Single-excitation guided sampling for particle-preserving fermionic Thouless states as an approach to determining the electronic structure ground state for single-reference molecules

February 23, 2026

Abstract

Diagonalisation of a coherent-state sample around the HF reference state is proposed as a post-HF method to estimate the molecular-electronic ground state. The $SU(M)$ fermionic Thouless states are described with regard to the relation between their parameter matrix and their overlap with single-excitation Slater determinants. For a RHF basis, a method which guides the sampling process based on the diagonalisation on zero- and single-excitation Slater determinant subspace is described and demonstrated on spin-zero molecules Li_2 and N_2. Generalisation of the sampling guidance process for an UHF basis is described and demonstrated on spin-one molecules BeH and NO.

Contents

1 Background

1.1 Fermionic $SU(M)$ Thouless states in electronic structure

Consider a Hilbert space described by M modes occupied in total by S fermions, where $S < M$. The occupancy basis of this Hilbert space consists of the Slater determinants

$$|n_1, n_2 \ldots n_M\rangle \quad \text{where} \quad \sum_{i=1}^{M} n_i = S \quad \text{and} \quad n_i \in \{0, 1\}$$

By defining the reference state as

$$|\phi_0\rangle = |1, 1 \ldots 1, 0, 0 \ldots 0\rangle \tag{1}$$

we may label every occupancy basis state as $|\phi_0 + a - b\rangle$, where a, b are equal-sized sets of indices with the constraints

$$
\begin{aligned}
a_i \in a \quad &: \quad 1 \leq a_i \leq M - S \\
b_i \in b \quad &: \quad 1 \leq b_i \leq S
\end{aligned}
$$

and the state $|\phi_0 + a - b\rangle$ describes the Slater determinant where modes occupied in ϕ_0 are occupied iff their index is not in b, and modes unoccupied in ϕ_0 are occupied iff their index is in a (with the first unoccupied mode, i.e. the $S + 1$-st mode, having index 1).

It will be helpful to define the integer sets

$$
\begin{aligned}
\pi_1 &= \{1, 2 \ldots S\} \\
\pi_0 &= \{1, 2 \ldots M - S\}
\end{aligned}
$$

which enumerate the occupied and unoccupied modes in $|\phi_0\rangle$, respectively.

Then, we can construct the $SU(M)$ Thouless states restricted to the Hilbert space spanned by the aforementioned basis from the reference state like so:

$$|Z\rangle = N(Z, Z^*) \exp\left(\sum_{i \in \pi_0} \sum_{j \in \pi_1} \hat{f}_{i+S}^\dagger \hat{f}_j \right) |\phi_0\rangle \tag{2}$$

where N is the normalisation coefficient and $\hat{f}^\dagger, \hat{f}$ are the fermionic creation and annihilation operators, respectively. Note that Z is therefore a $(M - S) \times S$ matrix of complex number.

These states form a non-orthogonal, overcomplete basis of the full Hilbert space.

It is then a known result that the decomposition of this Thouless state into the occupancy basis is

$$\langle \phi_0 + a - b | Z \rangle = (-1)^{\frac{1}{2}|a|(|a|+1)} \det\{Z_{a,b}\} \tag{3}$$

where $|a| = |b|$ is the number of excitations from the reference state and $Z_{a,b}$ is the submatrix of Z composed of rows and columns labelled by indices in a and b, respectively.

In other words, the minors of Z describe the decomposition coefficients of specific occupancy basis states in $|Z\rangle$, with all minors of order r describing all occupancy basis states obtained by promoting exactly r fermions from modes occupied in $|\phi_0\rangle$ into modes unoccupied in $|\phi_0\rangle$.

1.2 Electronic structure described by pairs of Thouless states

Under the Born-Oppenheimer approximation, we can describe the wavefunction of the electrons of a particular molecule by a vector in a Hilbert space spanned by Slater determinants constructed on molecular spin-orbitals:

$$|\Psi\rangle = \left| n(\phi_1^\alpha), n(\phi_1^\beta), n(\phi_2^\alpha), n(\phi_2^\beta) \dots \right\rangle \tag{4}$$

where, for practicality, we terminate the sequence of molecular orbitals at some highest orbital ϕ_M, where we hope that the behaviour of the molecule in the energy range of interest as described on this finite basis closely matches its true behaviour.

As the labelling suggests, it is very natural to decompose this Hilbert space into a direct product of (spatial) molecular orbitals and the two spin states $|\alpha\rangle, |\beta\rangle$, which, assuming the Hamiltonian has full rotatinal symmetry, can be selected arbitrarily as any orthonormal basis of the spin-$\frac{1}{2}$ system (typically, these are associated with spin-up and spin-down states, i.e. eigenstates of $\hat{S}_z$ for the particular selection of the z-axis). Then, the occupancy basis vectors may be writ out like so:

$$|\Psi\rangle = |n_1, n_2 \dots n_M\rangle_\alpha \otimes |\bar{n}_1, \bar{n}_2 \dots \bar{n}_{\bar{M}}\rangle_\beta \tag{5}$$

where $n_i, \bar{n}_j$ are shorthands for the occupancy numbers of modes $\phi_i^\alpha, \phi_j^\beta$, respectively. Note that we do not implicitly assume that $\phi_i^\alpha = \phi_i^\beta$, as is seen by allowing even the total number of the MOs in each spin-subspace to be different. The case where this equivalence is assumed is referred to as a restricted Hartree-Fock (RHF) basis. Otherwise, we are working under the unrestricted Hartree-Fock (UHF) basis.

Note that in the non-relativistic case the Hamiltonian preserves the total number of electrons in each spin-subspace:

$$[\hat{H}, \hat{N}_\alpha] = [\hat{H}, \hat{N}_\beta] = 0 \quad \text{where} \quad \hat{N}_\alpha |\Psi\rangle = \sum_{i=1}^M n_i = N_\alpha \quad \text{and} \quad \hat{N}_\beta |\Psi\rangle = \sum_{i=1}^{\bar{M}} \bar{n}_i = N_\beta \tag{6}$$

Equivalently, $\hat{H}$ commutes with the total number operator $\hat{N} = \hat{N}_\alpha + \hat{N}_\beta$ and the total spin z-projection operator $\hat{S}_z = \hat{N}_\alpha - \hat{N}_\beta$. As a result, N_α, N_β, S_z are good quantum numbers.

Consequently, the two spin-subspaces of the full Hilbert space are spanned exactly by the kind of basis which admits the construction of fermionic $SU(M)$ Thouless states. Hence, we can span the Hilbert space by states of the form

$$|Z\alpha, Z_\beta\rangle = |Z_\alpha\rangle_\alpha \otimes |Z_\beta\rangle_\beta \tag{7}$$

where $|Z_\alpha\rangle, |Z_\beta\rangle$ are Thouless states with mode and total occupancy numbers being M, N_α and $\bar{M}, N_\beta$, respectively.

1.3   Approximating the electronic ground state

For a single-reference molecule, a reasonable assumption (which forms the basis of Hartree-Fock theory) is that its electronic ground state has a significant overlap with its reference state. To go beyond the reference state is to design a post-HF method, which takes the mean-field MOs and the Slater determinants constructed on them as the starting point.

A good method to find a state lower in energy than the reference state is to use the fact that the reference state's large overlap with the true ground state can be rephrased as them being in close proximity in the Z-parameter space. The ground state, of course, is not necessarily a pure Thouless state, but, utilising the over-completeness of the Thouless state basis, it should be sufficient to sample a small set of states around the reference state in the parameter space (making sure it also contains the reference state itself) and then diagonalise the Hamiltonian matrix formed on them.

By Cauchy's interlacing theorem, the lowest eigenvalue of the Hamiltonian matrix on our sample basis must be strictly lower than the reference state energy. The issue is that the improvement may be minuscule compared to $E_{\phi_0} - E_g$. In order to quickly converge to E_g without the need of a sample of a size comparable to the full Hilbert space dimension, we need to design a smart method of sampling the states. The method proposed in this article is called single-excitation guided sampling, or SEGS for short.

1.4   Electronic ground state spin-symmetry

It appears that the number of occupancy basis vector in the $\hat{S}_z$-degenerate Hilbert space is equal to

$$\binom{M}{N_\alpha}\binom{\bar{M}}{N_\beta}$$

However, there is one further restriction to be placed on this Hilbert space if our only goal is to find the energy states. Because the non-relativistic Hamiltonian contains no spin coordinates, it commutes not only with $\hat{S}_z$, but also with the total spin operator $\hat{S}^2$.

Because of this, total spin S is another good quantum number, and if we know its value for the ground state, we can further restrict ourselves to a subspace of the full Hilbert space which is degenerate in $\hat{S}^2$ with the same eigenvalue as the ground state.

In general, Slater determinants are not eigenstates of $\hat{S}^2$, and hence a more suitable basis would be formed of their careful linear combinations which are referred to as configuration state functions (CSFs), or as spin-adapted configurations (SACs).

For an RHF basis, some specific Slater determinants are eigenstates of $\hat{S}^2$. These are states where all open shells have the same spin state (i.e. all open shells are in either the α subspace or the β subspace). Other $\hat{S}^2$ eigenstates for a specific total spin number are constructed as superpositions of Slater determinants in a hierarchical fashion, which can be characterised by the number of open shells. For example, a singlet CSF with two open shells can be constructed like so:

$$|\text{state with } S = 0\rangle = \frac{1}{\sqrt{2}}\left(|1,0\rangle_\alpha \otimes |0,1\rangle_\beta + |0,1\rangle_\alpha \otimes |1,0\rangle_\beta\right) \tag{8}$$

Note that the above construction can be generalised to larger mode and total particle number cases if all other shells are closed. The expressions for CSF singlets with four, six etc. open shells are more complicated.

We cannot limit the Thouless states to be pure eigenstates of $\hat{S}^2$ by the nature of their construction, but this is not necessary–an aspect of SEGS is to naturally guide the sampling by considering the CSFs of the same spin as the ground state in the hierarchical order.

2 Results

The main idea of SEGS is to favour the particular excitations which are more strongly present in the ground state. A single excitation characterised by a pair of numbers $(i \rightarrow j)$ does not correspond to a single Slater determinant, but is characteristic of a subset of CSFs which span the subspace which is known to contain the ground state.

Even if we do not know the ground state beforehand, we can estimate the prevalence of a particular excitation by taking very small samples from the CSF-subspace, diagonalising them to obtain a ground state estimate, and then projecting it onto states in the sample which are characterised by that excitation. A hierarchy of Hilbert states emerges:

$\{$CSFs characterised by $(i \rightarrow j)\} \subset \{$Singly-excited CSFs$\} \subset$

$\{$CSFs with total spin matching the ground state$\} \subset \{N_\alpha, N_\beta$-preserving states$\}$

We shall start by stating more approximations and assumptions to outline the basic version of SEGS, and then propose how to generalise it in case those approximations and assumptions cease to hold.

2.1 SEGS on RHF

In case of a singlet molecule with restricted MOs, the reference state is

$$|\phi_0\rangle = |1, 1 \ldots 1, 0, 0 \ldots 0\rangle_\alpha \otimes |1, 1 \ldots 1, 0, 0 \ldots 0\rangle_\beta \quad \text{with} \quad N_\alpha = N_\beta, M = \bar{M}, \phi_i^\alpha = \phi_i^\beta$$

Then, the CSF space can be partitioned into singlets with no open shells ($\mathcal{H}_0$), singlets with two open shells ($\mathcal{H}_1$) etc. We shall assume that the ground state has a projection into $\mathcal{H}_0$ with norm nearing one.

In $\mathcal{H}_0$, the CSF with the largest overlap with the ground state is expected to be $|\phi_0\rangle$ itself. It is a reasonable assumption that larger number of excitations correlates with higher energy expectation value. Let us then consider the states

$$|(i \rightarrow j)\rangle = |\phi_0 + \{j\} - \{i\}\rangle_\alpha \otimes |\phi_0 + \{j\} - \{i\}\rangle_\beta \tag{9}$$

Based on our previous assumptions, we would expect these states (referred to as SECSs, short for single-excitation closed shells) to contain most of the ground state sans its $|\phi_0\rangle$ component. Actually,

we can get a good first estimate of the ground state energy by taking the basis set $\{|\phi_0\rangle\} + \{\text{SECS}\}$ and diagonalising its Hamiltonian matrix. The ground state $|\text{SECS g.s.}\rangle$ energy estimate on this basis set will be referred to as the SECS ground state energy. The number of SECS states is $(M - S)S$.

Now, we shall use the Thouless sampling method to improve further on the SECS ground state energy estimate by considering the overlap

$$\langle \text{SECS g.s.}|(i \to j)\rangle = \eta_{ij} \tag{10}$$

The magnitude η_{ij} is related to the component of the ground state consisting of states that feature excitation $(i \to j)$. Ergo, it is related to the expected value of the magnitude of Z_{ji} in the decomposition of the ground state into the Thouless state basis. However, just constructing a state $\left|\eta^T\right\rangle_\alpha \otimes \left|\eta^T\right\rangle_\beta$ does not help, since in this state the excitation $(i \to j)$ in the α subspace is not correlated with the same excitation in the β subspace, and the total state, having zero entanglement between spin subspaces, has a large projection onto states with one or more open shells.

However, the magnitude $|\eta_{ij}|$ gives us a good estimate for the expected amplitude of Z_{ji} in the random sample. Therefore, we propose two methods of sampling Thouless states:

- from a normal distribution with mean zero and standard deviation on Z_{ji} equal to η_{ij}.

- Taking $Z_{ij} = \eta_{ji}$ and multiplying every one of its elements by a random number from the complex unit circle.

The two methods both perform well on Li_2 and N_2, see figure below.

2.2 SEGS on UHF

On an UHF basis, no Slater determinant is a true eigenstate of $\hat{S}^2$ in general. However, if we assume that the reference state is still close to the ground state, the method readily generalises for an UHF basis, by simply considering all possible pairs of excitations on each spin subspace, including no excitation:

$$
\begin{aligned}
|(i \to j)\rangle_\alpha \otimes |\phi_0\rangle_\beta &= |\phi_0 + \{j\} - \{i\}\rangle_\alpha \otimes |\phi_0\rangle_\beta \\
|\phi_0\rangle_\alpha \otimes |(i \to j)\rangle_\beta &= |\phi_0\rangle_\alpha \otimes |\phi_0 + \{j\} - \{i\}\rangle_\beta \\
|(p \to q)\rangle_\alpha \otimes |(r \to s)\rangle_\beta &= |\phi_0 + \{q\} - \{p\}\rangle_\alpha \otimes |\phi_0 + \{s\} - \{r\}\rangle_\beta
\end{aligned}
$$

Once again, we can build a basis out of all possible states of this fashion, plus the reference state. The size of this basis is $(1 + (M - N_\alpha))N_\alpha(1 + (\bar{M} - N_\beta)N_\beta)$, which is a polynome of the fourth order. We can find the ground state on this sample, denoting it $|\text{SE g.g.}\rangle$, since it is formed on a basis where within each spin subspace at most a single excitation is considered.

Then, the spin-correlated single excitation weight consists of the decomposition coefficients for every pair of single excitations $(p \to q), (r \to s)$ in the decomposition of $|\text{SECS g.g.}\rangle$, including no excitation for either of the two spin subspaces.

The basic construction of Thouless states as discussed above has no entanglement between the two spin subspaces, and thus we cannot correlate the excitations between the two spin subspaces. For simplicity's sake, we may obtain the two single-excitation weights like so:

$$
\eta_{ij}^{\alpha} = \langle \text{SECS g.g.} | \left(|(i \to j)\rangle_{\alpha} \otimes |\phi_0\rangle_{\beta} + \sum_{k \in \pi_1^{\beta}} \sum_{l \in \pi_0^{\beta}} |(i \to j)\rangle_{\alpha} \otimes |(k \to l)\rangle_{\beta} \right) \tag{11}
$$

$$
\eta_{ij}^{\beta} = \langle \text{SECS g.g.} | \left(|\phi_0\rangle_{\alpha} \otimes |(i \to j)\rangle_{\beta} + \sum_{k \in \pi_1^{\alpha}} \sum_{l \in \pi_0^{\alpha}} |(k \to l)\rangle_{\alpha} \otimes |(i \to j)\rangle_{\beta} \right) \tag{12}
$$

In other words, we are projecting $|\text{SECS g.g.}\rangle$ onto the single-excitation subspace for a particular excitation in either spin-subspace. Although it may seem reasonable to restrict ourselves to one excitation in one spin-subspace and zero excitation in the other to restrict the basis size, this could not be expected to have a big overlap with the ground state because it disregards approximately closed shells. If the basis is too big, it may be helpful to only consider states where the two excitations are from and to modes of similar energy.

Once we obtain $\eta^{\alpha}, \eta^{\beta}$, we can use the same sampling techniques as for RHF, using each matrix as a guidance for Thouless component in the corresponding spin subspace.