Electronic Supplementary Material (ESI) for RSC Advances.

Compact wavefunctions from compressed imaginary time evolution

Jarrod R. McClean and Alán Aspuru-Guzik*

Department of Chemistry and Chemical Biology, Harvard University, Cambridge MA, 02138

E-mail: aspuru@chemistry.harvard.edu

1 Supplemental Information

1.1 Orthogonal Matching Pursuit

In this section, we offer some additional details on the implementation of Orthogonal Matching Pursuit[1] with imaginary-time evolution in quantum systems. Given a quantum state $|\Psi(\tau')\rangle$ that one wishes to reconstruct, orthogonal matching pursuit is a greedy decomposition algorithm (i.e. one that selects the best component at each iteration without regard to future iterations) that approximates the sparse problem of finding $|\tilde{\Psi}(\tau')\rangle$ such that

$$\begin{aligned} \min_{|\tilde{\Psi}(\tau')\rangle} || \, |\Psi(\tau')\rangle - |\tilde{\Psi}(\tau')\rangle \, ||_2^2 \\ \text{subject to } || \, |\tilde{\Psi}(\tau')\rangle \, ||_0 < N \end{aligned} \tag{1}$$

---

*To whom correspondence should be addressed

This is done by considering an overcomplete dictionary $\{|\Phi^i(\tau')\rangle\}$ that can express $|\tilde{\Psi}\rangle$ as

$$|\tilde{\Psi}(\tau')\rangle = \sum_i c_i(\tau')\,|\Phi^i(\tau')\rangle \tag{2}$$

and at each stage selecting selecting the $|\Phi^i(\tau')\rangle$ which maximizes the overlap with the residual with respect to the target signal $|\Psi(\tau')\rangle$,

$$\max_{|\Phi^i\rangle(\tau')} \frac{|\langle\Phi^i(\tau')|\Psi(\tau')\rangle - \sum_{j<i} c_j(\tau')\langle\Phi^i(\tau')|\Phi^j(\tau')\rangle|}{\sqrt{\langle\Phi^i(\tau')|\Phi^i(\tau')\rangle}} \tag{3}$$

In practice for quantum systems, the dictionary $\{|\Phi^i(\tau')\rangle\}$ can be any overcomplete basis for the N-particle Hilbert space, and the location of the optimal $|\Phi^i(\tau')\rangle$ can be done with a few different methods such as discrete enumeration of all basis states, stochastic search, and direct non-linear optimization. While discrete enumeration is commonly used in the orthogonal matching pursuit literature, the high dimensional nature of quantum systems does not readily allow it. Among the remaining options, we find that direct non-linear optimization is superior to stochastic search methods for the problems we considered. Specifically, we utilized a quasi-Newton BFGS procedure with analytic gradients and inexact line search satisfying the strong Wolfe conditions.[2] The implementation closely follows that discussed in the classic text by Nocedal and Wright with a modified Cholesky regularization to protect again instabilities in the approximate Hessian.

After selection of the optimal $|\Phi^i(\tau')\rangle$, the full set of coefficients $\{c_i(\tau')\}_{i=0}^{j}$ are re-determined by orthogonal projection of the selected basis functions on the signal $|\Psi(\tau')\rangle$. This is equivalent to solving the linear equation

$$Sc = v \tag{4}$$

for the coefficient vector c, where $S_{ij} = \langle\Phi^i(\tau')|\Phi^j(\tau')\rangle$, $v_i = \langle\Phi^i(\tau')|\Psi(\tau')\rangle$, and $c_i = c_i(\tau')$.

Throughout this procedure, one also has a choice of how to represent the target signal

$|\Psi(\tau')\rangle$. In some cases, it is feasible to construct $|\Psi(\tau')\rangle$ explicitly from a previous time step and imaginary time propagator G, and doing so could potentially facilitate the optimization procedure by examining properties of the state. However, exact expansion of the state $|\Psi(\tau')\rangle$ using G can have many terms for even modestly sized quantum systems, negating the potential benefits of compressing the wavefunction. In practice, we found that a much better approach is to directly with $G\,|\Psi(\tau)\rangle$ without first expanding the wavefunction explicitly. When using the linearized propagator $G(\lambda) \approx (I - d\tau(H-\lambda))$, this means that Hamiltonian and overlap matrix elements and their derivatives are sufficient for the implementation of the procedure.

In principle, at any time step, one may continue to add elements $|\Phi^i(\tau')\rangle$ until an arbitrary convergence tolerance is reached, i.e. $||\,|\Psi(\tau')\rangle - |\tilde{\Psi}(\tau')\rangle\,||_2 < \epsilon$ for some $\epsilon > 0$. However, as only the final state in the large τ limit is of interest, and any state that is not completely orthogonal to this state will eventually converge to it, some errors in intermediate steps are permissible. Thus a more economical approach, is to terminate the addition of states $|\Phi^i\rangle$ at intermediate time steps according to some proxy, such as sufficient decrease in the energy $\tilde{E}(\tau') = \langle\tilde{\Psi}(\tau')|\,H\,|\tilde{\Psi}(\tau')\rangle$ from the previous time step.

1.2 NOMAGIC Algorithm

In this section we detail our implementation of the NOMAGIC procedure, including how values of algorithmic parameters may be chosen. Consider a physical system defined by the Hamiltonian H with a number of particles N, each represented on a discrete basis of M functions. The NOMAGIC algorithm begins by selecting an initial component $|\tilde{\Psi}(0)\rangle = |\Phi^0(0)\rangle$ from some approximation procedure. In this work we utilize a mean-field procedure that minimizes the energy, namely the Hartree-Fock algorithm,[3] to find an initial state, however other procedures such as a tensor cross approximation may be used.[4] Using the selected initial component, the initial expectation value of energy is computed as

$$\tilde{E}(0) = \langle \tilde{\Psi}(0) | \, H \, | \tilde{\Psi}(0) \rangle \, . \tag{5}$$

Additionally, this initial component is also used to estimate a safe value for the imaginary timestep τ. The bound on τ that guarantees a correct final state[5] under exact propagation is given by

$$d\tau \leq \frac{2}{E_{\max} - \lambda} \tag{6}$$

where $E_{\max}$ is the maximum eigenvalue of H. However, one does not expect to know the eigenvalues in advance, and moreover the problem is changed by the fact that propagation may be performed to some finite, economical precision. As a result, we use a fast estimate to determine an approximate suitable timestep $d\tau$. This is done by constructing the corresponding mean-field Hamiltonian H_{mf} from the original Hamiltonian and component function by performing a partial trace on all but a target particle. In the distinguishable case, this will result in N uncoupled Hamiltonians (one for each particle) and in the indistinguishable case, a single mean-field Hamiltonian, which is the Fock matrix for electronic systems. The maximum eigenvalue of the mean-field Hamiltonian is easily found, and we define $\Delta_{mf} = E^{mf}_{\max} - E(0)$, with the constant shift set to the expectation value of the energy. From this, the value of the timestep is set to

$$d\tau = \frac{1.8}{\Delta_{mf}}. \tag{7}$$

We note that more efficient and adaptive schemes are possible for timestep selection in problems of this type, however these were not utilized in the current work.

Once the timestep has been selected, the algorithm propagates forward through imaginary time with the orthogonal matching pursuit algorithm, with a termination threshold based on the energy. For computational practicality, we thus set several threshold values. The maximum number of component functions allows, $N_{\text{c-max}}$, the maximum imaginary time $\tau_{\max}$, and the minimal improvement in energy ϵ_E. We also introduce the function N that counts the number of components present in a wavefunction. This procedure is detailed in Alg. 1.

```
$\epsilon_E \leftarrow 0$
while $\tau < \tau_{\max}$ do
  $\tau' \leftarrow \tau + d\tau$
  $i \leftarrow 0$
  while $i < N_{\text{c-max}}$ do
    Find $|\Phi^i(\tau')\rangle$ satisfying eq 3 on state $G(\lambda)\,|\tilde{\Psi}(\tau)\rangle$
    Determine $c_i(\tau')$ via eq 4
    $|\tilde{\Psi}(\tau')\rangle_i \leftarrow \sum_j^i c_j\,|\Phi^j(\tau')\rangle$
    Calculate $\Delta = (\langle\tilde{\Psi}(\tau')|_i\, H\,|\tilde{\Psi}(\tau')\rangle_i - E(\tau))$
    if $\Delta < -d\tau\epsilon_E$ then break
    end if
    $i \leftarrow i + 1$
  end while
  if $i = N_{\text{c-max}}$ and $\Delta > -d\tau\epsilon_E$ then break
  else if $i > N(|\tilde{\Psi}(\tau)\rangle)$ then $\epsilon_E \leftarrow \max(\Delta/(ed\tau), 10^{-7})$
  end if
  $\lambda \leftarrow \langle\tilde{\Psi}(\tau')|\, H\,|\tilde{\Psi}(\tau')\rangle$
  $\tau \leftarrow \tau + d\tau$
end while
```

ALGORITHM 1: NOMAGIC Algorithm

Note that the update for the threshold value $\epsilon_E \leftarrow \max(\Delta/(ed\tau)), 10^{-7}$, where e is Euler's number, is based on a heuristic that if perfect evolution was achieved, the energy would decay to the ground state exponentially in imaginary time. At the termination of this imaginary time procedure, a final variational relaxation is performed on the wavefunction $|\tilde{\Psi}(\tau)\rangle$ with respect to both component functions $|\Phi_i(\tau)\rangle$ and coefficients $c_i(\tau)$ to relax the greedy constraint on the fitting procedure. It is known that direct minimization of canonical tensor decompositions can suffer from numerical issues if care is not taken to constraint the length of the individual components.[4] In particular, the space of canonical rank$-k$ decompositions is not closed, however the addition of a constraint on the norm of components remedies this situation.[6] In practice, we find a loose penalty term sufficient to enforce this constraint and mitigate the potential numerical difficulties from this problem without introducing the complexities of sophisticated constrained optimizations. Specifically we variationally minimize the auxiliary functional

$$\mathcal{L} = \frac{\langle \tilde{\Psi}(\tau) | \, H \, | \tilde{\Psi}(\tau) \rangle - \gamma (\max(0, \sum_i \langle \Phi^i(\tau) | \Phi^i(\tau) \rangle - D))^2}{\langle \tilde{\Psi}(\tau) | \tilde{\Psi}(\tau) \rangle} \tag{8}$$

where D controls the maximum length of components and γ is the penalty parameter. In this work we choose $D = 4.0$ and $\gamma = 1.0$, however little dependence is observed in the final result on these parameters unless extreme values are taken. Note that despite the presence of the max() function, this penalty term is differentiable and introduces no substantial additional difficulty in implementation.

1.3 Electronic Wavefunction Parameterization

Here we detail the electronic wavefunction parametrization used in this work, as well as the expressions used for the implementation of orthogonal matching pursuit and variational relaxation in electronic systems.

In quantum chemistry, frequently one first chooses a suitable single particle spin-orbital basis for the description of the electrons, which we denote $\{|\phi_i\rangle\}$. This basis typically consists of atom-centered contracted Gaussian type functions with a spin function, and are in general non-orthogonal such that they have an overlap matrix defined by

$$S_{ij} = \langle \phi_i | \phi_j \rangle \tag{9}$$

Linear combinations of these atomic orbitals are used to form molecular orbital functions

$$|\chi_m\rangle = \sum_i c_m^i \, |\phi_i\rangle \tag{10}$$

which have an inner product

$$\langle\chi_m|\chi_n\rangle = \sum_{i,j} c_m^{i*} c_n^j \langle\phi_i|\phi_j\rangle = \sum_i c_m^{i*} c_n^j S_{ij} \tag{11}$$

In our implementation, the $N-$electron component wavefunctions may be formed from the anti-symmetrized $N-$fold product of molecular orbital functions, also known as Slater determinants.

$$|\Phi^k\rangle = \mathcal{A}\left(|\chi_0^k\rangle\,|\chi_1^k\rangle\,...\,|\chi_{N-1}^k\rangle\right) \tag{12}$$

where $\mathcal{A}$ is the anti-symmetrizing operator. A convenient computational representation of an anti-symmetric component function $|\Phi^k\rangle$ is given by the coefficient matrix

$$T^K = \left(c_0^K|c_1^K|...|c_{N-1}^K\right) \tag{13}$$

which denotes an $M \times N$ matrix whose m'th column are the coefficients defining the m'th molecular orbital $|\chi_m^k\rangle$. This yields a convenient construction for the overlap between two component functions

$$\langle\Phi^K|\Phi^L\rangle = M_{KL} = \det\left(V_{KL}\right) = \det\left(T^{K\dagger} S T^L\right) \tag{14}$$

One quantity of convenience is the so-called transition density matrix defined between determinants K and L

$$P^{KL} = T^K \left(T^{L\dagger} S T^K\right)^{-1} T^{L\dagger} \tag{15}$$

Hamiltonian matrix elements may be written as

$$H_{KL} = M_{KL} \left(\mathrm{Tr}\left[P^{KL} \hat{h} \right] + \frac{1}{2} \mathrm{Tr}\left[P^{KL} G^{KL} \right] \right) \tag{16}$$

where $\hat{h}$ are the single electron integrals,

$$h_{\mu\nu} = \int d\sigma \; \phi_\mu^*(\sigma) \left(-\frac{\nabla_r^2}{2} - \sum_i \frac{Z_i}{|R_i - r|} \right) \phi_\nu(\sigma) \tag{17}$$

$$\tag{18}$$

where $\sigma = (r, s)$ denotes electronic spatial and spin variables and the nuclear positions and charges are R_i and Z_i. G^{KL} is given by

$$G_{\mu\nu}^{KL} = \left(\sum_{\lambda\sigma} P_{\lambda\sigma}^{KL} (g_{\mu\nu\lambda\sigma} - g_{\mu\lambda\nu\sigma}) \right) \tag{19}$$

with the two electron integrals $g_{\mu\nu\lambda\sigma}$

$$g_{\mu\nu\lambda\sigma} = \int d\sigma_1 \; d\sigma_2 \; \frac{\phi_\mu^*(\sigma_1)\phi_\nu(\sigma_1)\phi_\lambda^*(\sigma_2)\phi_\sigma(\sigma_2)}{|r_1 - r_2|} \tag{20}$$

From the description of orthogonal matching pursuit, we see that to utilize non-linear optimization of the component functions $|\Phi^k\rangle$ with analytic first derivatives, one needs the variations of H_{KL} and M_{KL} with respect to T^K. Allowing variations δT^K, the required expressions in the non-orthogonal spin orbital basis are as follows:

$$\delta M_{KL} = M_{KL}\,\mathrm{Tr}\left[ST^{L}(V^{KL})^{-1}\delta T^{K\dagger}\right] \tag{21}$$

$$\delta P^{KL} = [1 - P^{KL}S]\delta T^{K}(V^{KL\dagger})^{-1}T^{L\dagger} \tag{22}$$

$$\delta G^{KL}_{\mu\nu} = \left(\sum_{\lambda\sigma} \delta P^{KL}_{\lambda\sigma}(g_{\mu\nu\lambda\sigma} - g_{\mu\lambda\nu\sigma})\right) \tag{23}$$

$$A_{KL} = \mathrm{Tr}\left[P^{KL}G^{KL}\right] \tag{24}$$

$$\delta A_{KL} = \mathrm{Tr}\left[(1 - SP^{KL\dagger})G^{KL\dagger}T^{L}(V^{KL})^{-1}\delta T^{K\dagger}\right] \tag{25}$$

One must take care in implementing this expression, as it is a special case of the adjugate relations that is only strictly valid when V^{KL} is non-singular. To use this expression in evaluating cases when V^{KL} is singular, techniques developed elsewhere utilizing the singular value decomposition of V^{KL} and exact interpolation can be used.[7] Note also that numerical simplifications are possible by explicitly considering spin (α, β) and noting that $T^{K} = T^{K\alpha} \oplus T^{K\beta}$. These reductions of the above equations are straightforward and we do not give them here.

1.4 Renormalization of determinants

The form of the functional used in all optimizations formally ensures their values are independent of total normalization of the wavefunction and normalization of individual columns defining the single particle portions of the wavefunctions. While this is true in exact arithmetic, there can be practical numerical issues if these values are allowed to become unbounded throughout the course of the simulation. For this reason, it is convenient to occasionally renormalize single particle functions as well as the total wavefunction. An efficient way to perform the renormalization at the level of a single determinant T with corresponding coefficient c defined on a non-orthogonal single particle basis with overlap matrix S is

$$\begin{aligned} Q, R &= \text{QRDecomp}(S^{1/2}T) \\ T' &= S^{-1/2}Q \\ c' &= \det(R)c \end{aligned} \tag{26}$$

where Q and R are the output from the well known QR decomposition of matrices, the columns of T' are orthonormal with respect to the overlap matrix S, and c' is its new coefficient in the wave function expansion. An alternative to this approach is to utilize an exponential parameterization of the coefficient space, which guarantees the preservation of normalization. The cost and benefits of using such a parameterization within this method are a subject of current research.

1.5 Nuclear Union Configuration Interaction

In this section we give some of the details of the nuclear union configuration interaction method used to improve the description of reaction coordinates. In the study of a set of related problems, such as set of electronic Hamiltonians differing only by the positions of the nuclei, one would like to describe each configuration with an equivalent amount of accuracy, to get the best relative features possible. In multi-reference methods, this is often done by selecting the same active space at each configuration, and rotating the orbitals and coefficients at each geometry accordingly. In the nuclear union configuration interaction method, we propose each reuse of the components(determinants) found locally at other geometries to give a totally identical variational space for all nuclear configurations. As the wavefunctions produced by the NOMAGIC method are especially compact, this introduces little extra overhead to the method as a whole.

Specifically, denote the component functions found at nuclear configuration R' with corresponding Hamiltonian $H(R')$ as $|\Phi^k_{R'}\rangle = |\Phi^i\rangle$ where i is now an index set variable that runs over all the component functions at all the geometries being considered. This could be a whole reaction coordinate, or simply neighboring points depending on computational restrictions or chemical/physical considerations. At each nuclear configuration R we find new coefficients $c_i(R)$ by solving

$$H(R)C = SCE \tag{27}$$

for its ground state eigenvector, and we define

$$H(R)_{ij} = \langle \Phi^i | \, H(R) \, | \Phi^j \rangle \tag{28}$$

$$S_{ij} = \langle \Phi^i | \Phi^j \rangle \tag{29}$$

Note that the overlap matrix may become singular, as configurations from nearby geometries are often very similar. This can be handled either through canonical orthogonalization[3] or by removing redundant configurations before attempting the diagonalization procedure. Moreover, one might expect that additional compression is possible in this space, and this is the subject of current research.

1.6 H_2

In this subsection we include an extra system to further understanding of the performance of the method of canonical test problems. Although the physical system is simple with respect to the systems considered in the main text, it remains illustrative of the method with its convergence towards the exact solution with respect to the number of determinants, and demonstrates how accurate the method can be numerically. The results of the calculation are shown in Fig. 1. In this case, we utilize the Dunning cc-pVTZ basis, and see that by 16 determinants the error in the energy is less than 10^{-10} E_h, which is far beyond the threshold for chemical accuracy.

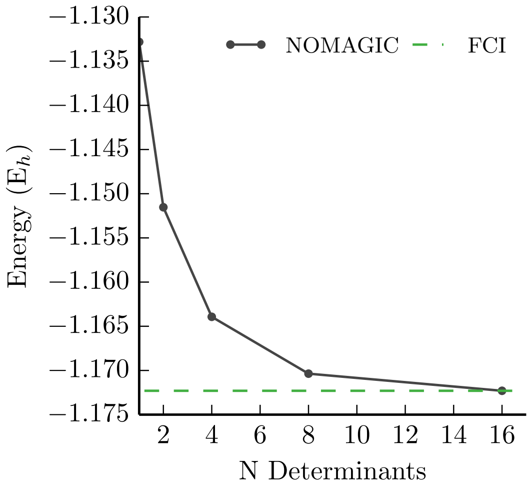

Figure 1: The ground state energy as a function of the number of determinants in the NOMAGIC expansion for H_2 at a separation of $R = 0.75\mathring{A}$ in a cc-pVTZ basis. The dashed line represents the exact solution within this basis, and we find rapid convergence with respect to the number of determinants. The error in the energy with respect to the exact solution in this basis by 16 determinants is less than 10^{-10} Hartrees

References

1. Tropp, J. A.; Gilbert, A. C. Signal recovery from random measurements via orthogonal matching pursuit. *IEEE Trans. Inf. Theory* **2007**, *53*, 4655–4666.

2. Nocedal, J.; Wright, S. J. *Numerical Optimization (2nd edition)*; Springer, 2006.

3. Helgaker, T.; Jorgensen, P.; Olsen, J. *Molecular Electronic Structure Theory*; Wiley, Sussex, 2002.

4. Kolda, T. G.; Bader, B. W. Tensor decompositions and applications. *SIAM review* **2009**, *51*, 455–500.

5. Trivedi, N.; Ceperley, D. M. Ground-state correlations of quantum antiferromagnets: A Green-function Monte Carlo study. *Phys. Rev. B* **1990**, *41*, 4552–4569.

6. Espig, M.; Hackbusch, W. A regularized Newton method for the efficient approximation

of tensors represented in the canonical tensor format. *Numerische Mathematik* **2012**, *122*, 489–525.

7. Amovilli, C. In *Quantum systems in chemistry and physics. trends in methods and applications*; McWeeny, R., Maruani, J., Smeyers, Y., Wilson, S., Eds.; Topics in Molecular Organization and Engineering; Springer Netherlands, 1997; Vol. 16; pp 343–347.