Adaptive Gaussian basis for (post) Hartree-Fock calculation

Jiaxi Zhao

December 4, 2024

Many-body Schrödinger equation

After Born–Oppenheimer approximation, the many-body Schrödinger equation is given by:

$$\hat{H}^e \Psi(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N) = E\Psi(\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_N),$$

$$\hat{H}^e = -\frac{1}{2}\Delta_{\mathbf{r}} - \sum_{i=1}^{N}\sum_{j=1}^{n_A} \frac{Z_j}{\|\mathbf{r}_i - \mathbf{R}_j\|_2} + \frac{1}{2}\sum_{i \neq j}^{N} \frac{1}{r_{ij}}$$

$$= \sum_{i=1}^{N} \widehat{h}(i) + \frac{1}{2}\sum_{i \neq j}^{N} \frac{1}{r_{ij}}$$

Solving this equation, we can obtain the ground state energy, potential energy surface (for geometry optimization and transition states search), band structure for the solid state system, and etc.

Mean-field approximation: Slater determinant

Pauli exclusion principle requires following antisymmetric constraint on the many-body wave function:

$$\Psi_{\mathsf{HF}}(..., \mathbf{r}_j, ..., \mathbf{r}_i, ...) = -\Psi_{\mathsf{HF}}(..., \mathbf{r}_i, ..., \mathbf{r}_j, ...)$$

The many-body wave function of the Hartree-Fock theory is given by a single Slater determinant:

$$\Psi_{\mathsf{HF}}(\mathbf{r}_1, ..., \mathbf{r}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(\mathbf{r}_1) & \phi_1(\mathbf{r}_2) & \cdots & \phi_1(\mathbf{r}_N) \\ \phi_2(\mathbf{r}_1) & \phi_2(\mathbf{r}_2) & \cdots & \phi_2(\mathbf{r}_N) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_N(\mathbf{r}_1) & \phi_N(\mathbf{r}_2) & \cdots & \phi_N(\mathbf{r}_N) \end{vmatrix}$$

This part is still a bit confusing for math audience.

Total energy

$$E[\Psi^{HF}] = \left\langle \Psi^{HF}|\hat{H}^e|\Psi^{HF}\right\rangle = \sum_{i=1}^{N}\int d\mathbf{x}_i\phi_i^*(\mathbf{x}_i)\widehat{h}(i)\phi_i(\mathbf{x}_i)$$

$$+ \frac{1}{2N(N-1)}\sum_{i\neq j}^{N}\sum_{k\neq l}^{N}\iint d\mathbf{x}_i d\mathbf{x}_j\phi_k^*(\mathbf{x}_i)\phi_l^*(\mathbf{x}_j)\frac{1}{|\mathbf{x}_i-\mathbf{x}_j|}\phi_k(\mathbf{x}_i)\phi_l(\mathbf{x}_j)$$

$$- \frac{1}{2N(N-1)}\sum_{i\neq j}^{N}\sum_{k\neq l}^{N}\iint d\mathbf{x}_i d\mathbf{x}_j\phi_k^*(\mathbf{x}_i)\phi_l^*(\mathbf{x}_j)\frac{1}{|\mathbf{x}_i-\mathbf{x}_j|}\phi_k(\mathbf{x}_j)\phi_l(\mathbf{x}_i)$$

$$= E(\phi_1,\phi_2,\cdots,\phi_N).$$

Minimizing the total energy will give the ground state configuration which is consistent with the self-consisten field approach. The double-electron integrations $O(N^4)$ are the dominant part of computational cost.

Basis set[1]

Slater basis functions:

$$\varphi^{S}_{a,A,\alpha} \equiv (x - A_x)^{a_x}(y - A_y)^{a_y}(z - A_z)^{a_z} e^{-\alpha \|\mathbf{r}-\mathbf{A}\|_2}.$$

Gaussian basis functions:

$$\varphi^{G}_{a,A,\alpha} \equiv (x - A_x)^{a_x}(y - A_y)^{a_y}(z - A_z)^{a_z} e^{-\alpha \|\mathbf{r}-\mathbf{A}\|_2^2}.$$

Contracted Gaussian basis functions:

$$\varphi^{G}_{a,A,\alpha,k} \equiv \sum_{k=1}^{K_A} D_{aAk}(x - A_x)^{a_x}(y - A_y)^{a_y}(z - A_z)^{a_z} e^{-\alpha_k \|\mathbf{r}-\mathbf{A}\|_2^2}.$$

Gaussian Lobe functions:

$$\varphi^{G}_{a,A,\alpha,k} \equiv \sum_{k=1}^{K_A} D_{aAk} e^{-\alpha \|\mathbf{r}-\mathbf{A}\|_2^2}.$$

---

[0]Gill, Peter MW. "Molecular integrals over gaussian basis functions." Advances in quantum chemistry. Vol. 25. Academic Press, 1994. 141-205.

A primer on molecular integral over Gaussian basis functions

It suffices to evaluate the fundamental integrals:

$$\iint e^{-\alpha\|\mathbf{r}_1-\mathbf{A}\|_2^2} e^{-\beta\|\mathbf{r}_2-\mathbf{B}\|_2^2} f(r_{12}) e^{-\gamma\|\mathbf{r}_1-\mathbf{C}\|_2^2} e^{-\delta\|\mathbf{r}_2-\mathbf{D}\|_2^2} d\mathbf{r}_1 d\mathbf{r}_2.$$

Integrations involving higher angular momentum can be calculated by taking the derivation w.r.t. $\mathbf{r}_1, \mathbf{r}_2$ and recursive formula.

Caveat: the recursive formula becomes very complex for higher angular momentum and thus computationally expensive.

Basis set

In this work, we propose to use the adaptive Gaussian basis with tunable mean and covariance:

$$\phi_i = \sum_{j \in I} c_{ij} \mathcal{N}(\mathbf{r}; \mu_j, \Sigma_j).$$

In general, all the optimizable parameters are $\mu \in \mathbb{R}^{|I| \times 3}, \Sigma \in \left(\mathbb{S}_+^3\right)^{|I|}, C \in \mathbb{C}^{|I| \times N}$. $|I|$ is the size of the basis set, N is the number of electrons (ground state + excited state).

---

[1]Kerbl, Bernhard, et al. "3D Gaussian Splatting for Real-Time Radiance Field Rendering." ACM Trans. Graph. 42.4 (2023): 139-1.

orthognormality and overlap matrix

   The orthgnormality of the orbitals in SCF is guaranteed by the eigensolver. In total energy minimization, we have to handle this explicitly.

   As the orbitals are supposed to be orthognormal, it can be written as

$$C^\dagger S C = I.$$

where S is the overlap matrix defined as:

$$S_{ij} = \langle \mathcal{N}\left(\mathbf{r}; \mu_\mathbf{i}, \Sigma_i\right)| \, \mathcal{N}\left(\mathbf{r}; \mu_\mathbf{j}, \Sigma_j\right)\rangle.$$

The orthognormality is enforced by QR decomposition modified by the Cholesky factorization of S.

Coulomb and external energy

Our primary idea is to approximate the Coulomb kernel by a series of Gaussian modes:

$$\frac{1}{r} = \sum_i c_i e^{-\alpha_i r^2},$$

which is done by minimizing the following

$$\min_{c_i, \alpha_i} \int_{B_M} \left(\frac{1}{|\mathbf{r}|} - \sum_i c_i e^{-\alpha_i |\mathbf{r}|^2} \right)^2 d\mathbf{r} = 4\pi \int_0^M \left(1 - \sum_i c_i r e^{-\alpha_i r^2} \right)^2 dr$$

Then we fix the number of Gaussian modes and optimize the analytic loss function via gradient-based optimization.
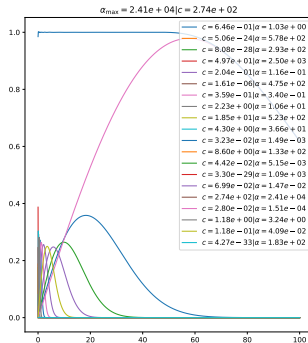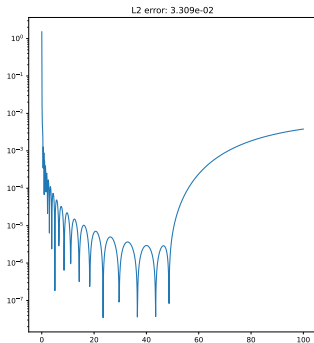
Analysis of the nonlinear optimization problem

$$\int_0^M \left(1 - \sum_i c_i r e^{-\alpha_i r^2}\right)^2 dr \overset{u=kr}{=} \frac{1}{k}\int_0^{kM}\left(1 - \sum_i \frac{c_i}{k}u e^{-\frac{\alpha_i}{k^2}u^2}\right)^2 du.$$

The optimal solution for different M can be obtained by rescaling the optimal coefficient and variance of the Gaussian modes.

However, optimizing such problem to a relatively low threshold, e.g. 10^{-5} is quite difficult.
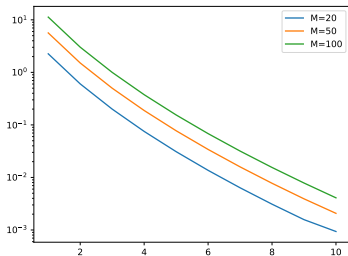
Coulomb kernel decomposition

More on double electron integrals

$$\langle \mathcal{N}(\mathbf{r}_1|\mu_1, \Sigma_1)| \frac{1}{|\mathbf{r}_1 - \mathbf{r}_2|} |\mathcal{N}(\mathbf{r}_2|\mu_2, \Sigma_2)\rangle$$

$$= \int \int d\mathbf{r}_1 d\mathbf{r}_2 \frac{1}{\sqrt{(2\pi)^6 \det(\Sigma_1) \det(\Sigma_2)}}$$

$$\sum_i c_i \exp\left[-\frac{1}{2} \left((\mathbf{r}_1 - \mu_1)^T \Sigma_1^{-1}(\mathbf{r}_1 - \mu_1) \right. \right.$$

$$\left. \left. + (\mathbf{r}_2 - \mu_2)^T \Sigma_2^{-1}(\mathbf{r}_2 - \mu_2)) - (\mathbf{r}_1 - \mathbf{r}_2)^T \alpha_i \mathsf{I}(\mathbf{r}_1 - \mathbf{r}_2) \right]$$

$$= \int \int d\mathbf{r} \frac{1}{\sqrt{(2\pi)^6 \det(\Sigma_1) \det(\Sigma_2)}}$$

$$\sum_i c_i s(i) \exp\left[-\frac{1}{2}(\mathbf{r} - \mu(i))^T \begin{pmatrix} \Sigma_1^{-1} + 2\alpha_i \mathsf{I} & -2\alpha_i \mathsf{I} \\ -2\alpha_i \mathsf{I} & \Sigma_2^{-1} + 2\alpha_i \mathsf{I} \end{pmatrix} (\mathbf{r} - \mu(i)) \right].$$

Potential improvement

1. Using greedy algorithm to optimize for more Gaussian modes, i.e. one can use the optimization results for less modes as the warm-start for more modes.

2. The approximation interval $[0, M]$ can be done adaptively according to the relative position of two Gaussian modes.

3. Motivated by the Boys function, can we design other decomposition of the Coulomb kernel?

$$\frac{1}{r} = C \int_{\mathbb{R}} e^{-r^2 t^2} dt.$$

Parameterization of Gaussian basis

We parameterize the Gaussian basis as

$$\Sigma_1 = U_1 D_1 U_1^T, \quad \Sigma_2 = U_2 D_2 U_2^T,$$

where U_i is obtained from the QR decomposition and the positivity of the diagonal element of D_i is enforced by the softplus function.

This is proved to be more robust than parametrization by Cholesky factorization.

Numerical experiments

	e_kin + e_ext	e_kin	e_ext	e_nuc	e_coul + e_exc	e_coul	e_exc	e_tot
OUR (H_2), 10	-2.5062	1.1254	-3.6316	0.7138	0.6591	1.3182	-0.6591	-1.1334
HF (STO-3G, 2)	-2.5049	*	*	0.7138	0.6745	*	*	-1.1167
HF (6-31G, 4)	-2.4902	*	*	0.7138	0.6497	*	*	-1.1267
HF (6-311G, 6)	-2.4924	*	*	0.7138	0.6507	*	*	-1.1280
OUR (CH_4), 22	-79.6713	40.0987	-119.7700	13.4477	26.1010	32.6936	-6.5926	-40.1225
HF (STO-3G, 9)	-79.3617	*	*	13.4477	26.1872	*	*	-39.7267
HF (6-31G, 17)	-79.6901	*	*	13.4477	26.0619	*	*	-40.1804
HF (6-311G, 25)	-79.6872	*	*	13.4477	26.0515	*	*	-40.1880
OUR (H_2O), 22	-122.9802	75.9462	-199.0371	9.1895	37.8786	46.8579	-8.9601	-76.0036
HF (STO-3G, 7)	-122.3614	*	*	9.1895	38.2089	*	*	-74.9630
HF (6-31G, 13)	-122.9701	*	*	9.1895	37.7966	*	*	-75.9839
HF (6-311G, 19)	-123.0146	*	*	9.1895	37.8157	*	*	-76.0094

Next step

1. Compare the scalability of the double-electron integral calculation over the number of basis functions with the SOTA method, e.g. MD, HGP(OS), Rys.

2. Conduct numerical experiments to check if the adaptive Gaussian basis can achieve the same accuracy with a smaller number of basis.

3. Implement post HF method, e.g. CCSD, CCSD(T).

Q & A

SCF

Self-consistent field approach to HF: adding Lagrangian multiplier (molecular orbital energy) and the KKT condition translates to an eigenvalue problem:

$$\delta E[\phi_k^*(x_k)] = \delta E(\phi_1, \phi_2, \cdots, \phi_N) - \delta \left[\sum_{i=1}^{N} \sum_{j=1}^{N} \lambda_{ij} \left(\langle \phi_i, \phi_j \rangle - \delta_{ij} \right) \right]$$

$$\widehat{F}(k)\phi_k(\mathbf{x}_k) \equiv \left[\hat{h}(k) + \hat{J}(k) - \hat{K}(k) \right] \phi_k(\mathbf{x}_k) = \epsilon_k \phi_k(\mathbf{x}_k),$$

$$\widehat{J}(k) \equiv \sum_{j=1}^{N} \int d\mathbf{x}_j \frac{\phi_j^*(\mathbf{x}_j)\phi_j(\mathbf{x}_j)}{|\mathbf{x}_k - \mathbf{x}_j|} = \sum_{j=1}^{N} \int d\mathbf{x}_j \frac{\rho(\mathbf{x}_j)}{|\mathbf{x}_k - \mathbf{x}_j|},$$

$$\widehat{K}(k)\phi_k(\mathbf{x}_k) \equiv \sum_{j=1}^{N} \phi_j(\mathbf{x}_k) \int d\mathbf{x}_j \frac{\phi_j^*(\mathbf{x}_j)\phi_k(\mathbf{x}_j)}{|\mathbf{x}_k - \mathbf{x}_j|}.$$

Motivation

The effect cutoff used by the plane-wave method to solve large system is unaffordable. Moreover, given the nature of the electron density which will be concentrated around the nucleus, while away from the nucleus which is the case of majority of the computational region, the electron density is relatively smooth, the plane-wave method is not efficient. We need to survey the literature for some multiscale basis such as the wavelet basis.

Literature

In literature, there has been lots of work on exploring the adaptive basis for solving DFT.

Use fixed analytic transformation mapping depending on the position of the atom. (Similar to the r type mesh refinement in numerical PDE.)

Solve some OT-type problems to predetermine the basis before the SCF loop.

However, our method is different from the above method in several aspects:

Our basis is continuously optimized during the optimization porcess of DFT.

We use Monte-Carlo method to estimate the density related energy, differs from the traditional PDE solving method.

Advantages

There are several advantages of using the distorted plane-wave method.

The orthognormality of the basis is preserved.

Calculation of the density related energy can be estimated efficiently by the Monte-Carlo method instead of solving the PDE.

References

📄 Gill, Peter MW.
Molecular integrals over gaussian basis functions.
Advances in quantum chemistry. Vol. 25. Academic Press, 1994. 141-205.

📄 Kerbl, Bernhard, et al
3D Gaussian Splatting for Real-Time Radiance Field Rendering
ACM Trans. Graph. 42.4 (2023): 139-1.

📄 S. Arisaka and Q. Li (2024)
Accelerating Legacy Numerical Solvers by Non-intrusive Gradient-based
Meta-solving
International Conference on Machine Learning 2024