tilde{\mathbf{b}}\\ \tilde{\mathbf{b}} & \tilde{\mathbf{a}}\end{pmatrix}\begin{pmatrix}\tilde{\mathbf{x}} & \tilde{\mathbf{y}}\\ \tilde{\mathbf{y}} & \tilde{\mathbf{x}}\end{pmatrix} - \begin{pmatrix}\mathbf{1} & \mathbf{0}\\ \mathbf{0} & -\mathbf{1}\end{pmatrix}\begin{pmatrix}\tilde{\mathbf{x}} & \tilde{\mathbf{y}}\\ \tilde{\mathbf{y}} & \tilde{\mathbf{x}}\end{pmatrix}\begin{pmatrix}\tilde{\omega} & \mathbf{0}\\ \mathbf{0} & -\tilde{\omega}\end{pmatrix} = \begin{pmatrix}\tilde{\mathbf{p}} & \tilde{\mathbf{q}}\\ \tilde{\mathbf{q}} & \tilde{\mathbf{p}}\end{pmatrix}. \tag{A30}$$

This symplectic linear problem may further be transformed to a linear problem of half the dimension by decomposing $\tilde{\mathbf{a}} - \tilde{\mathbf{b}}$ according to Eq. (A20),

$$\mathbf{MZ} - \mathbf{Z}\tilde{\omega}^2 = \mathbf{T}. \tag{A31}$$

The symmetric coefficient matrix $\mathbf{M}$ is given by (A22), and the RHS

$$\mathbf{T} = \mathbf{G}^T(\tilde{\mathbf{p}} + \tilde{\mathbf{q}}) + \mathbf{G}^{-1}(\tilde{\mathbf{p}} - \tilde{\mathbf{q}})\tilde{\omega} \tag{A32}$$

is frequency-dependent. Equation (A31) is a Sylvester equation of the form (A10) and the solution $\mathbf{Z}$ may be determined by the methods outlined in Subsection 2 of the Appendix. The solutions of Eq. (A30) are thus

$$\begin{aligned}\tilde{\mathbf{x}} + \tilde{\mathbf{y}} &= \mathbf{GZ},\\ \tilde{\mathbf{x}} - \tilde{\mathbf{y}} &= \mathbf{G}^{-T}\mathbf{Z},\end{aligned} \tag{A33}$$

while the solutions of original problem (A25) are

$$\begin{aligned}\mathbf{x} + \mathbf{y} &= \mathbf{d}^{-1/2}\mathbf{L}^{-T}(\tilde{\mathbf{x}} + \tilde{\mathbf{y}})(\mathbf{t} - \mathbf{u})^T\\ &= \mathbf{d}^{-1/2}\mathbf{L}^{-T}\mathbf{GZ}(\mathbf{t} - \mathbf{u})^T,\\ \mathbf{x} - \mathbf{y} &= \mathbf{d}^{-1/2}\mathbf{U}^{-1}(\tilde{\mathbf{x}} + \tilde{\mathbf{y}})(\mathbf{t} + \mathbf{u})^T\\ &= \mathbf{d}^{-1/2}\mathbf{U}^{-1}\mathbf{G}^{-T}\mathbf{Z}(\mathbf{t} + \mathbf{u})^T.\end{aligned} \tag{A34}$$

[1]T. Helgaker, S. Coriani, P. Jørgensen, K. Kristensen, J. Olsen, and K. Ruud, Chem. Rev. **112**, 543 (2012).
[2]J. Olsen and P. Jørgensen, J. Chem. Phys. **82**, 3235 (1985).
[3]O. Christiansen, P. Jørgensen, and C. Hättig, Int. J. Quantum Chem. **68**, 1 (1998).
[4]P. Norman, Phys. Chem. Chem. Phys. **13**, 20519 (2011).
[5]Y. Saad, *Iterative Methods for Sparse Linear Systems* (SIAM, 2003).
[6]E. R. Davidson, J. Comput. Phys. **17**, 87 (1975).
[7]B. Roos, Chem. Phys. Lett. **15**, 153 (1972).
[8]T. D. Bouman, A. E. Hansen, B. Voigt, and S. Rettrup, Int. J. Quantum Chem. **23**, 595 (1983).
[9]M. Feyereisen, J. Nichols, J. Oddershede, and J. Simons, J. Chem. Phys. **96**, 2978 (1992).
[10]J. Olsen, H. J. A. Jensen, and P. Jørgensen, J. Comput. Phys. **74**, 265 (1988).
[11]H. Weiss, R. Ahlrichs, and M. Häser, J. Chem. Phys. **99**, 1262 (1993).
[12]J. Almlöf, K. Faegri, and K. Korsell, J. Comput. Chem. **3**, 385 (1982).
[13]M. Häser and R. Ahlrichs, J. Comput. Chem. **10**, 104 (1989).
[14]M. Challacombe and E. Schwegler, J. Chem. Phys. **106**, 5526 (1997).
[15]C. Fonseca Guerra, J. G. Snijders, G. te Velde, and E. J. Baerends, Theor. Chem. Acc. **99**, 391 (1998).
[16]R. Bauernschmitt and R. Ahlrichs, Chem. Phys. Lett. **256**, 454 (1996).
[17]R. E. Stratmann, G. E. Scuseria, and M. J. Frisch, J. Chem. Phys. **109**, 8218 (1998).
[18]S. J. A. van Gisbergen, J. G. Snijders, and E. J. Baerends, Comput. Phys. Commun. **118**, 119 (1999).
[19]F. Furche and D. Rappoport, "Density functional methods for excited states: Equilibrium structure and electronic spectra," in *Computational Photochemistry* (Elsevier, Amsterdam, 2005), pp. 93–128.
[20]A. Halkier, H. Koch, O. Christiansen, P. Jørgensen, and T. Helgaker, J. Chem. Phys. **107**, 849 (1997).
[21]R. Bauernschmitt, M. Häser, O. Treutler, and R. Ahlrichs, Chem. Phys. Lett. **264**, 573 (1997).
[22]D. Rappoport and F. Furche, J. Chem. Phys. **122**, 064105 (2005).
[23]F. Neese and G. Olbrich, Chem. Phys. Lett. **362**, 170 (2002).
[24]M. Beer and C. Ochsenfeld, J. Chem. Phys. **128**, 221102 (2008).
[25]J. Kussmann and C. Ochsenfeld, J. Chem. Phys. **127**, 204103 (2007).
[26]S. Coriani, S. Host, B. Jansik, L. Thogersen, J. Olsen, P. Jørgensen, S. Reine, F. Pawlowski, T. Helgaker, and P. Salek, J. Chem. Phys. **126**, 154108 (2007).
[27]M. J. Lucero, A. M. N. Niklasson, S. Tretiak, and M. Challacombe, J. Chem. Phys. **129**, 064114 (2008).
[28]S. Tretiak, C. M. Isborn, A. M. N. Niklasson, and M. Challacombe, J. Chem. Phys. **130**, 054111 (2009).
[29]T. Touma, M. Kobayashi, and H. Nakai, Chem. Phys. Lett. **485**, 247 (2010).
[30]T. Petrenko, S. Kossmann, and F. Neese, J. Chem. Phys. **134**, 054116 (2011).
[31]C. Ko, D. K. Malick, D. A. Braden, R. A. Friesner, and T. J. Martínez, J. Chem. Phys. **128**, 104103 (2008).
[32]T. M. Maier, H. Bahmann, and M. Kaupp, J. Chem. Theory Comput. **11**, 4226 (2015).
[33]D. Cremer and J. Gauss, J. Comput. Chem. **7**, 274 (1986).
[34]C. Ochsenfeld, C. A. White, and M. Head-Gordon, J. Chem. Phys. **109**, 1663 (1998).
[35]E. Schwegler, M. Challacombe, and M. Head-Gordon, J. Chem. Phys. **106**, 9708 (1997).
[36]E. R. Davidson, Comput. Phys. Commun. **53**, 49 (1989).
[37]A. Dreuw, J. L. Weisman, and M. Head-Gordon, J. Chem. Phys. **119**, 2943 (2003).
[38]Y. Zhao and D. G. Truhlar, J. Phys. Chem. A **110**, 13126 (2006).
[39]D. Jacquemin, E. A. Perpete, G. E. Scuseria, I. Ciofini, and C. Adamo, J. Chem. Theory Comput. **4**, 123 (2008).
[40]R. Send, M. Kühn, and F. Furche, J. Chem. Theory Comput. **7**, 2376 (2011).
[41]E. Tapavicza, G. D. Bellchambers, J. C. Vincent, and F. Furche, Phys. Chem. Chem. Phys. **15**, 18336 (2013).
[42]S. Hirata and M. Head-Gordon, Chem. Phys. Lett. **314**, 291 (1999).
[43]J. B. Foresman, M. Head-Gordon, J. A. Pople, and M. J. Frisch, J. Phys. Chem. **96**, 135 (1992).
[44]M. E. Casida, "Time-dependent density functional response theory for molecules," in *Recent Advances in Density Functional Methods* (World Scientific, Singapore, 1995), pp. 155–192.
[45]M. Crouzeix, B. Philippe, and M. Sadkane, SIAM J. Sci. Comput. **15**, 62 (1994).
[46]J. A. Pople, R. Krishnan, H. B. Schlegel, and J. S. Binkley, Int. J. Quantum Chem. **16**, 225 (1979).
[47]S. M. Colwell, C. W. Murray, N. C. Handy, and R. D. Amos, Chem. Phys. Lett. **210**, 261 (1993).
[48]P. Deglmann, F. Furche, and R. Ahlrichs, Chem. Phys. Lett. **362**, 511 (2002).
[49]A. McLachlan and M. Ball, Rev. Mod. Phys. **36**, 844 (1964).
[50]D. L. Yeager and P. Jørgensen, Chem. Phys. Lett. **65**, 77 (1979).
[51]E. S. Nielsen, P. Jørgensen, and J. Oddershede, J. Chem. Phys. **73**, 6328 (1980).
[52]T. B. Pedersen, H. Koch, and C. Hättig, J. Chem. Phys. **110**, 8318 (1999).
[53]F. Furche, J. Chem. Phys. **114**, 5982 (2001).
[54]T. Ziegler, M. Krykunov, and J. Autschbach, J. Chem. Theory Comput. **10**, 3980 (2014).
[55]J. P. Perdew, M. Ernzerhof, and K. Burke, J. Chem. Phys. **105**, 9982 (1996).
[56]F. Weigend and R. Ahlrichs, Phys. Chem. Chem. Phys. **7**, 3297 (2005).
[57]D. Rappoport and F. Furche, J. Chem. Phys. **133**, 134105 (2010).
[58]O. Treutler and R. Ahlrichs, J. Chem. Phys. **102**, 346 (1995).
[59]C. Lanczos, J. Res. Natl. Bur. Stand. **45**, 255 (1950).
[60]M. R. Hestenes and E. Stiefel, J. Res. Natl. Bur. Stand. **49**, 409 (1952).
[61]J. H. van Lenthe and P. Pulay, J. Comput. Chem. **11**, 1164 (1990).
[62]V. Chernyak, M. F. Schulz, S. Mukamel, S. Tretiak, and E. V. Tsiper, J. Chem. Phys. **113**, 36 (2000).
[63]A. Stathopoulos, Y. Saad, and K. Wu, SIAM J. Sci. Comput. **19**, 227 (1998).
[64]K. Wu and H. Simon, SIAM. J. Matrix Anal. Appl. **22**, 602 (2000).
[65]V. Simoncini and D. B. Szyld, Numer. Math. **100**, 711 (2005).
[66]J. Kauczor, P. Jørgensen, and P. Norman, J. Chem. Theory Comput. **7**, 1610 (2011).
[67]C. C. Paige, IMA J. Appl. Math. **10**, 373 (1972).
[68]J. E. Bates and F. Furche, J. Chem. Phys. **137**, 164105 (2012).
[69]A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).

[70]M. Kattannek, "Entwicklung und Implementierung optimierter Algorithmen für molekulare Hartree-Fock- und Dichtefunktional-Rechnungen," Ph.D. thesis, Universität Karlsruhe (TH), 2006.
[71]D. Jacquemin, E. A. Perpète, G. Scalmani, M. J. Frisch, X. Assfeld, I. Ciofini, and C. Adamo, J. Chem. Phys. **125**, 164324 (2006).
[72]D. Jacquemin, J. Preat, V. Wathelet, M. Fontaine, and E. A. Perpète, J. Am. Chem. Soc. **128**, 2072 (2006).
[73]E. A. Badaeva, T. V. Timofeeva, A. Masunov, and S. Tretiak, J. Phys. Chem. A **109**, 7276 (2005).
[74]J. F. Corbey, D. H. Woen, C. T. Palumbo, M. E. Fieser, J. W. Ziller, F. Furche, and W. J. Evans, Organometallics **34**, 3909 (2015).
[75]G. S. He, T.-C. Lin, J. Dai, P. N. Prasad, R. Kannan, A. G. Dombroskie, R. A. Vaia, and L.-S. Tan, J. Chem. Phys. **120**, 5275 (2004).
[76]A. P. Gamiz-Hernandez, I. N. Angelova, R. Send, D. Sundholm, and V. R. I. Kaila, Angew. Chem., Int. Ed. **54**, 11564 (2015).
[77]M. R. Gill, J. Garcia-Lara, S. J. Foster, C. Smythe, G. Battaglia, and J. A. Thomas, Nat. Chem. **1**, 662 (2009).
[78]S. O. Nilsson Lill, Phys. Chem. Chem. Phys. **13**, 16022 (2011).
[79]F. D'Souza, P. M. Smith, M. E. Zandler, A. L. McCarty, M. Itou, Y. Araki, and O. Ito, J. Am. Chem. Soc. **126**, 7898 (2004).
[80]K. Eichkorn and R. Ahlrichs, Chem. Phys. Lett. **288**, 235 (1998).
[81]E. Badaeva, M. R. Harpham, R. Guda, O. Suzer, C.-Q. Ma, P. Bauerle, T. Goodson III, and S. Tretiak, J. Phys. Chem. B **114**, 15808 (2010).
[82]F. Furche, R. Ahlrichs, C. Hättig, W. Klopper, M. Sierka, and F. Weigend, WIREs: Comput. Mol. Sci. **4**, 91 (2014).
[83]See supplementary material at http://dx.doi.org/10.1063/1.4947245 for atomic coordinates in XYZ format.
[84]A. D. Becke, J. Chem. Phys. **98**, 1372 (1993).
[85]F. Weigend, F. Furche, and R. Ahlrichs, J. Chem. Phys. **119**, 12753 (2003).
[86]F. Weigend, Phys. Chem. Chem. Phys. **8**, 1057 (2006).
[87]Y. Takimoto, F. D. Vila, and J. J. Rehr, J. Chem. Phys. **127**, 154114 (2007).
[88]K. Lopata and N. Govind, J. Chem. Theory Comput. **7**, 1344 (2011).
[89]G. L. Sleijpen and H. A. Van der Vorst, SIAM. J. Matrix Anal. Appl. **17**, 401 (1996).
[90]M. R. Wall and D. Neuhauser, J. Chem. Phys. **102**, 8011 (1995).
[91]W. Liang, S. A. Fischer, M. J. Frisch, and X. Li, J. Chem. Theory Comput. **7**, 3540 (2011).
[92]P. Norman, D. M. Bishop, H. J. A. Jensen, and J. Oddershede, J. Chem. Phys. **123**, 194103 (2005).
[93]L. Jensen, J. Autschbach, and G. C. Schatz, J. Chem. Phys. **122**, 224115 (2005).
[94]K. Hirao and H. Nakatsuji, J. Comput. Phys. **45**, 246 (1982).
[95]J. F. Stanton and R. J. Bartlett, J. Chem. Phys. **98**, 7029 (1993).
[96]T. Yanai, R. J. Harrison, and N. C. Handy, Mol. Phys. **103**, 413 (2005).
[97]D. Rocca, R. Gebauer, Y. Saad, and S. Baroni, J. Chem. Phys. **128**, 154105 (2008).
[98]P. Plessow and F. Weigend, J. Comput. Chem. **33**, 810 (2012).
[99]G. H. Golub and C. F. Van Loan, *Matrix Computations*, 4th ed. (JHU Press, 2013).
[100]J. Čížek and J. Paldus, J. Chem. Phys. **47**, 3976 (1967).