Analytical energy gradient in variational calculations of the two lowest ^3P states of the carbon atom with explicitly correlated Gaussian basis functions

Keeper L. Sharkey,[1] Sergiy Bubin,[2] and Ludwik Adamowicz[1,3,a)]
[1]*Department of Chemistry, University of Arizona, Tucson, Arizona 85721, USA*
[2]*Department of Physics and Astronomy, Vanderbilt University, Nashville, Tennessee 37235, USA*
[3]*Department of Physics, University of Arizona, Tucson, Arizona 85721, USA*

(Received 8 February 2010; accepted 9 April 2010; published online 12 May 2010)

Variational calculations of ground and excited bound states on atomic and molecular systems performed with basis functions that explicitly depend on the interparticle distances can generate very accurate results provided that the basis function parameters are thoroughly optimized by the minimization of the energy. In this work we have derived the algorithm for the gradient of the energy determined with respect to the nonlinear exponential parameters of explicitly correlated Gaussian functions used in calculating n-electron atomic systems with two p-electrons and $(n-2)$ s-electrons. The atomic Hamiltonian we used was obtained by rigorously separating out the kinetic energy of the center of mass motion from the laboratory-frame Hamiltonian and explicitly depends on the finite mass of the nucleus. The advantage of having the gradient available in the variational minimization of the energy is demonstrated in the calculations of the ground and the first excited ^3P state of the carbon atom. For the former the lowest energy upper bound ever obtained is reported.
 [doi:10.1063/1.3419931]

I. INTRODUCTION

Very accurate quantum mechanical calculations of the electronic structures of small atoms have always provided the testing ground for new computational methods for calculating the ground and excited bound states. The testing has been possible due to the availability of very accurate gas-phase spectra of these systems. Previously, the only atoms for which the accuracy of the calculations matched the accuracy of the experiments were the one-, two-, and three-electron ones (see, for example, Refs. 1–8). Recently, it was demonstrated that with the use of the variational method and of the explicitly correlated Gaussian functions (ECGFs), one can also calculate the energies of ground and excited states of four-electron atoms and corresponding electronic transition frequencies with experimental accuracy.[9–14]

In order to extend the range of the atomic states that can be calculated with ECGFs, we have recently derived algorithms for the Hamiltonian matrix elements to calculate states with one and two p-electrons.[15,16] The development of algorithms for calculating states of atoms that include electrons with higher angular momenta using ECGFs is currently in progress. All algorithms have been developed for an arbitrary number of electrons and using the Hamiltonian that explicitly includes the finite mass of the atomic nucleus. This Hamiltonian is obtained by separating out the kinetic energy of the center of mass motion from the total nonrelativistic laboratory-frame Hamiltonian. As this separation is rigorous, the total energies obtained in the calculations correspond to true internal bound states of the studied system. Also, by setting the mass of the nucleus to values corresponding to the different isotopes of the system, the isotope spectral shifts can be calculated. Another advantage of having a variable nuclear mass in the Hamiltonian is the possibility of calculating infinite-nuclear-mass energies. These energies can be directly compared to the energies obtained in the standard atomic calculations performed assuming the Born–Oppenheimer (BO) approximation, which is the way the majority of the atomic calculations have been done. We should add that as some of those calculations have been performed using the reduced electron mass, not all the energies from literature are directly comparable with our results.

Achieving high accuracy in the atomic (and molecular) calculations with ECGFs is possible provided that the nonlinear exponential parameters of Gaussians are extensively optimized based on the variational principle. This usually is a process that takes large amounts of computer time. To accelerate the basis set optimization in the ECGF calculations we have derived and implemented analytical derivatives (i.e., gradient) of the energy with respect to the nonlinear parameters[15,17–21] involved in the Gaussian functions. The gradient-based approach has enabled us to perform very accurate BO and non-BO calculations of atomic and molecular systems with accuracy unmatched by any previous calculation.[22–27]

In this work we continue the development of the gradient algorithms for calculating the atomic energy levels. The work is focused on the first derivatives of the energy with respect to the parameters of explicitly correlated Gaussians describing atomic states with two p-electrons. The derivation of formulas for the gradient presented here involves an approach based on the powerful technique called the matrix

[a)]Electronic mail: ludwik@u.arizona.edu.

differential calculus. This elegant and very useful technique is described in Magnus and Neudecker[28] and, unfortunately, is not well known outside of the disciplines of mathematical statistics and econometrics. Some steps involved in the present derivations will be useful in deriving algorithms for the energy gradients with atomic ECGFs representing states involving electrons with higher angular momenta.

The formulas for the energy gradient derived in this work have been implemented using the Message Passing Interface (MPI) protocol and the calculations have been run on a multiprocessor computer system. Some details concerning the implementation are described in Sec. IV. In the next step we tested the code in calculations concerning the two lowest ^3P states of the carbon atom, which is the smallest system with two p electrons in the ground state. These tests are also described in Sec. IV.

II. THE METHOD

A. The Hamiltonian

For an atom with N particles (i.e., $N-1$ electrons and a nucleus) the laboratory-frame nonrelativistic Hamiltonian has the following form:

$$\hat{H}_{\text{lab}} = -\sum_{i=1}^{N} \frac{1}{2M_i} \nabla_{\mathbf{R}_i}^2 + \sum_{i>j=1}^{N} \frac{Q_i Q_j}{R_{ij}}, \tag{1}$$

where M_i are the masses of the particles, Q_i are their charges, $\mathbf{R}_i$ are the Cartesian position vectors in the coordinate frame, and $\nabla_{\mathbf{R}_i}$ are the gradients with respect to $\mathbf{R}_i$. R_{ij} is the distance between the ith and jth particles, $R_{ij}=|\mathbf{R}_j-\mathbf{R}_i|$. After separating out the center of mass motion of the system from Eq. (1) an internal Hamiltonian, $\hat{H}$, is obtained. This internal Hamiltonian in the internal Cartesian coordinate system with the center at the nucleus has the following form:

$$\hat{H} = -\frac{1}{2}\left(\sum_{i=1}^{n} \frac{1}{\mu_i} \nabla_{\mathbf{r}_i}^2 + \sum_{\substack{i,j=1 \\ i\neq j}}^{n} \frac{1}{m_0} \nabla'_{\mathbf{r}_i} \nabla_{\mathbf{r}_j}\right) + \sum_{i=1}^{n} \frac{q_0 q_i}{r_i} + \sum_{i>j=1}^{n} \frac{q_i q_j}{r_{ij}}, \tag{2}$$

where $n=N-1$ is the number of particles the internal Hamiltonian describes, $\mathbf{r}_i=\mathbf{R}_{i+1}-\mathbf{R}_1$ are their internal Cartesian coordinates, m_0 is the nucleus mass, q_0 is its charge, $q_i=Q_{i+1}$ are the electron charges, and $\mu_i=m_0 m_i/(m_0+m_i)$ are their reduced masses, where $m_i=M_{i+1}$. We call the particles described by Hamiltonian (2) pseudoelectrons because even though they have the same charges as the electrons, their masses are reduced electron masses. Thus, Hamiltonian (2) describes the motion of n pseudoelectrons in the central field of the charge of the nucleus. This motion is coupled through the Coulombic interactions between the pseudoelectrons and their interactions with the nucleus charge,

$$\sum_{i=1}^{n} \frac{q_0 q_i}{r_i} + \sum_{i>j=1}^{n} \frac{q_i q_j}{r_{ij}},$$

where $r_{ij}=|\mathbf{r}_j-\mathbf{r}_i|$, and through the mass polarization term,

$$-\frac{1}{2}\sum_{\substack{i,j=1 \\ i\neq j}}^{n} (1/m_0) \nabla'_{\mathbf{r}_i} \nabla_{\mathbf{r}_j}.$$

The prime in the mass polarization term indicates the matrix/vector transpose; this notation is used throughout this work.

B. The basis functions

The atomic system considered here has two p-electrons and $(n-2)$ s-electrons. The standard procedure for adding angular momenta can be applied to construct basis functions for such a case. In our previous work[16] we showed that for an atom with two p-electrons and the total angular momentum quantum number $L=1$ and its projection on the z-axis $M=0$, such as for the ground state of the carbon atom, the angular factor in the basis functions should be

$$x_i y_j - x_j y_i. \tag{3}$$

Thus, the suitable explicitly correlated Gaussian basis functions for calculating the ground ^3P state of the carbon atom are

$$\phi_k = (x_{i_k} y_{j_k} - x_{j_k} y_{i_k}) \exp[-\mathbf{r}'(A_k \otimes I_3)\mathbf{r}], \tag{4}$$

where i_k and j_k are integers that indicate the label of the p electrons, A_k is an $n\times n$ symmetric matrix, the subscript k reflects the fact that the matrix is unique for each basis function, $\otimes$ is the Kronecker product, I_3 is a 3×3 identity matrix, and $\mathbf{r}$ is a $3n$ vector of the electron (pseudoelectron) coordinates that has the form

$$\mathbf{r} = \begin{pmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \\ \vdots \\ \mathbf{r}_n \end{pmatrix} = \begin{pmatrix} x_1 \\ y_1 \\ z_1 \\ \vdots \\ x_n \\ y_n \\ z_n \end{pmatrix}. \tag{5}$$

In a simplified form, basis function (4) can be written as

$$\phi_k = (x_{i_k} y_{j_k} - x_{j_k} y_{i_k}) \exp[-\mathbf{r}'\mathbf{A}_k\mathbf{r}], \tag{6}$$

where $\mathbf{A}_k=A_k\otimes I_3$. In the approach described in this work we use yet another representation of basis function (4), which is more convenient in the computational implementation and more general in terms of providing a representation for other types of angular ECGFs. The representation utilizes the sparse $3n\times 3n$ symmetric matrix $\mathbf{W}_k$, which for function (4) comprises only four nonzero off-diagonal elements, two of which have values of 1/2 the other two $-1/2$. The 1/2 elements are placed in the $(3i_k-2,3j_k-1)$ and $(3j_k-1,3i_k-2)$ positions, while the $-1/2$ elements are placed in $(3j_k-2,3i_k-1)$ and $(3i_k-1,3j_k-2)$ positions. With such $\mathbf{W}_k$ basis function (4) has the following form:

$$\phi_k = (\mathbf{r}'\mathbf{W}_k\mathbf{r}) \exp[-\mathbf{r}'\mathbf{A}_k\mathbf{r}]. \tag{7}$$

To make function (7) square integrable it is convenient to represent A_k in a Cholesky factored form as $A_k=L_k L'_k$, where

L_k is a lower triangular matrix. With this ϕ_k is automatically square integrable regardless of the values of the L_k matrix elements. The convenience of using the Cholesky representation of A_k comes in the variational minimization of the energy with respect to the exponential parameters of ϕ_k. When those parameters are the L_k matrix elements, the optimization can be carried out without any restrictions concerning the values of the variables. Such restrictions would be necessary if the optimization parameters were the A_k matrix elements.

The lower triangular matrix, L_k, representing the parameters of ϕ_k is stored in a vector form, vech L_k, meaning "vector half." This particular vector operation utilizes the structure of a lower triangular matrix. The vech operation is built by stacking the columns of the matrix one on top of the other, but by using only the lower triangular portion of the matrix, i.e., by starting with the diagonal element of the column and including all elements below the diagonal element of that column. The vech operation transforms an $n\times n$ matrix to an $n(n+1)/2$ vector. This operation effectively eliminates the storage of zeros from the upper triangular of the L_k matrix. The vech operation is also used in the gradient derivation in this work to store the derivatives of the Hamiltonian and overlap matrix elements calculated with respect to the elements of vech L_k.

C. Total energy and energy gradient

The optimization of the L_k Gaussian parameters in this work is performed through the minimization of the Rayleigh–Ritz variational energy functional. The spatial part of the wave function of the system, Ψ, is approximated as a linear combination of K basis functions ϕ_k,

$$\Psi(\mathbf{r}) = \sum_{k=1}^{K} \mathsf{c}_k \hat{Y}\phi_k(\mathbf{r}), \tag{8}$$

where c_k are the linear variational parameters and $\hat{Y}$ is a permutational symmetry projector represented as a linear combination of permutational operators. As the Hamiltonian is spin independent, the calculations can be carried out using the spin-free approach. This requires that the basis functions have the appropriate spatial symmetry. In practice, the implementation of the symmetry is accomplished at the level of calculating the overlap and Hamiltonian matrix elements as it was described in our previous work.[16] In brief, the ket functions in those matrix elements are operated on with the permutation operator $\hat{P}=\hat{Y}^\dagger\hat{Y}$ (the dagger stands for conjugate), where the $\hat{Y}$ operator can be derived using a Young tableaux (see our previous work[16]) suitable for the state under consideration.

The minimization of the Rayleigh–Ritz functional with respect to the c_k coefficients leads to the secular equation

$$(\mathsf{H}-\varepsilon\mathsf{S})\mathsf{c}=0, \tag{9}$$

where H and S are $K\times K$ Hermitian matrices of the Hamiltonian and overlap integrals, with the elements $\mathsf{H}_{kl}=\langle\phi_k|\hat{H}\hat{Y}^\dagger\hat{Y}|\phi_l\rangle$ and $\mathsf{S}_{kl}=\langle\phi_k|\hat{Y}^\dagger\hat{Y}|\phi_l\rangle$, respectively, and c is a K-component vector of the linear expansion coefficients c_k. The lowest energy solutions of Eq. (9) represent the ground state and the higher energy solutions represent excited states. All of them remain upper bounds to the corresponding exact energies regardless of any particular choice of the basis functions and the values of the linear coefficients.

The differential of the secular Eq. (9) is

$$\mathrm{d}(\mathsf{H}-\varepsilon\mathsf{S})\mathsf{c} = (\mathrm{d}\mathsf{H})\mathsf{c} - (\mathrm{d}\varepsilon)\mathsf{S}\mathsf{c} - \varepsilon(\mathrm{d}\mathsf{S})\mathsf{c} + (\mathsf{H}-\varepsilon\mathsf{S})\mathrm{d}\mathsf{c}. \tag{10}$$

Multiplying the above equation by $\mathsf{c}^\dagger$ from the left we obtain

$$\mathrm{d}\varepsilon = \mathsf{c}^\dagger(\mathrm{d}\mathsf{H}-\varepsilon\mathrm{d}\mathsf{S})\mathsf{c}. \tag{11}$$

To get Eq. (11) we used Eq. (9) and assumed that the wave function is normalized, $\mathsf{c}^\dagger\mathsf{S}\mathsf{c}=1$. Relation (11) is essentially the same as the well known Hellmann–Feynman theorem.

Let α_t be a vector of the nonlinear parameters the basis function φ_t depends on. With that we can now determine the complete energy gradient. As the tth row and tth column of matrices H and S depend on α_t, the derivative of any arbitrary element belonging to that row or that column of either of the two matrices can be written as

$$\frac{\partial \mathsf{H}_{kl}}{\partial\alpha_t} = \frac{\partial \mathsf{H}_{kl}}{\partial\alpha_t}(\delta_{kt}+\delta_{lt}-\delta_{kt}\delta_{lt}), \quad k,l,t=1,\ldots,K, \tag{12}$$

and

$$\frac{\partial \mathsf{S}_{kl}}{\partial\alpha_t} = \frac{\partial \mathsf{S}_{kl}}{\partial\alpha_t}(\delta_{kt}+\delta_{lt}-\delta_{kt}\delta_{lt}), \quad k,l,t=1,\ldots,K. \tag{13}$$

Further, using Eqs. (11)–(13), the derivative of the total energy, ε, with respect to the parameters α_t is

$$\begin{aligned}\frac{\partial\varepsilon}{\partial\alpha_t} &= \mathsf{c}_t^*\sum_{l=1}^{K}\mathsf{c}_l\left(\frac{\partial \mathsf{H}_{tl}}{\partial\alpha_t}-\varepsilon\frac{\partial \mathsf{S}_{tl}}{\partial\alpha_t}\right)+\mathsf{c}_t\sum_{l=1}^{K}\mathsf{c}_l^*\left(\frac{\partial \mathsf{H}_{lt}}{\partial\alpha_t}-\varepsilon\frac{\partial \mathsf{S}_{lt}}{\partial\alpha_t}\right)\\ &\quad-\mathsf{c}_t\mathsf{c}_t^*\left(\frac{\partial \mathsf{H}_{tt}}{\partial\alpha_t}-\varepsilon\frac{\partial \mathsf{S}_{tt}}{\partial\alpha_t}\right)\\ &= 2\Re\left[\mathsf{c}_t^*\sum_{l=1}^{K}\mathsf{c}_l\left(\frac{\partial \mathsf{H}_{tl}}{\partial\alpha_t}-\varepsilon\frac{\partial \mathsf{S}_{tl}}{\partial\alpha_t}\right)\right]-\mathsf{c}_t\mathsf{c}_t^*\left(\frac{\partial \mathsf{H}_{tt}}{\partial\alpha_t}-\varepsilon\frac{\partial \mathsf{S}_{tt}}{\partial\alpha_t}\right).\end{aligned} \tag{14}$$

By calculating all such derivatives for all $\alpha_k(k=1,\ldots,K)$ parameters, the complete energy gradient is obtained.

To make the calculations efficient it is best to evaluate all derivatives of ε with respect to the entire vech L_k vector in a single step rather than performing separate differentiation for individual parameters $(L_k)_{11},(L_k)_{21},\ldots,(L_k)_{nn}$ because many of the operations in calculating the derivatives are identical. With that, the calculation of Eq. (14) requires the following derivatives of the H and S matrix elements:

$$\frac{\partial \mathsf{H}_{kl}}{\partial(\mathrm{vech}\ L_k)}, \quad \frac{\partial \mathsf{H}_{kl}}{\partial(\mathrm{vech}\ L_l)}, \quad \frac{\partial \mathsf{S}_{kl}}{\partial(\mathrm{vech}\ L_k)}, \quad \frac{\partial \mathsf{S}_{kl}}{\partial(\mathrm{vech}\ L_l)}. \tag{15}$$

Now, we will derive the expressions for these derivatives.

D. Matrix elements

The overlap matrix element, S_{kl}, for Gaussian basis functions ϕ_k and ϕ_l is[16]

$$\mathsf{S}_{kl} = \langle \phi_k | \phi_l \rangle = \tfrac{1}{2}\pi^{3n/2}|A_{kl}|^{-3/2}\left\{\tfrac{1}{2}\eta_1\eta_2 + \eta_3\right\}, \tag{16}$$

where $\eta_1 = \mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_k]$, $\eta_2 = \mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_l]$, $\eta_3 = \mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l]$, and $A_{kl} = A_k + A_l$. Here and below vertical bars around a matrix are used to denote the determinant of the matrix.

For code efficiency, it is advantageous to determine if any of the traces involved in the formula for calculating an overlap (or Hamiltonian) matrix element are equal to zero. Having a zero trace can occur for two reasons: The first is due to the sum of the diagonal elements simply adding to zero; the second is due to all diagonal elements being zeros. The latter case can be easily verified and, when it occurs, appropriate simplifications can be made in the computer code to avoid calculating such a trace. The verification involves the analysis of the structure of the matrices whose product is used to calculate the trace.

For example, by using the sparsity of $\mathbf{W}_k$ [Eq. (7)], its product with the $\mathbf{A}_{kl}^{-1}$ matrix is found to be traceless. By explicitly implementing the information regarding the sparsity in coding the formulas for the matrix elements and their derivatives, one can effectively eliminate multiplications by zeros and make the calculations significantly faster. While only off-diagonal elements exist in $\mathbf{A}_{kl}^{-1}\mathbf{W}_k$ and $\mathbf{A}_{kl}^{-1}\mathbf{W}_l$, the product $\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l$ is not traceless. There exist sets of nonzero elements in $\mathbf{A}_{kl}^{-1}\mathbf{W}_k(i,k)$ and $\mathbf{A}_{kl}^{-1}\mathbf{W}_l(k,i)$ that contribute to the diagonal of $\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l(i,i)$. In this case the trace is not zero and has to be calculated.

By considering the above, the formula for the overlap integral (16) in our case simplifies to

$$\mathsf{S}_{kl} = \langle \phi_k | \phi_l \rangle = \tfrac{1}{2}\pi^{3n/2}|A_{kl}|^{-3/2}\eta_3. \tag{17}$$

This is the overlap integral formula used in calculating the energy gradient. It should be noted that this formula cannot be reduced or simplified any further.

The Hamiltonian matrix element, H_{kl}, consists of two terms, the kinetic energy term, T_{kl}, and the potential energy term, V_{kl}, where $\mathsf{H}_{kl} = \mathsf{T}_{kl} + \mathsf{V}_{kl}$. The formula for kinetic energy term is

$$\begin{aligned}\mathsf{T}_{kl} = \langle \phi_k | -\nabla_{\mathbf{r}}'\mathbf{M}\nabla_{\mathbf{r}}' | \phi_l \rangle = \pi^{3n/2}|A_{kl}|^{-3/2}\Big\{&\tfrac{1}{2}\eta_1\eta_2\,\mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l] + \eta_3\,\mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l] + \eta_1(\mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l] - \mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{W}_l]) \\ &+ \eta_2(\mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l] - \mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{M}\mathbf{A}_l]) + 2(\mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l] + \mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{M}\mathbf{W}_l] \\ &- \mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{W}_l] - \mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{M}\mathbf{A}_l] + \mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l])\Big\},\end{aligned} \tag{18}$$

where $\mathbf{M}$ is the mass matrix and $\mathbf{M} = M \otimes I_3$. In the M matrix, the diagonal elements are set to $1/(2m_1), 1/(2m_2), \ldots, 1/(2m_n)$, while the off-diagonal elements are set to $1/(2m_0)$. Again, m_0 is the mass of the nucleus and $m_1, \ldots$, and m_n are the electron masses. The terms containing $\eta_1 = \mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_k]$ and $\eta_2 = \mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_k]$ are again zero reducing the formula for the kinetic energy contribution to the Hamiltonian matrix element to

$$\begin{aligned}\mathsf{T}_{kl} &= \langle \phi_k | -\nabla_{\mathbf{r}}'\mathbf{M}\nabla_{\mathbf{r}}' | \phi_l \rangle \\ &= \pi^{3n/2}|A_{kl}|^{-3/2}\{\eta_3\,\mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l] \\ &\quad + 2(\mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l] + \mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{M}\mathbf{W}_l] \\ &\quad - \mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{W}_l] - \mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{M}\mathbf{A}_l] \\ &\quad + \mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l])\}.\end{aligned} \tag{19}$$

The potential energy matrix element is

$$\begin{aligned}\mathsf{V}_{kl} &= \langle \phi_k | \frac{1}{r_{ij}} | \phi_l \rangle = 2\pi^{(3n-1)/2}|A_{kl}|^{-3/2}\lambda^{-1/2} \\ &\quad \times \Bigg\{ \frac{1}{4}\eta_1\eta_2 + \frac{1}{2}\eta_3 - \frac{1}{12}\lambda^{-1}(\eta_1\,\mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{J}] \\ &\quad + \eta_2\,\mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{J}] + 2\,\mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{J}] \\ &\quad + 2\,\mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{J}]) \\ &\quad + \frac{1}{20}\lambda^{-2}(2\,\mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{J}\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{J}] \\ &\quad + \mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{J}]\mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{J}])\Bigg\},\end{aligned} \tag{20}$$

where $\mathbf{J}$ is a $3n \times 3n$ symmetric matrix, $\mathbf{J} = J \otimes I_3$, with the matrix J defined as

$$J = \begin{cases} E_{ii}, & i = j \ \text{ for } \ r_i \\ E_{ii} + E_{jj} - E_{ij} - E_{ji}, & i \neq j \ \text{ for } \ r_{ij}, \end{cases} \tag{21}$$

where E_{ij} is a matrix with one in the i,jth position and zeroes elsewhere, and $\lambda = \mathrm{tr}[A_{kl}^{-1}J]$. Applying again the same analysis of traces as in the case of the overlap integral, the formula for the potential energy matrix element reduces to

$$\begin{aligned}\langle \phi_k | \frac{1}{r_{ij}} | \phi_l \rangle &= 2\pi^{(3n-1)/2}|A_{kl}|^{-3/2}\lambda^{-1/2} \\ &\quad \times \Bigg\{ \frac{1}{2}\eta_3 - \frac{1}{6}\lambda^{-1}(\mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{J}] \\ &\quad + \mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{J}])\Bigg\}.\end{aligned} \tag{22}$$

Additionally the significant sparsity of the **J** matrix can be explicitly considered in writing the computer code. With that the number of operations used can be further reduced.

III. THE GRADIENT DERIVATION

Before differentiating the Hamiltonian matrix elements with respect to vech L_k, it is useful to write the differential of the A_k matrix,

$$\mathrm{d}A_k = (\mathrm{d}L_k)L_k' + L_k\mathrm{d}L_k', \tag{23}$$

and the differential of A_l,

$$\mathrm{d}A_l = P'(\mathrm{d}L_l)L_l'P + P'L_l\mathrm{d}L_l'P, \tag{24}$$

as these matrices contain the elements with which the derivatives are determined. The differential of A_{kl} is a sum of the above two differentials,

$$\mathrm{d}A_{kl} = (\mathrm{d}L_k)L_k' + L_k\mathrm{d}L_k' + P'(\mathrm{d}L_l)L_l'P + P'L_l(\mathrm{d}L_l')P. \tag{25}$$

This differential can be extended to determine the differential of $\mathbf{A}_{kl}$ as

$$\mathrm{d}\mathbf{A}_{kl} = \mathrm{d}\mathbf{L}_k\mathbf{L}_k' + \mathbf{L}_k\mathrm{d}\mathbf{L}_k' + \mathbf{P}'\mathrm{d}\mathbf{L}_l\mathbf{L}_l'\mathbf{P} + \mathbf{P}'\mathbf{L}_l\mathrm{d}\mathbf{L}_l'\mathbf{P} \equiv \mathrm{d}A_{kl}\otimes I_3. \tag{26}$$

All formulas for the S_{kl}, T_{kl}, and V_{kl} matrix elements contain the determinant of A_k and traces of matrix products involving $\mathbf{A}_{kl}^{-1}$. The differential of the determinant of an arbitrary matrix X, the differential of its inverse, and the differential of its trace are

$$\mathrm{d}|X| = |X|\mathrm{tr}[X^{-1}\mathrm{d}X], \tag{27}$$

$$\mathrm{d}X^{-1} = -X^{-1}(\mathrm{d}X)X^{-1}, \tag{28}$$

and

$$\mathrm{d}\,\mathrm{tr}[X] = \mathrm{tr}[\mathrm{d}X]. \tag{29}$$

To show the gradient derivation, the overlap integral formula (17) will be used, as it is the simplest of all three formulas. Then the approach used there will be extended to Eqs. (19) and (22). Using Eqs. (27)–(29), the differential of the overlap matrix element (17) can be written as

$$\mathrm{d}\mathsf{S}_{kl} = \mathrm{d}\langle\phi_k|\phi_l\rangle = -\tfrac{1}{2}\pi^{3n/2}|A_{kl}|^{-3/2}\left\{\tfrac{3}{2}\mathrm{tr}[A_{kl}^{-1}\mathrm{d}A_{kl}]\eta_3 + \mathrm{d}\eta_3\right\}, \tag{30}$$

where

$$\begin{aligned}\mathrm{d}\eta_3 &= \mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{A}_{kl}^{-1}\mathrm{d}\mathbf{A}_k\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l] + \mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathrm{d}\mathbf{A}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_l]\\ &= \mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathrm{d}\mathbf{A}_k] + \mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathrm{d}\mathbf{A}_l].\end{aligned} \tag{31}$$

As it is seen in the derivation of Eq. (31) the property of the trace of a product of matrices was utilized. This property allows a cyclic permutation of the matrices to be applied and move the $\mathrm{d}\mathbf{A}_k$ and $\mathrm{d}\mathbf{A}_l$ differentials to the rightmost positions in the respective matrix products. This is possible because a cyclic permutation of matrices in the product under the trance does not change the trace, i.e., $\mathrm{tr}[XY]=\mathrm{tr}[YX]$, or more generally, $\mathrm{tr}[XYZ,\ldots,PQ]=\mathrm{tr}[YZ,\ldots,PQX]=\mathrm{tr}[Z,\ldots,PQXY]=\cdots$.

Plugging in Eq. (26) to Eq. (31),

$$\begin{aligned}\mathrm{d}\eta_3 &= \mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}(\mathrm{d}\mathbf{L}_k\mathbf{L}_k' + \mathbf{L}_k\mathrm{d}\mathbf{L}_k' + \mathrm{d}\mathbf{L}_l\mathbf{L}_l' + \mathbf{L}_l\mathrm{d}\mathbf{L}_l')] + \mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}(\mathrm{d}\mathbf{L}_k\mathbf{L}_k' + \mathbf{L}_k\mathrm{d}\mathbf{L}_k' + \mathrm{d}\mathbf{L}_l\mathbf{L}_l' + \mathbf{L}_l\mathrm{d}\mathbf{L}_l')]\\ &= \mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathrm{d}\mathbf{L}_k\mathbf{L}_k'] + \mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{L}_k\mathrm{d}\mathbf{L}_k'] + \mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathrm{d}\mathbf{L}_l\mathbf{L}_l'] + \mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{L}_l\mathrm{d}\mathbf{L}_l']\end{aligned} \tag{32}$$

is obtained by using

$$\mathrm{tr}[A+B] = \mathrm{tr}[A] + \mathrm{tr}[B]. \tag{33}$$

To combine the $\mathrm{d}L_k'$ term with the $\mathrm{d}L_k$ term and the $\mathrm{d}\mathbf{L}_k'$ term with the $\mathrm{d}\mathbf{L}_k$ the following property of the trace involving matrix transposition is applied:

$$\mathrm{tr}[XYZ] = \mathrm{tr}[Z'Y'X']. \tag{34}$$

Using relation (34) in Eq. (32) and transforming $\mathrm{tr}[A_{kl}^{-1}\mathrm{d}A_{kl}]$, also present in the overlap integral formula (30), in the same way $\mathrm{d}\eta_3$ was transformed we obtain

$$\begin{aligned}\mathrm{d}\mathsf{S}_{kl} = \mathrm{d}\langle\phi_k|\phi_l\rangle = -\tfrac{1}{2}\pi^{3n/2}|A_{kl}|^{-3/2}\Big\{&\tfrac{3}{2}(\mathrm{tr}[L_k'(A_{kl}^{-1}+A_{kl}^{-1\prime})\mathrm{d}L_k] + \mathrm{tr}[L_l'(A_{kl}^{-1}+A_{kl}^{-1\prime})\mathrm{d}L_l])\eta_3 + \mathrm{tr}[\mathbf{L}_k'(\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\\ &+ (\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1})')\mathrm{d}\mathbf{L}_k] + \mathrm{tr}[\mathbf{L}_l'(\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1})')\mathrm{d}\mathbf{L}_l] + \mathrm{tr}[\mathbf{L}_k'(\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\\ &+ (\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1})')\mathrm{d}\mathbf{L}_k] + \mathrm{tr}[\mathbf{L}_l'(\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1})')\mathrm{d}\mathbf{L}_l]\Big\}.\end{aligned} \tag{35}$$

Let us now list some useful properties of the vec and vech operators. If X and Y are arbitrary square matrices and L is a lower triangular matrix, then

$$(\mathrm{vec}\ X)'\mathrm{vec}\ Y = \mathrm{tr}[X'Y], \tag{36}$$

$$(\mathrm{vec}\ X)'\mathrm{vec}\ L = (\mathrm{vech}\ X)'\mathrm{vech}\ L. \tag{37}$$

Both vec and vech operations transform a matrix into a vector. vec stacks the columns of a matrix one underneath the other. Thus, it transforms an $n\times n$ matrix into an n^2-component vector. For example, if X is a 2×2 matrix with elements X_{ij}, then vec X is the following four-component vector:

$$\mathrm{vec}\ X = \begin{pmatrix} X_{11} \\ X_{12} \\ X_{21} \\ X_{22} \end{pmatrix}. \tag{38}$$

As described before, the second operator, vech, elements transform an $n\times n$ matrix into an $n(n+1)/2$-component vector. For example, if X is a 3×3 matrix with elements X_{ij}, then vech X is the following six-component vector:

$$\mathrm{vech}\ X = \begin{pmatrix} X_{11} \\ X_{12} \\ X_{13} \\ X_{22} \\ X_{23} \\ X_{33} \end{pmatrix}. \tag{39}$$

Using the vech operator, Eq. (30) can be written as

$$\begin{aligned} \mathrm{d}\langle\phi_k|\phi_l\rangle = \tfrac{1}{2}\pi^{3n/2}|A_{kl}|^{-3/2}\{&(\mathrm{vech}((A_{kl}^{-1}+A_{kl}^{-1'})L_k)'\mathrm{d}\ \mathrm{vech}\ L_k + \mathrm{vech}((A_{kl}^{-1}+A_{kl}^{-1'})L_l)'\mathrm{d}\ \mathrm{vech}\ L_l)\eta_3 + \mathrm{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1} \\ &+ (\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1})')\mathbf{L}_k)'\mathrm{d}\ \mathrm{vech}\ \mathbf{L}_k + \mathrm{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1})')\mathbf{L}_l)'\mathrm{d}\ \mathrm{vech}\ \mathbf{L}_l \\ &+ \mathrm{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1})')\mathbf{L}_k)'\mathrm{d}\ \mathrm{vech}\ \mathbf{L}_k + \mathrm{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1} \\ &+ (\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1})')\mathbf{L}_l)'\mathrm{d}\ \mathrm{vech}\ \mathbf{L}_l\}. \end{aligned} \tag{40}$$

Equation (40) contains derivatives with respect to the vech $\mathbf{L}_k$ and vech $\mathbf{L}_l$ vectors. However, it is computationally advantageous that the derivatives are taken with respect to vech L_k and vech L_l. To transform the expression to the correct form, a transformation matrix, $\mathcal{T}$, must be applied to the appropriate terms. This transformation matrix has the dimension of $[3n(3n+1)/2]\times[n(n+1)/2]$ and is defined as

$$\mathcal{T} = \frac{\mathrm{d}\ \mathrm{vech}\ \mathbf{L}_k(\mathrm{vech}\ L_k)}{\mathrm{d}(\mathrm{vech}\ L_k)'}. \tag{41}$$

The notations vech $\mathbf{L}_k$(vech L_k) indicate that the $3n(3n+1)/2$-dimensional vector vech $\mathbf{L}_k$ is a function of the $n(n+1)/2$-dimensional vector vech L_k. The derivative of vech $\mathbf{L}_k$ with respect to vech L_k is defined as the $3n(3n+1)/2\times n(n+1)/2$ matrix of partial derivatives whose ijth element is the partial derivative of the ith component of vech $\mathbf{L}_k$ (a column vector) with respect to the jth element of (vech L_k)$'$ (a row vector). It should be noted that $\mathcal{T}$ is independent of index k and is a matrix consisting of zeroes and ones. Equation (41) can be rearranged to the following form:

$$\mathrm{d}\ \mathrm{vech}\ \mathbf{L}_k = \mathcal{T}\mathrm{d}\ \mathrm{vech}\ L_k, \tag{42}$$

and a similar form for d vech $\mathbf{L}_l$. Substituting d vech $\mathbf{L}_k$ and d vech $\mathbf{L}_l$ in Eq. (40) with Eq. (42) and grouping terms with vech L_k separately from vech L_l terms the following two derivatives can be written:

$$\frac{\partial\langle\phi_k|\phi_l\rangle}{\partial\ \mathrm{vech}\ L_k} = -\frac{1}{2}\pi^{3n/2}|A_{kl}^{-1}|^{3/2}\Bigg\{\frac{3}{2}\mathrm{vech}((A_{kl}^{-1}+A_{kl}^{-1'})L_k)\mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l] + \mathrm{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1})')\mathbf{L}_k)\mathcal{T} + \mathrm{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1})')\mathbf{L}_k)\mathcal{T}\Bigg\}, \tag{43}$$

$$\frac{\partial\langle\phi_k|\phi_l\rangle}{\partial\ \mathrm{vech}\ L_l} = -\frac{1}{2}\pi^{3n/2}|A_{kl}^{-1}|^{3/2}\Bigg\{\frac{3}{2}\mathrm{vech}((A_{kl}^{-1}+A_{kl}^{-1'})L_l)\mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l] + \mathrm{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1})')\mathbf{L}_l)\mathcal{T} + \mathrm{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1})')\mathbf{L}_l)\mathcal{T}\Bigg\}. \tag{44}$$

Applying the same operations as used to derive the derivatives for the overlap matrix element to the matrix elements for kinetic and potential energies, T_{kl} and V_{kl}, their derivatives can be expressed. Nothing additional is required other than meticulousness. These derivatives are

$$\begin{aligned}\frac{\partial\mathsf{T}_{kl}}{\partial\ \mathrm{vech}\ L_k} &= \frac{\partial\langle\phi_k| - \nabla_{\mathbf{r}}'\mathbf{M}\nabla_{\mathbf{r}}'|\phi_l\rangle}{\partial\ \mathrm{vech}(L_k)}\\ &= -\pi^{3n/2}|A_{kl}|^{-3/2}\Bigg\{\frac{3}{2}\mathrm{vech}((A_{kl}^{-1}+A_{kl}^{-1'})L_k)(\eta_3\ \mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l] + 2(\mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l]\\ &+ \mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{M}\mathbf{W}_l] + \mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{W}_l] + \mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{M}\mathbf{A}_l] + \mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l]))\\ &+ \mathrm{tr}[\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l](\mathrm{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1})')\mathbf{L}_k) + \mathrm{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\\ &+ (\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1})')\mathbf{L}_k))\mathcal{T} + \eta_3(\mathrm{vech}((\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1})')\mathbf{L}_k) - \mathrm{vech}((\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1}\\ &+ (\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1})')\ \mathbf{L}_k))\mathcal{T} + 2(\mathrm{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1})')\mathbf{L}_k)\\ &+ \mathrm{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1})')\mathbf{L}_k)\\ &+ \mathrm{vech}((\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1})')\mathbf{L}_k) - \mathrm{vech}((\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\\ &+ (\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1})')\mathbf{L}_k) + \mathrm{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1})')\mathbf{L}_k)\\ &+ \mathrm{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1})')\mathbf{L}_k)\\ &+ \mathrm{vech}((\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1})')\mathbf{L}_k) - \mathrm{vech}((\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\\ &+ (\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1})')\mathbf{L}_k) + \mathrm{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{W}_l\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{W}_l\mathbf{A}_{kl}^{-1})')\mathbf{L}_k)\\ &+ \mathrm{vech}((\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1})')\mathbf{L}_k) - \mathrm{vech}((\mathbf{M}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\\ &+ (\mathbf{M}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1})')\mathbf{L}_k) + \mathrm{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1})')\mathbf{L}_k)\\ &+ \mathrm{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1})')\mathbf{L}_k) + \mathrm{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{M}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\\ &+ (\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{M}\mathbf{W}_l\mathbf{A}_{kl}^{-1})')\mathbf{L}_k))\mathcal{T}\Bigg\}\end{aligned} \tag{45}$$

and

$$\frac{\partial \mathsf{T}_{kl}}{\partial \text{ vech } L_l} = \frac{\partial \langle \phi_k| - \nabla'_{\mathbf{r}}\mathbf{M}\nabla'_{\mathbf{r}}|\phi_l\rangle}{\partial \text{ vech } L_l}$$

$$= -\pi^{3n/2}|A_{kl}|^{-3/2}\Bigg\{\frac{3}{2}\text{vech}((A_{kl}^{-1} + A_{kl}^{-1\prime})L_l)(\eta_3 \text{ tr}[\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l] + 2(\text{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l]$$
$$+ \text{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{M}\mathbf{W}_l] + \text{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{W}_l] + \text{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{M}\mathbf{A}_l] + \text{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l]))$$
$$+ \text{tr}[\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l](\text{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1})')\mathbf{L}_l) + \text{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}$$
$$+ (\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1})')\mathbf{L}_l))\mathcal{T} + \eta_3(\text{vech}((\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1})')\mathbf{L}_l) - \text{vech}((\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}$$
$$+ (\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M})')\mathbf{L}_l))\mathcal{T} + 2(\text{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1})')\mathbf{L}_l)$$
$$+ \text{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1})')\mathbf{L}_l)$$
$$+ \text{vech}((\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1})')\mathbf{L}_l) - \text{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}$$
$$+ (\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M})')\mathbf{L}_l) + \text{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1})')\mathbf{L}_l)$$
$$+ \text{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1})')\mathbf{L}_l)$$
$$+ \text{vech}((\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1})')\mathbf{L}_l) - \text{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}$$
$$+ (\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M})')\mathbf{L}_l) + \text{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{W}_l\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{W}_l\mathbf{A}_{kl}^{-1})')\mathbf{L}_l)$$
$$+ \text{vech}((\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{A}_k\mathbf{M}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1})')\mathbf{L}_l) + \text{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1}$$
$$+ (\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1})')\mathbf{L}_l) + \text{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{M}\mathbf{A}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1})')\mathbf{L}_l)$$
$$- \text{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{M} + (\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{M})')\mathbf{L}_l) + \text{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{M}\mathbf{W}_l\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{M}\mathbf{W}_l\mathbf{A}_{kl}^{-1})')\mathbf{L}_l))\mathcal{T}\Bigg\}. \tag{46}$$

The derivative of the potential energy matrix elements is determined to be

$$\frac{\partial \mathsf{V}_{kl}}{\partial \text{ vech } L_k} = \frac{\partial \langle \phi_k|\frac{1}{r_{ij}}|\phi_l\rangle}{\partial \text{ vech } L_k}$$

$$= 2\pi^{(3n-1)/2}|A_{kl}|^{-3/2}\lambda^{-1/2}\Bigg\{\left(\frac{1}{2}\lambda^{-1} \text{ vech}((A_{kl}^{-1}JA_{kl}^{-1} + (A_{kl}^{-1}JA_{kl}^{-1})')L_k) - \frac{3}{2}\text{vech}((A_{kl}^{-1} + A_{kl}^{-1\prime})L_k)\right)$$
$$\times\left(\frac{1}{2}\text{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l] - \frac{1}{6}\lambda^{-1}(\text{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{J}] + \text{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{J}])\right)$$
$$+ \frac{1}{2}(\text{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1})')\mathbf{L}_k) + \text{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}$$
$$+ (\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1})')\mathbf{L}_k))\mathcal{T} + \frac{1}{6}\lambda^{-1}(\text{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{J}\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{J}\mathbf{A}_{kl}^{-1})')\mathbf{L}_k)$$
$$+ \text{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{J}\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{J}\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1})')\mathbf{L}_k) + \text{vech}((\mathbf{A}_{kl}^{-1}\mathbf{J}\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}$$
$$+ (\mathbf{A}_{kl}^{-1}\mathbf{J}\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1})')\mathbf{L}_k) + \text{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{J}\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{J}\mathbf{A}_{kl}^{-1})')\mathbf{L}_k)$$
$$+ \text{vech}((\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{J}\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{J}\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1})')\mathbf{L}_k) + \text{vech}((\mathbf{A}_{kl}^{-1}\mathbf{J}\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}$$
$$+ (\mathbf{A}_{kl}^{-1}\mathbf{J}\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1})')\mathbf{L}_k))\mathcal{T} - \frac{1}{6}\lambda^{-2} \text{ vech}((A_{kl}^{-1}JA_{kl}^{-1} + (A_{kl}^{-1}JA_{kl}^{-1})')L_k)(\text{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{J}]$$
$$+ \text{tr}[\mathbf{A}_{kl}^{-1}\mathbf{W}_l\mathbf{A}_{kl}^{-1}\mathbf{W}_k\mathbf{A}_{kl}^{-1}\mathbf{J}])\Bigg\}, \tag{47}$$

$$\frac{\partial \mathbf{V}_{kl}}{\partial \text{ vech } L_l} = \frac{\partial \langle \phi_k | \frac{1}{r_{ij}} | \phi_l \rangle}{\partial \text{ vech } L_l}$$

$$= 2\pi^{(3n-1)/2} |A_{kl}|^{-3/2} \lambda^{-1/2} \left\{ \left(\frac{1}{2}\lambda^{-1} \text{ vech}((A_{kl}^{-1} J A_{kl}^{-1} + (A_{kl}^{-1} J A_{kl}^{-1})') L_l) - \frac{3}{2} \text{vech}((A_{kl}^{-1} + A_{kl}^{-1\prime}) L_l) \right) \right.$$

$$\times \left(\frac{1}{2} \text{tr}[\mathbf{A}_{kl}^{-1} \mathbf{W}_k \mathbf{A}_{kl}^{-1} \mathbf{W}_l] - \frac{1}{6}\lambda^{-1} (\text{tr}[\mathbf{A}_{kl}^{-1} \mathbf{W}_k \mathbf{A}_{kl}^{-1} \mathbf{W}_l \mathbf{A}_{kl}^{-1} \mathbf{J}] + \text{tr}[\mathbf{A}_{kl}^{-1} \mathbf{W}_l \mathbf{A}_{kl}^{-1} \mathbf{W}_k \mathbf{A}_{kl}^{-1} \mathbf{J}]) \right) + \frac{1}{2} (\text{vech}((\mathbf{A}_{kl}^{-1} \mathbf{W}_k \mathbf{A}_{kl}^{-1} \mathbf{W}_l \mathbf{A}_{kl}^{-1}$$

$$+ (\mathbf{A}_{kl}^{-1} \mathbf{W}_k \mathbf{A}_{kl}^{-1} \mathbf{W}_l \mathbf{A}_{kl}^{-1})') \mathbf{L}_l) + \text{vech}((\mathbf{A}_{kl}^{-1} \mathbf{W}_l \mathbf{A}_{kl}^{-1} \mathbf{W}_k \mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1} \mathbf{W}_l \mathbf{A}_{kl}^{-1} \mathbf{W}_k \mathbf{A}_{kl}^{-1})') \mathbf{L}_l)) \mathcal{T}$$

$$+ \frac{1}{6}\lambda^{-1} (\text{vech}((\mathbf{A}_{kl}^{-1} \mathbf{W}_k \mathbf{A}_{kl}^{-1} \mathbf{W}_l \mathbf{A}_{kl}^{-1} \mathbf{J} \mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1} \mathbf{W}_k \mathbf{A}_{kl}^{-1} \mathbf{W}_l \mathbf{A}_{kl}^{-1} \mathbf{J} \mathbf{A}_{kl}^{-1})') \mathbf{L}_l) + \text{vech}((\mathbf{A}_{kl}^{-1} \mathbf{W}_l \mathbf{A}_{kl}^{-1} \mathbf{J} \mathbf{A}_{kl}^{-1} \mathbf{W}_k \mathbf{A}_{kl}^{-1}$$

$$+ (\mathbf{A}_{kl}^{-1} \mathbf{W}_l \mathbf{A}_{kl}^{-1} \mathbf{J} \mathbf{A}_{kl}^{-1} \mathbf{W}_k \mathbf{A}_{kl}^{-1})') \mathbf{L}_l) + \text{vech}((\mathbf{A}_{kl}^{-1} \mathbf{J} \mathbf{A}_{kl}^{-1} \mathbf{W}_k \mathbf{A}_{kl}^{-1} \mathbf{W}_l \mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1} \mathbf{J} \mathbf{A}_{kl}^{-1} \mathbf{W}_k \mathbf{A}_{kl}^{-1} \mathbf{W}_l \mathbf{A}_{kl}^{-1})') \mathbf{L}_l)$$

$$+ \text{vech}((\mathbf{A}_{kl}^{-1} \mathbf{J} \mathbf{A}_{kl}^{-1} \mathbf{W}_k \mathbf{A}_{kl}^{-1} \mathbf{W}_l \mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1} \mathbf{J} \mathbf{A}_{kl}^{-1} \mathbf{W}_k \mathbf{A}_{kl}^{-1} \mathbf{W}_l \mathbf{A}_{kl}^{-1})') \mathbf{L}_l) + \text{vech}((\mathbf{A}_{kl}^{-1} \mathbf{W}_l \mathbf{A}_{kl}^{-1} \mathbf{W}_k \mathbf{A}_{kl}^{-1} \mathbf{J} \mathbf{A}_{kl}^{-1}$$

$$+ (\mathbf{A}_{kl}^{-1} \mathbf{W}_l \mathbf{A}_{kl}^{-1} \mathbf{W}_k \mathbf{A}_{kl}^{-1} \mathbf{J} \mathbf{A}_{kl}^{-1})') \mathbf{L}_l) + \text{vech}((\mathbf{A}_{kl}^{-1} \mathbf{W}_k \mathbf{A}_{kl}^{-1} \mathbf{J} \mathbf{A}_{kl}^{-1} \mathbf{W}_l \mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1} \mathbf{W}_k \mathbf{A}_{kl}^{-1} \mathbf{J} \mathbf{A}_{kl}^{-1} \mathbf{W}_l \mathbf{A}_{kl}^{-1})') \mathbf{L}_l)$$

$$+ \text{vech}((\mathbf{A}_{kl}^{-1} \mathbf{J} \mathbf{A}_{kl}^{-1} \mathbf{W}_l \mathbf{A}_{kl}^{-1} \mathbf{W}_k \mathbf{A}_{kl}^{-1} + (\mathbf{A}_{kl}^{-1} \mathbf{J} \mathbf{A}_{kl}^{-1} \mathbf{W}_l \mathbf{A}_{kl}^{-1} \mathbf{W}_k \mathbf{A}_{kl}^{-1})') \mathbf{L}_l)) \mathcal{T} - \frac{1}{6}\lambda^{-2} \text{ vech}((A_{kl}^{-1} J A_{kl}^{-1} + (A_{kl}^{-1} J A_{kl}^{-1})') L_l)$$

$$\left. \times (\text{tr}[\mathbf{A}_{kl}^{-1} \mathbf{W}_k \mathbf{A}_{kl}^{-1} \mathbf{W}_l \mathbf{A}_{kl}^{-1} \mathbf{J}] + \text{tr}[\mathbf{A}_{kl}^{-1} \mathbf{W}_l \mathbf{A}_{kl}^{-1} \mathbf{W}_k \mathbf{A}_{kl}^{-1} \mathbf{J}]) \right\}. \quad (48)$$

IV. NUMERICAL ILLUSTRATION

The carbon in the ground ^3P state has four *s*- and two *p*-electrons (the electron configuration: $1s^2 2s^2 2p^2$). As this is the smallest atom with two *p*-electrons in the ground electronic state, it is a good model system to test the procedure for calculating the energy gradient described in this work. The calculations performed in this work have concerned, apart from the ground state, also the first excited ^3P state of this system and the corresponding transition energy with respect to the ground state. This energy is very well established experimentally as 71 352.51 cm^{-1} (the transition between the $J=0$ sublevels).[29]

Before the calculations for the carbon atom were started, the gradient procedure was tested for correctness. In the testing we first compared the values of the derivatives of individual Hamiltonian and overlap matrix elements determined with respect to the elements of the L_k matrix with the derivatives calculated using the numerical differentiation. After this test was successfully completed, the derivatives of the total energy were checked also against the values obtained using the numerical differentiation.

The first test calculation for the carbon atom was performed for the ground state of the ^{12}C isotope with only ten basis functions. For the initial values for the nonlinear parameters of these functions we used the orbital exponents taken from the standard 3-21 G orbital Gaussian basis set. Squares of these exponents were used as the diagonal elements of the L_k matrices in the initial guesses for the basis functions and the off-diagonal elements of these matrices were set to zero. Next a gradient-based optimization was performed where the values of all L_k matrix elements of the ten basis functions were simultaneously optimized. This and the other optimizations performed in this work have been done with the Broyden–Fletcher–Goldfarb–Shanno (BFGS) algorithm. This algorithm is usually somewhat more efficient for small and medium size optimization problems than the conjugate gradient and limited-memory BFGS algorithms. The full BFGS algorithm requires storing the complete Hessian (or more precisely, the Cholesky factor of it), which becomes prohibitively expensive for problems with many thousands of variables. However, for the calculations performed in this work the storage requirements for the Hessian update are moderate and the BFGS algorithm works efficiently.

In Table I we show how the total energy and the norm of the energy gradient were changing in that optimization as functions of the iteration number (only values for interaction numbers being whole multiples of 200 are shown). As one can see, it took the optimization procedure 1677 iterations to

TABLE I. The convergence of the total ^{12}C atom energy and the gradient norm in a global optimization calculation with ten ECGFs in the basis set. The energy is in hartrees.

Iteration No.	Energy	Gradient norm
0	−34.707 161 49	0.46×10^{-2}
200	−36.621 903 54	0.30×10^{-4}
400	−36.782 963 59	0.21×10^{-5}
600	−36.893 867 67	0.21×10^{-5}
800	−36.913 420 84	0.15×10^{-5}
1000	−36.919 845 57	0.13×10^{-6}
1200	−36.922 242 62	0.52×10^{-6}
1400	−36.923 386 06	0.30×10^{-7}
1600	−36.923 439 51	0.51×10^{-11}
1677	−36.923 439 52	0.38×10^{-15}

lower the gradient norm from about 10^{-2} to 10^{-15} hartree bohr. When the gradient norm approached 10^{-15} the variational energy in the given basis was converged to ten significant figures. This type of optimization progress is not unusual when a small basis set is optimized. For larger basis set the progress of the optimization is usually much slower due to a much larger number of the parameters optimized (in the carbon calculations there are 21 parameters per Gaussian; 10×21 total in the calculation with ten basis functions) and due to a stronger interdependency between the basis functions. The ten-function test showed to us that the gradient procedure works correctly and, as expected, the lowering of the gradient norm is accompanied by a systematic convergence of the energy.

A basis set of ten functions is also used to show the advantage of the basis set optimization that involves the gradient versus the optimization without the gradient. Both optimizations in the test have been performed for the ground state of the ^{12}C atom and both were started using the same initial ten-function basis set. To quantity, which is compared, is the total energy obtained after running the optimization for one CPU hour. In the optimization that does not utilize the gradient, this amount of CPU time allowed to lower the total energy from the initial −33.505 631 97 to −35.132 739 46 hartree. In the optimization with the gradient, during the same amount of time the energy was lowered to −35.926 912 85 hartree clearly showing the advantage of utilizing the gradient in the variational energy minimization.

The next test was performed for the two ^3P carbon states mentioned above. Both states, i.e., the $2s^22p^2\ ^3P$ ground state considered in the first test and the $2s^22p3p\ ^3P$ state have the same spatial symmetry. For each of the two states the basis set was grown to the size of 500 functions by incremental additions of sets of 20 (in the early stage of the optimization) or 50 (when the number of functions exceeded 200) functions. In generating the new functions we used the functions already included in the set. By randomly perturbing the exponential parameters of these already included functions and selecting those from the modified function set that contribute the most to the energy the new subset of functions to be added to the basis set was generated. The new functions were subsequently optimized using the gradient-based procedure. Two types of optimizations have been performed. When the basis set contained less than 80 functions, we used a "global" optimization approach where we simultaneously optimized all functions in the basis set. When the basis set exceeded 80 functions, the optimization involved tuning the parameters of only one function at the time and cycling over the whole basis set several times. We found this type of strategy to be more time efficient and robust for larger basis sets than the simultaneous optimization of all nonlinear parameters.

We should mention that besides the L_k parameters, there are two integer parameters, i_k and j_k, in each Gaussian (6) that can be optimized. Optimization of these parameters usually has a small effect on the energy, but for the sake of completeness of the basis set, it needs to be performed especially if very accurate energy is sought for. In this work we only optimized the value of the second parameter, j_k, setting the first to be equal to one in all basis functions ($i_k=1$). This optimization was only done once for each function just after the function was added to the basis set.

TABLE II. The convergence of the total nonrelativistic energies of the ground and first excited ^3P states ($2s^22p^2\ ^3P$ and $2s^22p3p\ ^3P$) of ^{12}C and $^{\infty}$C and the corresponding transition energy with the number of basis functions. Total energies are in hartrees and transition energies are in cm^{-1}.

Basis size	$2s^22p^2\ ^3P$	$2s^22p3p\ ^3P$	Transition energy
		^{12}C	
100	−37.810 500 84	−37.471 761 87	74 344.61
200	−37.828 749 14	−37.499 145 61	72 339.61
300	−37.834 299 63	−37.506 664 64	71 907.57
400	−37.837 007 63	−37.509 749 61	71 824.83
500	−37.838 416 07	−37.511 548 21	71 739.20
Experiment[a]			71 352.51
		$^{\infty}$C	
500	−37.840 128 79	−37.513 262 95	71 738.76

[a]Reference 29.

In Table II we show how the total energies of the two ^3P states of ^{12}C converge with the number of basis functions. Results obtained using the basis sets ranging from 100 functions to 500 functions in increments of 100 are shown. We also show the convergence of the transition energy. As one can see, 500 functions are definitely not enough to achieve a high level of convergence for the energy of either of the two states. Thousands of Gaussians are needed to achieve the spectroscopic accuracy in this case. However the results obtained with basis sets with sizes up to 500 functions quite adequately illustrate that the gradient procedure works efficiently in lowering the energies of both states. We should add that the number of the optimization cycles for smaller basis sets was larger than for the larger sets (only three for the set with 500 functions). This was related to the amount of the computer time we were able to allocate for this project. This resulted in the energies for the smaller sets to be better optimized than the energies obtained with the larger sets. As the energy of the upper state ($2s^22p3p\ ^3P$) requires more optimization cycles than the energy of the ground state to be converged to a similar level, the transition energies obtained with the basis sets containing 400 and 500 basis functions are not as low as they would be, if more cycles were performed. In Table II we also show the energies of the two states calculated with 500 functions for $^{\infty}$C. This was done by setting the mass of the nucleus to infinity. In those calculations we used the ^{12}C basis functions without reoptimizing their exponential parameters (only the linear parameters of basis functions were recomputed).

The ground state energy converging faster than the energy of the excited state is due to more complicated nature of the spatial wave function of latter state than the former. The difference in the convergence patterns causes the transition energy calculated as the difference between the total energies of the two states to be an upper bound to the exact value. Thus, as the number of basis functions increases the transition energy decreases. The target value of the experimental transition energy of 71 352.51 cm^{-1} is still off by less than

400 cm^{-1} at the basis set size of 500 (where the result is 71 739.20 cm^{-1}), but the energy seems to be converging to the right value. We should also mention that the contribution of the relativistic and quantum electrodynamics effects (not considered in the present work) to the $2s^22p^2 \rightarrow 2s^22p3p$ transition energy may be quite significant—of the order of 10^2 cm^{-1}.

As for the total energy of the ground state, our lowest result obtained with 500 Gaussians for $^{\infty}$C is −37.840 128 79 hartree. This value is significantly lower than the previous best variational ground state upper bound of −37.792 860 hartree obtained by Sundholm and Olsen[30] in their finite-element multi-configuration Hartree-Fock calculations with a large atomic orbital basis set. In our calculations the Sundholm and Olsen energy value was already reached before the basis set size was increased to 80 ECGFs. Our result can also be compared to the complete-basis-set limit for the carbon ground state energy estimated by Woon and Dunning[31] to be −37.838 16 hartree based on their configuration interaction results obtained with the aug-cc-pCVXZ type basis sets. The comparison suggests that Woon and Dunning slightly underestimated the magnitude of the correlation energy contribution.

We should note that, in general, the results of a multiparameter variational optimization can be dependent on the initial guess for the basis functions used to start the growing of the basis set. However, as the growing progresses and the basis set becomes more complete, the dependency of the results on the initial guess should decrease to eventually completely disappear at the limit of completeness. Another problem, which seems to also become less severe as the basis set becomes larger, are linear dependencies among ECGFs. This problem is addressed in the calculations by examining each new basis function before and after the function is optimized and added to the basis set for possible linear dependency with all other functions already included in the basis set. If a linear dependency appears, the new function is discarded and a new function is generated. This process continues until a linearly independent function is found.

V. SUMMARY

In this work we presented formulas for calculating the analytical derivatives of the total energy of an atom with two *p*-electrons with respect to the exponential parameters of explicitly correlated all-electron Gaussian functions used to expand the wave function of the system. The derivatives, which are collected in the gradient vector, have been employed in the variational calculations concerning the ground and the first excited ^3P state of the carbon atom. As only 500 Gaussians have been used in the calculations, the results concerning the total energies of the two states, as well as the transition energy, are not converged to the spectroscopic accuracy. However they clearly demonstrate the advantage of using the analytical gradient in the variational energy minimization. Even with only 500 Gaussians in the basis set, our lowest energy of the carbon ground state is significantly lower that the best literature result.

The present calculations have been carried out on a multiprocessor computer system using the MPI protocol. There are two ways this project will further evolve in the future. The first involves redesigning the procedure for the gradient-based energy minimization procedure so it can be executed with high efficiency using hundreds or even thousands of processors. With that calculations for the carbon atom and for other systems of similar size can be performed using large basis sets capable of producing results with the spectroscopic accuracy. The second way involves calculations of excited states of four- and five-electron atomic systems. Those states can either be states with two *p*-electrons or states with one *d*-electron, as the present approach can easily be extended to those latter types of states.

ACKNOWLEDGMENTS

This work was supported by the National Science Foundation. We are grateful to the University of Arizona Information Technology Services (UITS) for the use of their computer resources.

[1] G. W. F. Drake, Can. J. Phys. **80**, 1195 (2002).
[2] Z. C. Yan and G. W. F. Drake, Phys. Rev. Lett. **81**, 774 (1998).
[3] B. A. Bushaw, W. Nortershauser, G. Ewald, A. Dax, and G. W. F. Drake, Phys. Rev. Lett. **91**, 043004 (2003).
[4] K. Pachucki, Phys. Rev. A **74**, 062510 (2006).
[5] K. Pachucki and V. Yerokhin, Phys. Rev. A **79**, 062516 (2009).
[6] Z.-C. Yan, W. Nortershauser, and G. Drake, Phys. Rev. Lett. **100**, 243002 (2008).
[7] M. Puchalski and K. Pachucki, Phys. Rev. A **78**, 052511 (2008).
[8] M. Puchalski and K. Pachucki, Phys. Rev. A **79**, 032510 (2009).
[9] K. Pachucki and J. Komasa, Phys. Rev. A **73**, 052502 (2006).
[10] K. Pachucki and J. Komasa, J. Chem. Phys. **125**, 204304 (2006).
[11] M. Stanke D. Kędziera, S. Bubin, and L. Adamowicz, Phys. Rev. Lett. **99**, 043001 (2007).
[12] M. Stanke D. Kędziera, S. Bubin, and L. Adamowicz, Phys. Rev. A **75**, 052510 (2007).
[13] M. Stanke, J. Komasa, S. Bubin, and L. Adamowicz, Phys. Rev. A **80**, 022514 (2009).
[14] S. Bubin, J. Komasa, M. Stanke, and L. Adamowicz, J. Chem. Phys. **131**, 234112 (2009).
[15] S. Bubin and L. Adamowicz, J. Chem. Phys. **128**, 114107 (2008).
[16] K. Sharkey, M. Pavanello, S. Bubin, and L. Adamowicz, " Phys. Rev. A **80**, 062510 (2009).
[17] P. Kozlowski and L. Adamowicz, J. Chem. Phys. **96**, 9013 (1992).
[18] P. Kozlowski and L. Adamowicz, J. Chem. Phys. **97**, 5063 (1992).
[19] D. Kinghorn, Int. J. Quantum Chem. **57**, 141 (1996).
[20] D. Kinghorn and L. Adamowicz, J. Chem. Phys. **110**, 7166 (1999).
[21] S. Bubin, M. Cafiero, and L. Adamowicz, Adv. Chem. Phys. **131**, 377 (2005).
[22] S. Bubin, L. Adamowicz, and M. Molski, J. Chem. Phys. **123**, 134310 (2005).
[23] M. Stanke D. Kędziera, S. Bubin, and L. Adamowicz, J. Chem. Phys. **127**, 134107 (2007).
[24] S. Bubin D. Kędziera, M. Stanke, and L. Adamowicz, Phys. Rev. A **75**, 062504 (2007).
[25] S. Bubin and L. Adamowicz, J. Chem. Phys. **126**, 214305 (2007).
[26] S. Bubin and L. Adamowicz, Phys. Rev. **79**, 022501 (2009).
[27] M. Stanke, S. Bubin, and L. Adamowicz, Phys. Rev. A **79**, 060501(R) (2009).
[28] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics* (Wiley, Chichester, 1988).
[29] Y. Ralchenko, A. Kramida, J. Reader, and NIST ASD Team, NIST Atomic Spectra Database, version 3.1.5, National Institute of Standards and Technology, Gaithersburg, MD, 2008, available online at http://physics.nist.gov/asd3.
[30] D. Sundholm and J. Olsen, Chem. Phys. Lett. **182**, 497 (1991).
[31] D. E. Woon and T. H. Dunning, J. Chem. Phys. **103**, 4572 (1995).