

Differentiable programming and density matrix based Hartree-Fock method

Hong-Bin Ren(任宏斌), Lei Wang(王磊), and Xi Dai(戴希)

Citation: Chin. Phys. B, 2021, 30 (**6**): 060701. DOI: 10.1088/1674-1056/abeeed
Journal homepage: http://cpb.iphy.ac.cn; http://iopscience.iop.org/cpb

What follows is a list of articles you may be interested in

Accurate Deep Potential model for the Al-Cu-Mg alloy in the full concentration space

Wanrun Jiang(姜万润), Yuzhi Zhang(张与之), Linfeng Zhang(张林峰), and Han Wang(王涵)
Chin. Phys. B, 2021, 30 (**5**): 050706. DOI: 10.1088/1674-1056/abf134

Efficient sampling for decision making in materials discovery

Yuan Tian(田原), Turab Lookman, and Dezhen Xue(薛德祯)
Chin. Phys. B, 2021, 30 (**5**): 050705. DOI: 10.1088/1674-1056/abf12d

Silicon-based optoelectronic synaptic devices

Lei Yin(尹蕾), Xiaodong Pi(皮孝东), Deren Yang(杨德仁)
Chin. Phys. B, 2020, 29 (**7**): 070703. DOI: 10.1088/1674-1056/ab973f

An artificial synapse by superlattice-like phase-change material for low-power brain-inspired computing

Qing Hu(胡庆), Boyi Dong(董博义), Lun Wang(王伦), Enming Huang(黄恩铭), Hao Tong(童浩), Yuhui He(何毓辉), Ming Xu(徐明), Xiangshui Miao(缪向水)
Chin. Phys. B, 2020, 29 (**7**): 070701. DOI: 10.1088/1674-1056/ab892a

High-performance synaptic transistors for neuromorphic computing

Hai Zhong(钟海), Qin-Chao Sun(孙勤超), Guo Li(李果), Jian-Yu Du(杜剑宇), He-Yi Huang(黄河意), Er-Jia Guo(郭尔佳), Meng He(何萌), Can Wang(王灿), Guo-Zhen Yang(杨国桢), Chen Ge(葛琛), Kui-Juan Jin(金奎娟)
Chin. Phys. B, 2020, 29 (**4**): 040703. DOI: 10.1088/1674-1056/ab7806

Differentiable programming and density matrix based Hartree–Fock method*

Hong-Bin Ren(任宏斌)[1,2], Lei Wang(王磊)[1,3], and Xi Dai(戴希)[4,†]

[1] *Beijing National Laboratory for Condensed Matter Physics and Institute of Physics, Chinese Academy of Sciences, Beijing* 100190, *China*
[2] *School of Physical Sciences, University of Chinese Academy of Sciences, Beijing* 100049, *China*
[3] *Songshan Lake Materials Laboratory, Dongguan* 523808, *China*
[4] *Department of Physics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon* 999077, *Hong Kong, China*

(Received 4 February 2021; revised manuscript received 13 March 2021; accepted manuscript online 16 March 2021)

Differentiable programming is an emerging programming paradigm that allows people to take derivative of an output of arbitrary code snippet with respect to its input. It is the workhorse behind several well known deep learning frameworks, and has attracted significant attention in scientific machine learning community. In this paper, we introduce and implement a density matrix based Hartree–Fock method that naturally fits into the demands of this paradigm, and demonstrate it by performing fully variational ground state calculation on several representative chemical molecules.

Keywords: differentiable programming, quantum chemistry

PACS: 07.05.Mh, 63.20.dk **DOI:** 10.1088/1674-1056/abeeed

1. Introduction

Hartree–Fock (HF) method[1] is one of the fundamental methods in many body physics. It has been widely applied in quantum chemistry to explore the ground state properties of molecules. Due to its simplicity, it is also considered as a good starting point for more accurate and expensive configuration interaction or perturbative calculations. In physics perspective, HF is a variational method whose variational space is determined by basis functions used to represent the molecular orbital. And we can achieve desired accuracy by including sufficient amount of basis functions. However, traditional HF method has $\mathcal{O}(K^4)$ (K is the number of basis functions used to expand the molecular orbital) computational complexity, which suggests that calculation on large basis set is only manageable for molecules in restricted scale. Instead of monotonically increasing the amount of basis functions in basis set, researchers indicate that one could also augments the variational space of HF by treating basis function related parameters, such as exponents, contraction coefficents and centers of Gaussian type orbital (GTO),[2] as variational parameters, and updating their values during HF's self consistent field (SCF).[3–7] Following this idea, Tachikawa *et al.* implemented fully variational HF method.[7] Their numerical results show that the augmented variational space can improve the total energy estimation with a small number of basis functions.

In previously reported works, gradients of total energy with respect to basis set parameters are manually derived for particular basis set and for specific molecule. This prevents the generalization of fully variational approach to practical situations. With the advent of the modern algorithm and computer technology, especially the boom of deep learning,[8] the process of manual derivation of the gradients could be replaced by automatic differentiation (AD).[9] Popular AD frameworks include Pytorch, TensorFlow[10] in Python, and ForwardDiff.jl,[11] Zygote.jl[12] in Julia. By using the state-of-the-art AD tools, Tamayo-Mendoza *et al.*[13] have successfully implemented the DiffiQult library. In DiffiQult, derivatives of HF energy with respect to GTO parameters can be obtained directly from AD, which greatly simplifies the heavy derivative evaluation process, and makes it possible to efficiently compare the effects of different basis sets. However, DiffiQult uses forward-mode AD, which is not efficient for taking derivative of a function whose input dimension is larger than the output (here, the output is energy, which is a scalar). Moreover, because the number of SCF cyclies needed for converged HF calculation is undetermined before calculation starts, their gradient evaluation may cost a large amount of computation resources.

In this paper, we design a mixed mode AD method and use it to implement fully variational density matrix HF (DMHF) method proposed by Helgaker *et al.*[14] and Hoai.[15] Unlike the original DMHF which uses purification method[16] to guarantee the validity of one body density matrix, we use Thouless theorem[17] to directly represent the one body density matrix in a set of non-redundant parameters. In this way, we eliminate the need to perform fixed point iterations separately, and can update the density matrix and GTO parameters simultaneously in a single optimization step. Our paper is or-

*Project supported by the Hong Kong Research Grants Council, China (Project No. GRF16300918), the National Key R&D Program of China (Grant Nos. 2016YFA0300603 and 2016YFA0302400), and the National Natural Science Foundation of China (Grant No. 11774398).

†Corresponding author. E-mail: daix@ust.hk

© 2021 Chinese Physical Society and IOP Publishing Ltd

ganized as follows. In Section 2, we introduce a theoretical description of DMHF method. In Section 3, we show results obtained from ground state calculation of some representative molecules, and compare the performance of our algorithm to previously published ones. At last, in Section 4, we present the conclusion and future perspective.

2. Method

2.1. Thouless theorem and parametrization of one body density matrix

The one body density matrix $\boldsymbol{\rho}$ is defined as

$$\rho_{ji} = \frac{\langle \Phi | \hat{c}_i^\dagger \hat{c}_j | \Phi \rangle}{\langle \Phi | \Phi \rangle}, \tag{1}$$

where $\hat{c}_i^\dagger$ and $\hat{c}_j$ are fermion creation and annihilation operators, and $|\Phi\rangle$ is a N-body Slater determinant state generated by another set of fermion operators $\hat{a}_1^\dagger, \ldots, \hat{a}_N^\dagger$,

$$\left\{\hat{c}_i, \hat{c}_j^\dagger\right\} = \left\{\hat{a}_i, \hat{a}_j^\dagger\right\} = \delta_{ij}.$$

Thouless[17] indicates that by introducing operator $\hat{b}_i^\dagger$ represented as

$$\hat{b}_i^\dagger = \hat{c}_i^\dagger + \sum_{j=N+1}^{D} M_{ji} \hat{c}_j^\dagger, \quad i \in [1, N],$$

where M_{ji} is unconstrained and is an element of a $(D-N)\times N$ matrix $\boldsymbol{M}$ (D is the dimension of Hilbert space of the single particle wave function associated with those fermion operators). One could construct a Slater determinant $|\Phi_b\rangle$ which is equivalent to $|\Phi\rangle$ up to a normalization factor.

Therefore, by substituting $|\Phi\rangle$ with $|\Phi_b\rangle$ in Eq. (1), we obtain

$$\rho_{ji} = \frac{\langle \Phi_b | \hat{c}_i^\dagger \hat{c}_j | \Phi_b \rangle}{\langle \Phi_b | \Phi_b \rangle} \tag{2}$$

$$= \left(\begin{pmatrix} \boldsymbol{I} \\ \boldsymbol{M} \end{pmatrix} \boldsymbol{S}^{-1} \begin{pmatrix} \boldsymbol{I} & \boldsymbol{M}^\dagger \end{pmatrix} \right)_{ji}, \tag{3}$$

where $\boldsymbol{S} = \boldsymbol{I} + \boldsymbol{M}^\dagger \boldsymbol{M}$ is a $N \times N$ overlap matrix between single particle wave functions $\phi_i^b(x)$ generated by $\hat{b}_i^\dagger$, $i \in [1, N]$. Detailed derivation can be found in Appendix A.

2.2. Variational density matrix based Hartree–Fock method

In this paper, we focus on restricted HF (RHF) method, where the molecule is a closed-shell system with all orbitals doubly occupied. In this case, the one body density matrix $\boldsymbol{\rho}$ is spin independent, therefore, we could omit the spin label, and write the total energy of a molecule as

$$E(\boldsymbol{\theta}, \boldsymbol{\rho}) = 2\sum_{ij} t_{ij}(\boldsymbol{\theta}) \rho_{ji} + \sum_{ijkl} (2 v_{iljk}(\boldsymbol{\theta}) - v_{ikjl}(\boldsymbol{\theta})) \rho_{il} \rho_{jk} + E_{\text{nuc-nuc}}, \tag{4}$$

where $\boldsymbol{\theta}$ is the collection of parameters of GTOs, t_{ij} is the sum of matrix element of kinetic and external potential operators under GTO, and v_{iljk} is the tensor element of electron–electron interaction operator under GTO, and $E_{\text{nuc-nuc}}$ is the nuclear repulsion energy inside the molecule.

From the above discussion, $\boldsymbol{\rho}$ can be parametrized by $\boldsymbol{M}$. Therefore, by variationally optimizing Eq. (4) with respect to $\boldsymbol{M}$ and $\boldsymbol{\theta}$

$$\boldsymbol{\theta}^*, \boldsymbol{M}^* = \arg\min_{\boldsymbol{M}, \boldsymbol{\theta}} E(\boldsymbol{\theta}, \boldsymbol{\rho}(\boldsymbol{M})), \tag{5}$$

one could obtain GTOs and one body density matrix that are suitable for describing the ground state of a given molecule.

We implemented our DMHF method in Julia with AD supported by Zygote.jl. Our computation graph is demonstrated in Fig. 1. As pointed out by Tamayo-Mendoza *et al.*, for most of the methods in quantum chemistry, one center integral matrix $\boldsymbol{t}$ and two-center integral tensor $\boldsymbol{v}$ are constructed by an element-wise array assignment, which is not feasible for reverse mode AD to manage. We solve this problem by wrapping the array assignment operation as a primitive function that maps $\boldsymbol{\theta}$ to $\boldsymbol{t}$ and $\boldsymbol{v}$. We define custom adjoint for this primitive that computes Jacobian matrix using forward mode AD inside, then pass the Jacobian to reverse mode AD to compute the vector-Jacobian product. And because we directly parameterize the density matrix, our calculation of energy only involves matrix multiplication, trace, and tensor contraction.

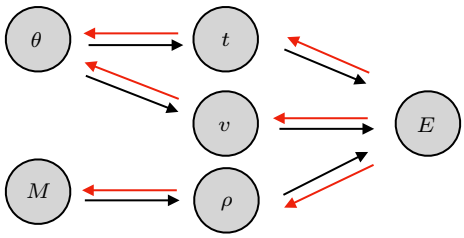


Fig. 1. The computation graph for our DMHF method. Black arrows indicate the forward function evaluation from inputs to outputs, and red arrows indicate backward steps for derivative propogation.

We demonstrate our method by the calculation of several small molecules. As a benchmark, we first perform fully variational (FV) calculation on hydrogen molecule, and compare our result to that obtained by DiffiQult. By FV, we can simultaneously optimize GTO parameters and the one body density matrix. The H–H bond length is set to 1.388 Bohr radius during the calculation. Figure 2(a) shows the variation of the ground state energy in the optimization process. We can see that our method converges much faster than DiffiQult, and could reaches a ground state energy comparable to that computed by 6-31G basis set which is two times larger than STO-3G basis we used in this case. If we switch off the variation of GTO centers, the ground state energy obtained by our method

will be the same as that by DiffiQult, but still has a faster convergence rate. Figure 2(b) shows the change of ground state electron density during the optimization process. The initial electron density is calculated from a randomly initialized density matrix. After optimization, the electron density becomes reasonable and approaches the one calculated by SCF method in Ref. [18] using STO-3G basis. Since we allow the variation of GTO centers, we also find that two peaks get close to each other suggesting that FV method could partly address the polarization of molecular orbital even with a minimal basis set.

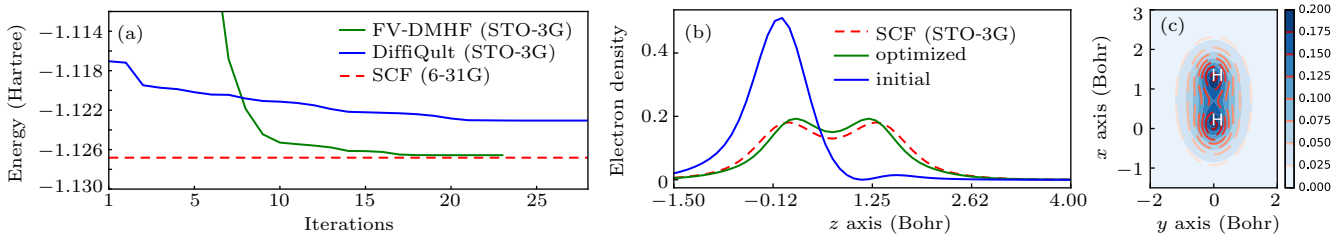


Fig. 2. Fully variational calculation of ground state of hydrogen molecule. (a) Optimization of ground state energy. The green line is calculated by FV DMHF with STO-3G basis set, the blue line is obtained using Tamayo-Mendoza *et al.* DiffiQult package, the dashed red line corresponds to SCF calculation using 6-31G basis set. (b) Electron density variation during the DMHF calculation, blue line corresponds to the initial electron density of hydrogen molecule before optimization, the green line is the final electron density after optimization, the dashed red line shows the SCF result with STO-3G basis. (c) Electron density distribution in the *yz* plane obtained from DMHF calculation, same result calculated from SCF with STO-3G is shown in dashed red line.

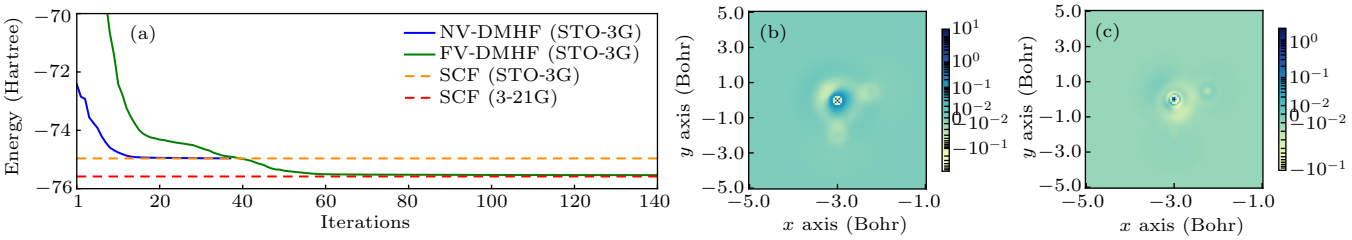


Fig. 3. Ground state calculation of water (SCF calculation is done with PySCF). (a) Optimization process. The blue line is the result obtained by using non variational DMHF in which basis set parameters are fixed, the green line corresponds to fully variational DMHF where we optimize the basis set parameters together with density matrix, both of them are calculated use STO-3G basis set. We compare our results with those obtained from SCF calculation with STO-3G basis set (dashed orange line) and 3-21G basis set (dashed red line). (b) Difference between electron density calculated by SCF method with STO-3G basis set and 3-21G basis set. (c) Differences between electron density calculated by FV DMHF with STO-3G and SCF with 3-21G basis set.

Table 1. GTO exponents of hydrogen and oxygen atoms in water, we compare the value in default setting of STO-3G and optimized STO-3G basis set to those in 3-21G basis set.

		1s	2s	2p		
Atom	Basis			$2p_x$	$2p_y$	$2p_z$
O	STO-3G (default)	130.709	5.033	5.033	5.033	5.033
		23.808	1.169	1.169	1.169	1.169
		6.443	0.380	0.380	0.380	0.380
	STO-3G (optimized)	286.139	11.244	7.681	7.901	7.543
		42.998	0.906	1.596	1.623	1.543
		9.224	0.319	0.367	0.363	0.321
	3-21G (default)	322.037	7.402	7.402	7.402	7.402
		48.430	1.576	1.576	1.576	1.576
		10.420	0.373	0.373	0.373	0.373
H	STO-3G (default)	3.425				
		0.623				
		0.168				
	STO-3G (optimized)	6.223				
		0.946				
		0.194				
	3-21G (default)	5.447				
		0.824				
		0.183				

We next apply DMHF method to determine the ground state of water molecule. Fully variational calculation of water using STO-3G basis requires simultaneously optimizing 73 parameters at different scales, compared to only 13 parameters in hydrogen molecule case. In Fig. 3(a), we compare the ground state energy calculated by FV DMHF and DMHF with GTO related parameters being fixed (non-variational (NV) DMHF). After optimization, NV DMHF converges to the ground state energy obtained by SCF method in PySCF, which is what we expect, while FV DMHF converges to a lower ground state energy that approximates the result obtained by SCF with 3-21G basis. Since 3-21G basis is a double zeta basis, its result is expected to be more accurate. In Fig. 4, we illustrate the variation of GTO exponents during the optimization process. We observe that starting from the default values of STO-3G basis set, these exponents finally evolve to values that are close to the default exponents of 3-21G basis set. The above results suggest that our fully variational approach makes better use of limited basis set and could effectively learn information encoded in higher dimensional Hilbert space. We also list the default and optimized GTO exponents explicitly in Table 1. In addition, in most of the commonly used basis sets, usually the same GTO exponents are applied to 2s and 2p orbitals, while in FV method, these parameters can be tuned independently which allows much larger variational space. Ground state electron density calculated from FV method is also compared to that obtained by SCF method un-

der 3-21G basis, and the difference between the two is shown in Fig. 3(c). Compared to the electron density calculated by the default STO-3G basis set (the difference between results of STO-3G and 3-21G basis set is shown in Fig. 3(b)), our method does not force electron to be on top of each atom, allows the electron density to adjust its spatial extent appropriate to each molecular.

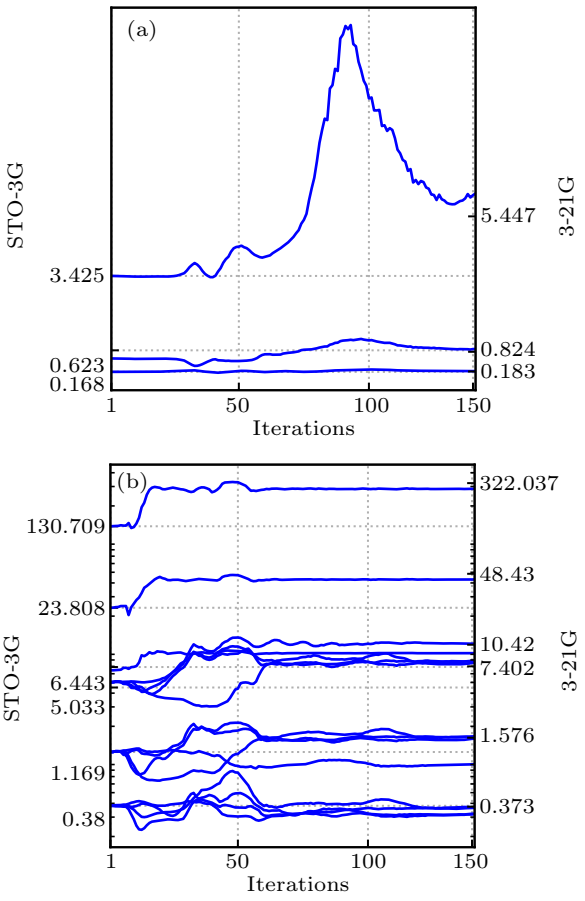


Fig. 4. Evolution of GTO exponents of hydrogen (a) and oxygen (b) in water molecule during the FV DMHF optimization process. The initial value is set to the default value of STO-3G basis set, our FV approach finally sends them to a value close to the default value of 3-21G, which performs better than original STO-3G basis in computing ground state energy and electron density. These values are also listed in Table 1.

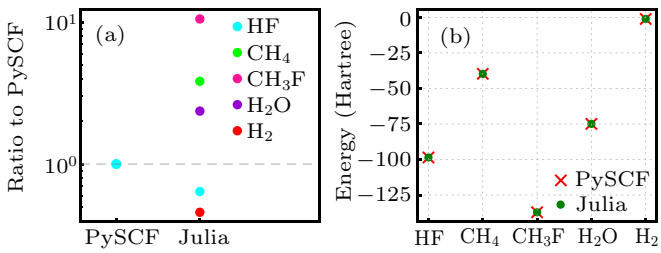


Fig. 5. Benchmark of our DMHF program with popular PySCF on several small molecules. (a) Walltime comparison between optimization process of our Julia code and SCF iterations of PySCF. The PySCF time is normalized to one, and Julia time is presented as its ratio to PySCF time. The results were obtained on single thread Intel(R) Xeon(R) E5-1650 v2 3.50GHz CPU with 16GB RAM, running macOS. (b) Comparison of the ground state energy obtained by PySCF and our Julia DMHF method.

We benchmark the performance of our DMHF implementation in non-variational setting against popular PySCF software on several small molecules, and show the benchmark results in Fig. 5. We perform calculation of ground state energy of H_2, HF, H_2O, CH_4, and CH_3F on Intel(R) Xeon(R) E5-1650 v2 3.50 GHz CPU with 16GB RAM machine, running macOS, and record the time needed for converged SCF in PySCF and optimization in our Julia DMHF. We normalize the time used by PySCF calculations to 1, and show ratio of our implementation over PySCF time in Fig. 5(a). For calculation of H_2 and HF, our implementation surpasses PySCF, and is slightly slower than PySCF in H_2O and CH_4 calculation. The worst case happens for CH_3F, where our implementation is 10 times slower. Comparing to PySCF which derives the initial density matrix from atomic orbitals, our density matrix is randomly initialized, therefore, may require more optimization steps.

3. Conclusion and perspective

In this work, we show that one body density matrix within HF approximation can be constructed using Thouless theorem, based on which a DMHF method can be implemented. Using density matrix directly in HF helps us to remove the usual SCF that contains diagonalization of a Hamiltonian matrix in energy calculation, and enables us to use reverse mode AD for efficient gradient evaluation. When applied to calculate the ground state of small molecules, with default setting, our DMHF could achieve the same performance as traditional SCF base HF. If we allow GTO related parameters to vary, FV DMHF can be used to obtain more accurate ground state energy with a smaller basis set, which suggests the information hidden in high dimensional Hilbert space could be successfully learned by the FV approach. In future, we tend to bring the density matrix treatment to density functional theory (DFT),[19,20] which is much widely applied in the computational physics. Since we are able to extract real space electron density from density matrix, we could just replace the HF's exchange energy calculation with the existing exchange–correlation functionals. By that, we could avoid solving Kohn–Sham equation, like what is done in orbital free DFT.

Acknowledgment

H. R. thanks Jinguo Liu for helpful discussion.

Appendix A: Proof of Thouless theorem

Consider the N-body Slater determinant $|\Psi\rangle$ generated by fermion operator $\hat{c}_1^\dagger, \ldots, \hat{c}_N^\dagger$, Thouless theorem states that

Theorem Any N-body Slater determinant $|\Phi\rangle$ which is not orthogonal to $|\Psi\rangle$ can be written in the form

$$|\Phi\rangle = \left[\Pi_{i=1}^{N}\Pi_{k=N+1}^{D}\left(1+M_{ki}\hat{c}_k^\dagger\hat{c}_i\right)\right]|\Psi\rangle, \qquad \text{(A1)}$$

where D is the dimension of the Hilbert space of one particle wavefunction associated with $\hat{c}_i^\dagger$, and M_{ki} are coefficients that could be uniquely determined.

To prove this, we introduce another set of fermion operator $\hat{a}_i^\dagger$, the single particle wavefunctions generated by $\hat{a}_i^\dagger$ are in the same Hilbert space as those generated by $\hat{c}_i^\dagger$. These two sets of fermion operators can be connected by a unitary transformation $\boldsymbol{U}$

$$\hat{a}_l^\dagger = \sum_{j=1}^{D} U_{jl}\hat{c}_j^\dagger.$$

We partition $\boldsymbol{U}$ into four blocks:

$$\boldsymbol{U} = \begin{pmatrix} \boldsymbol{U}_{11} & \boldsymbol{U}_{12} \\ \boldsymbol{U}_{21} & \boldsymbol{U}_{22} \end{pmatrix},$$

in which $\boldsymbol{U}_{11}$ is a $N\times N$ matrix, $\boldsymbol{U}_{21}$ is $(D-N)\times N$ matrix.

If the N body Slater determinant $|\Phi\rangle$ and $|\Psi\rangle$ generated by these fermion operators

$$|\Phi\rangle = \Pi_{i=1}^{N}\hat{a}_i^\dagger|0\rangle, \quad |\Psi\rangle = \Pi_{i=1}^{N}\hat{c}_i^\dagger|0\rangle$$

are not orthogonal $\langle\Phi|\Psi\rangle \neq 0$. We can scale $|\Phi\rangle$ by the inverse of $\langle\Phi|\Psi\rangle$

$$\begin{aligned} |\Phi'\rangle &= \frac{1}{\langle\Phi|\Psi\rangle}|\Phi\rangle \\ &= \det\left(\boldsymbol{U}_{11}^{-1}\right)\Pi_{i=1}^{N}\hat{a}_i^\dagger|0\rangle \\ &= \Pi_{i=1}^{N}\sum_{m=1}^{N}\left(\boldsymbol{U}_{11}^{-1}\right)_{mi}\left(\sum_{j=1}^{N}\boldsymbol{U}_{jm}\hat{c}_j^\dagger + \sum_{k=N+1}^{D}\boldsymbol{U}_{km}\hat{c}_k^\dagger\right) \\ &= \Pi_{i=1}^{N}\left(\hat{c}_i^\dagger + \sum_{k=N+1}^{D}\boldsymbol{M}_{ki}\hat{c}_k^\dagger\right)|0\rangle, \end{aligned} \tag{A2}$$

where $\boldsymbol{M} \equiv \boldsymbol{U}_{21}\boldsymbol{U}_{11}^{-1}$, and $\langle\Phi|\Psi\rangle = \det(\boldsymbol{U}_{11})$ is proved using Wick's theorem as

$$\begin{aligned} \langle\Phi|\Psi\rangle &= \langle 0|\hat{a}_N\ldots\hat{a}_1\hat{c}_1^\dagger\ldots\hat{c}_N^\dagger|0\rangle \\ &= \sum_{i_1,\ldots,i_N}(-)^{P(i_1,\ldots,i_N)}\langle 0|\hat{a}_{i_N}\ldots\overbrace{\hat{a}_{i_1}\hat{c}_1^\dagger}\ldots\hat{c}_N^\dagger|0\rangle \\ &= \det(\boldsymbol{U}_{11}). \end{aligned}$$

Equation (A2) is equivalent to Eq. (A1) since

$$\begin{aligned} &\Pi_{i=1}^{N}\left(\hat{c}_i^\dagger + \sum_{k=N+1}^{D}M_{ki}\hat{c}_k^\dagger\right)|0\rangle \\ =\ &\Pi_{i=1}^{N}\left(\hat{c}_i^\dagger + \sum_{k=N+1}^{D}M_{ki}\hat{c}_k^\dagger\left\{\hat{c}_i,\hat{c}_i^\dagger\right\}\right)|0\rangle \\ =\ &\Pi_{i=1}^{N}\left\{\left(1+\sum_{k=N+1}^{D}M_{ki}\hat{c}_k^\dagger\hat{c}_i\right)\hat{c}_i^\dagger + \sum_{k=N+1}^{D}M_{ki}\hat{c}_k^\dagger\hat{c}_i^\dagger\hat{c}_i\right\}|0\rangle \\ =\ &\left[\Pi_{i=1}^{N}\Pi_{k=N+1}^{D}\left(1+M_{ki}\hat{c}_k^\dagger\hat{c}_i\right)\right]|\Psi\rangle. \end{aligned}$$

So, this proves the density matrix parameterization mentioned in Section 2.

Appendix B: Detailed proof of density matrix representation

Following notation in Section 2, the Laplace expansion of the determinant of $N\times N$ matrix $\boldsymbol{S}$ is known as

$$\det(\boldsymbol{S}) = \sum_{m,n=1}^{N}(-)^{m+n}(\boldsymbol{S})_{mn}\det\left(\left(\bar{\boldsymbol{S}}\right)_{mn}\right),$$

where $\left(\bar{\boldsymbol{S}}\right)_{mn}$ denotes a matrix generated from $\boldsymbol{S}$ by removing its m-th row and n-th column. Dividing $\det(\boldsymbol{S})$ on both sides, we obtain

$$\sum_{m,n=1}^{N}(\boldsymbol{S})_{mn}\times(-)^{m+n}\frac{\det\left(\left(\bar{\boldsymbol{S}}\right)_{mn}\right)}{\det(\boldsymbol{S})} = 1,$$

which suggests that

$$\left(\boldsymbol{S}^{-1}\right)_{nm} = (-)^{m+n}\frac{\det\left(\left(\bar{\boldsymbol{S}}\right)_{mn}\right)}{\det(\boldsymbol{S})}. \tag{B1}$$

Expanding Eq. (2) and pluging in Eq. (B1), we obtain

$$\begin{aligned} \boldsymbol{\rho}_{ji} &= \frac{\langle\Phi_b|\hat{c}_i^\dagger\hat{c}_j|\Phi_b\rangle}{\langle\Phi_b|\Phi_b\rangle} \\ &= \frac{1}{\det(\boldsymbol{S})}\langle 0|\hat{b}_N\ldots\hat{b}_1\hat{c}_i^\dagger\hat{c}_j\hat{b}_1^\dagger\ldots\hat{b}_N^\dagger|0\rangle \\ &= \left(\begin{pmatrix}\boldsymbol{I}\\ \boldsymbol{M}\end{pmatrix}\boldsymbol{S}^{-1}\begin{pmatrix}\boldsymbol{I} & \boldsymbol{M}^\dagger\end{pmatrix}\right)_{ji}. \end{aligned}$$

We now prove that density matrix generated by Eq. (3) is a valid density matrix. Noting that $\left(\boldsymbol{S}^{-1}\right)^\dagger = \boldsymbol{S}^{-1}$, $\boldsymbol{\rho}$ is therefore a Hermitian matrix. Plugging Eq. (3) into $\boldsymbol{\rho}^2 = \boldsymbol{\rho}$, we obtain

$$\begin{aligned} \boldsymbol{\rho}^2 &= \begin{pmatrix}\boldsymbol{I}\\ \boldsymbol{M}\end{pmatrix}\boldsymbol{S}^{-1}\begin{pmatrix}\boldsymbol{I} & \boldsymbol{M}^\dagger\end{pmatrix}\begin{pmatrix}\boldsymbol{I}\\ \boldsymbol{M}\end{pmatrix}\boldsymbol{S}^{-1}\begin{pmatrix}\boldsymbol{I} & \boldsymbol{M}^\dagger\end{pmatrix} \\ &= \begin{pmatrix}\boldsymbol{I}\\ \boldsymbol{M}\end{pmatrix}\boldsymbol{S}^{-1}\left(\boldsymbol{I}+\boldsymbol{M}^\dagger\boldsymbol{M}\right)\boldsymbol{S}^{-1}\begin{pmatrix}\boldsymbol{I} & \boldsymbol{M}^\dagger\end{pmatrix} = \boldsymbol{\rho}, \end{aligned}$$

which suggests that $\boldsymbol{\rho}$ is also an idempotent matrix.

The electron number conservation can be proved with the help of cyclic property of matrix trace

$$\begin{aligned} \mathrm{trace}(\boldsymbol{\rho}) &= \mathrm{trace}\left(\begin{pmatrix}\boldsymbol{I}\\ \boldsymbol{M}\end{pmatrix}\boldsymbol{S}^{-1}\begin{pmatrix}\boldsymbol{I} & \boldsymbol{M}^\dagger\end{pmatrix}\right) \\ &= \mathrm{trace}\left(\begin{pmatrix}\boldsymbol{I} & \boldsymbol{M}^\dagger\end{pmatrix}\begin{pmatrix}\boldsymbol{I}\\ \boldsymbol{M}\end{pmatrix}\boldsymbol{S}^{-1}\right) = N. \end{aligned}$$

References

[1] Helgaker T, Jorgensen P and Olsen J 2014 *Molecular electronic-structure theory* (John Wiley and Sons)
[2] Boys S F 1950 *Proc. R. Soc. Lond. Ser. A* **200** 542
[3] Hurley A C 1954 *Proc. R. Soc. Lond. Ser. A* **226** 170
[4] Frost A A 1967 *J. Chem. Phys.* **47** 3714
[5] Helgaker T and Almlöf J 1988 *J. Chem. Phys.* **89** 4889
[6] Jensen F 1999 *J. Chem. Phys.* **110** 6601

[7] Tachikawa M, Taneda K and Mori K 1999 *Int. J. Quantum Chem.* **75** 497
[8] LeCun Y, Bengio Y and Hinton G 2015 *Nature* **521** 436
[9] Baydin A G, Pearlmutter B A, Radul A A and Siskind J M 2017 *The Journal of Machine Learning Research* **18** 5595
[10] Abadi M, Barham P, Chen J, *et al.* 2016 *symposium on operating systems design and implementation* pp. 265–283
[11] Revels J, Lubin M and Papamarkou T 2016 arXiv:1607.07892
[12] Innes M 2018 arXiv:1810.07951
[13] Tamayo-Mendoza T, Kreisbeck C, Lindh R and Aspuru-Guzik A 2018 *ACS Cent. Sci.* **4** 559
[14] Helgaker T, Larsen H, Olsen J and Jorgensen P 2000 *Chem. Phys. Lett.* **327** 397
[15] Hoai L T 2015 *Product wave-functions in Fock-space* (M.S. thesis) (German Research School for Simulation Sciences)
[16] McWeeny R 1960 *Rev. Mod. Phys.* **32** 335
[17] Thouless D J 1960 *Nucl. Phys.* **21** 225
[18] Sun Q, Berkelbach T C and Blunt N S, *et al.* 2018 *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **8** e1340
[19] Kohn W and Sham L J 1965 *Phys. Rev.* **140** A1133
[20] Hohenberg P and Kohn W 1964 *Phys. Rev.* **136** B864