arXiv:2304.04260v2 [physics.comp-ph] 4 May 2023

基于原子簇展开的多电子¨薛定谔方程的多层次方法*

周德轩†、陈华杰‡、何卓轩§、和克里斯托夫·奥特纳¶

摘要。 原子簇展开（ACE） *(Drautz, 2019)* 为对称多项式提供了一种高效且易于解读的参数化方法，在模拟多粒子系统的性质方面取得了巨大成功。在本研究中，我们将 ACE 框架的实际适用范围扩展至多电子波函数的计算。为此，我们开发了一种定制的变分蒙特卡洛算法，利用 ACE 波函数的稀疏性和层次结构。我们通过一系列针对一维系统的概念验证应用，展示了该方法的可行性。

关键词。 多电子薛定谔方程，变分蒙特卡罗，原子簇展开，级联多层方法

MSC 分类号。 81-08, 65C05, 65N25

1. 引言。 计算多体薛定谔¨方程的基态，可以说是电子结构计算中最根本的问题。同时，由于两个原因，这也是最具挑战性的计算任务之一：第一，未知波函数是所有电子坐标的函数，并且由于电子间相互作用，通常无法因式分解。第二，要获得有关所关注系统的有意义信息，通常需要非常高的精度。实际中采用的主要方法在计算成本与准确度之间各有取舍：(i) 密度泛函理论 [39]能够处理数千个粒子，但对许多关注的系统而言精度不足，尤其是强关联系统；(ii) 后哈特里-福克方法 [28]能够准确描述量子系统，但“黄金标准”方法 CCSD(T) 和 MP2(4) 对大型系统而言仍然成本高得令人望而却步；(iii) 量子蒙特卡洛方法 [3] 通过随机算法近似求得基态，是本文工作的重点。具体而言，我们将介绍一种新的变分蒙特卡洛（VMC）方法变体，采用原子簇展开 [12, 14] 对波函数进行参数化，正如 [13], 中所提出的，并采用级联多层方法计算基态。

变分蒙特卡洛方法将量子期望值表示为一系列构型的形式，这些构型按照试探波函数的分布概率进行分布。通过优化试探波函数以最小化能量期望值，可以获得基态解。变分蒙特卡洛方法的准确性和效率取决于波函数试探形式、

*2023年5月5日**资助：** DZ和HC的工作得到了国家重点研发计划（编号：2020YFA0712900）和国家自然科学基金（编号：NSFC11971066）的支持。

†北京师范大学数学科学学院，中国北京 100875（zhoudex-uan@mail.bnu.edu.cn）。

‡北京师范大学数学科学学院，中国北京 100875（chen.huajie@bnu.edu.cn）。§加拿大不列颠哥伦比亚省温哥华市数学路1984号，不列颠哥伦比亚大学数学系，V6T 1Z2（jerryho528@math.ubc.ca）。

¶数学系，不列颠哥伦比亚大学，加拿大卑诗省温哥华 Mathematics Road 1984 号，邮编 V6T 1Z2（ortner@math.ubc.ca）。

采样和参数优化算法。变分蒙特卡洛中最常用的试探波函数是 Jastrow-Slater波函数 [49]，尽管其精度受到根本性限制，但它们已在许多应用中取得了成功。

近年来，以神经网络架构对波函数进行参数化，极大拓展了变分蒙特卡洛算法 [27, 29, 42, 43] 可处理的波函数设计空间。采用深度神经网络架构使参数数量远超以往的研究，但也让变分蒙特卡洛方法达到了前所未有的精度；例如 NetKet [8, 22], PauliNet [29]和 FermiNet [42] 架构。这些方法的成功可归因于神经网络的表达能力，以及这些研究所采用的反对称试探形式。

本工作建立在采用机器学习方法的思想之上，但我们不使用深度神经网络，而是利用了原子簇展开（ACE）模型的思想 [12, 14]，该模型最初在 [13]中针对波函数提出。从本质上讲，这一试探形式以高度稀疏化的对称多项式取代了PauliNet [29] 中采用的 $S(N)$-等变深度神经网络。ACE模型最初是为参数化原子间势而开发的，在这一领域中，它能够产生精确且高效的模型，其性能可与最先进的机器学习技术相媲美 [37]。作为一种稀疏多项式展开，ACE模型相较于神经网络模型的优势在于，它具有清晰且可解释的近似参数：基的选择、多项式次数以及稀疏化模式。多项式次数使得对收敛性进行理论和数值研究成为可能，而稀疏化模式似乎与多电子系统中的相关性有关。我们在更大的 [13] 设计空间内，分别基于范德蒙德试探形式和回流试探形式，提出并研究了两种具体的参数化方案。这两种波函数模型都有一个有趣的性质，即它们具有层次结构。本工作的第二项贡献是利用这一模型层次结构，设计了一种级联多层变分蒙特卡洛算法，从而显著加速并稳定了参数优化过程。

在本工作中，我们致力于通过开发一种稳健且高性能的多层变分蒙特卡洛算法，证明使用 ACE参数化在原则上的可行性。为此，我们将数值测试限制在一维电子和软库仑相互作用这一高度简化的设置中。对现实三维系统的应用将在另一项工作中研究。由于这一限制，我们还将对波函数参数化的描述限定为一维粒子，但需要指出的是，至少在概念上，推广到任意维度是直接的 [13]。

提纲。 本文其余部分的组织结构如下：在 第2 节中，我们回顾多电子薛¨罗丁格方程和变分蒙特卡洛方法的相关背景。在 第3节中，我们回顾原子簇展开模型，然后分别采用范德蒙德变换和回流变换，构造两种反对称波函数参数化。在 第4节中，我们回顾变分蒙特卡洛方法，并介绍一种级联多层变分蒙特卡洛算法。在 第 5 节中，我们针对一系列模型问题进行数值实验。

记号。在本文中，我们将用 $\#A$ 表示集合 A 的基数。设 $\mathbb{Z}_m := \{\sigma_1, \cdots \sigma_m\}$ 是一个基数为 m、并赋有序关系 $\sigma_1 < \cdots < \sigma_m$ 的离散集。特别地，我们将自旋量子数集合记为 $\mathbb{Z}_2 := \{\uparrow, \downarrow\}$，并赋有序关系 $\downarrow < \uparrow$。对于任意一对 $\nu_1 = (k_1, s_1), \nu_2 = (k_2, s_2) \in \mathbb{N} \times \mathbb{Z}_m$，我们按字典序定义 $\nu_1 \le \nu_2$，即 $\nu_1 < \nu_2$

当且仅当 $k_1 < k_2$ 或 $k_1 = k_2, s_1 < s_2$。对于 $n \in \mathbb{N}_+$，我们随后将由 $(\mathbb{N}\times\mathbb{Z}_m)^n_{\mathrm{ord}}$ 表示的所有有序 n-元组的集合记为，即，

$$(\mathbb{N}\times\mathbb{Z}_m)^n_{\mathrm{ord}} := \Big\{\boldsymbol{\nu} = (\nu_1,\cdots,\nu_n) \in (\mathbb{N}\times\mathbb{Z}_m)^n :\ \nu_1 \le \nu_2 \le \cdots \le \nu_n\Big\}. \tag{1.1}$$

2.多体薛定¨谔方程。 我们考虑含有 M 个原子核和 N 个电子的一维多粒子系统。令 $r_i \in \mathbb{R}$, $\sigma_i \in \mathbb{Z}_2$ $(i = 1,\cdots,N)$表示电子的位置和自旋坐标。系统的电子态由多电子波函数 $\Psi(\boldsymbol{x})$描述，其N电子构型为 $\boldsymbol{x} = (x_1,\cdots,x_N)$，且 $x_i = (r_i,\sigma_i) \in \mathbb{R}\times\mathbb{Z}_2$。波函数 Ψ 必须满足 反对称性条件

$$\Psi(x_{\pi(1)},\cdots,x_{\pi(N)}) = (-1)^{\epsilon_\pi}\Psi(x_1,\cdots,x_N) \tag{2.1}$$

对于任意置换 π，其中 ϵ_π 是 π的奇偶性。如果 $(-1)^{\epsilon_\pi}$ 中没有符号（2.1），则称其为对称条件。我们将N电子波函数的可容许类记为

$$\mathscr{X}_N := \Big\{\Psi \in L^2\big((\mathbb{R}\times\mathbb{Z}_2)^N,\mathbb{R}\big) : \nabla_{r_i}\Psi \in L^2\big((\mathbb{R}\times\mathbb{Z}_2)^N,\mathbb{R}\big), \Psi \text{ is anti-symmetric}\Big\},$$

这里我们仅考虑取实值的波函数。由于我们将通过最小化瑞利商来求得基态波函数（见（2.5）），因此忽略了归一化约束 $\|\Psi\|_{L^2} = 1$。

令 $R_I \in \mathbb{R}$和 $Z_I \in \mathbb{Z}_+$$(I = 1,\ \cdots,\ M)$分别表示原子核的位置和原子序数。在 Born–Oppenheimer 近似下，电子系统的非相对论哈密顿算符是作用于 $\mathscr{X}_N$上的厄米算符，定义为

$$\mathcal{H} := -\frac{1}{2}\sum_{i=1}^N \nabla_{r_i}^2 + \sum_{i=1}^N v_{\mathrm{ext}}(r_i) + \sum_{1\le i<j\le N} v_{\mathrm{ee}}\big(|r_i - r_j|\big), \tag{2.2}$$

其中 $v_{\mathrm{ext}}:\ \mathbb{R}\to\mathbb{R}$ 表示原子核产生的吸引作用，而 $v_{\mathrm{ee}}:\ \mathbb{R}\to\mathbb{R}$表示电子间的排斥作用。在本文中，我们采用软库仑相互作用 [9, 15, 50]，而非裸库仑相互作用；电子间排斥和原子核对电子的吸引分别表示为

$$v_{\mathrm{ee}}(u) = \frac{1}{\sqrt{1+u^2}} \qquad \text{and} \qquad v_{\mathrm{ext}}(r) = -\sum_{I=1}^M Z_I v_{\mathrm{ee}}(|r - R_I|). \tag{2.3}$$

在此模型中，基态波函数不会出现三维库仑体系中存在的尖点。然而，三维库仑体系中由长程相互作用和强关联引起的困难，在这些一维模型体系中同样存在 [48, 54]。

通过求解多电子定态薛定谔方程，可以确定该系统的电子基态 Ψ_0

$$\mathcal{H}\Psi_0 = E_0\Psi_0, \tag{2.4}$$

其中，基态能量 E_0 是哈密顿算符的最低特征值，使得相应的特征函数为 $\Psi_0 \in \mathscr{X}_N$。基态解还可以

可通过对可容许的波函数最小化能量泛函（或瑞利商）而获得

$$E_0 = \min_{\Psi \in \mathscr{X}_N} E(\Psi) \qquad \text{with} \quad E(\Psi) = \frac{\langle \Psi | \mathcal{H} | \Psi \rangle}{\langle \Psi | \Psi \rangle} = \frac{\int \Psi(\boldsymbol{x}) \big(\mathcal{H}\Psi\big)(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x}}{\int |\Psi(\boldsymbol{x})|^2 \, \mathrm{d}\boldsymbol{x}}, \tag{2.5}$$

其中我们使用了狄拉克括号；对电子构型 $\boldsymbol{x}$ 的积分 $\int_{\boldsymbol{x} \in \left(\mathbb{R} \times \mathbb{Z}_2\right)^N}$表示为：

$= \sum_{\sigma_1 \in \mathbb{Z}_2} \cdots \sum_{\sigma_N \in \mathbb{Z}_2} \int_{\mathbb{R}^3} \mathrm{d}r_1 \cdots \int_{\mathbb{R}^3} \mathrm{d}r_N$。

在变分蒙特卡洛（VMC）方法中， [3], 能量泛函被改写为

$$E(\Psi) = \int \frac{\big(\mathcal{H}\Psi\big)(\boldsymbol{x})}{\Psi(\boldsymbol{x})} \frac{|\Psi(\boldsymbol{x})|^2}{\langle \Psi | \Psi \rangle} \, \mathrm{d}\boldsymbol{x} =: \int E_{\mathrm{L}}(\boldsymbol{x}; \Psi) P(\boldsymbol{x}; \Psi) \, \mathrm{d}\boldsymbol{x} = \mathbb{E}_{\boldsymbol{x} \sim P(\cdot; \Psi)} \big[E_{\mathrm{L}}(\boldsymbol{x}; \Psi) \big], \tag{2.6}$$

其中，概率密度 $P(\boldsymbol{x}; \Psi)$ 和局部能量 $E_{\mathrm{L}}(\boldsymbol{x}; \Psi)$ 分别定义为

$$P(\boldsymbol{x}; \Psi) := \frac{\big|\Psi(\boldsymbol{x})\big|^2}{\int \big|\Psi(\boldsymbol{x})\big|^2 \, \mathrm{d}\boldsymbol{x}} \qquad \text{and} \tag{2.7}$$

$$E_{\mathrm{L}}(\boldsymbol{x}; \Psi) := \frac{\big(\mathcal{H}\Psi\big)(\boldsymbol{x})}{\Psi(\boldsymbol{x})} = -\frac{1}{2} \sum_{i=1}^{N} \frac{\nabla_{r_i}^2 \Psi(\boldsymbol{x})}{\Psi(\boldsymbol{x})} + \sum_{i=1}^{N} v_{\mathrm{ext}}(r_i) + \sum_{1 \le i < j \le N} v_{\mathrm{ee}}(|r_i - r_j|).$$

VMC 表述（2.6）使我们能够通过马尔可夫链蒙特卡罗采样，估算式（2.5）中的能量泛函（一个高维积分）。

3. 波函数表示。 波函数的参数化方式多种多样，通常是通过对 $S(N)$等变张量进行变换来实现，例如经典的回流试探形式，最近在 [42]中重新受到关注。本节将回顾并聚焦于一种波函数设计空间，其中等变张量是通过原子簇展开（ACE）[2, 14, 37, 53]构造的。在 ACE 框架下，多粒子对称函数首先在标准张量积基中展开，随后根据物理原理进行大幅稀疏化。此外，该框架有效利用对称性，实现了高效的求值算法。我们将 [13] 中提出的庞大 ACE 波函数设计空间缩小为两种具体实用的架构，并在第 5 节的一系列初步基准测试中实现并检验这些架构。我们主要关注基于回流变换的方法，同时也将介绍一种采用范德蒙德行列式的简单变体，以作比较。

3.1.单粒子基。令$\Omega := \mathbb{R} \times \mathbb{Z}_m$ 表示一个抽象的单粒子构型空间，其中 $\mathbb{R}$ 为空间坐标的空间， $\mathbb{Z}_m$ 为“自旋”的集合。我们将用单粒子基的张量积来展开多体函数，即用一个基 $\{\phi_\nu\}$ 表示函数 $f : \Omega \to \mathbb{R}$。一般而言， ϕ_ν的选择具有很大的自由度。由于我们关注的是一维粒子和光滑波函数，因此选择单粒子基函数 ϕ_ν 如下：

$$\phi_\nu(x) := \phi_{(k,s)}(x) := \mathscr{L}_k\left(\frac{2}{\pi} \arctan\left(\frac{r}{L}\right)\right) \cdot \chi_s(\sigma) \quad \text{for } x = (r, \sigma) \in \mathbb{R} \times \mathbb{Z}_m, \tag{3.1}$$

其中，索引是一个元组 $\nu = (k, s)$，其中 $k \in \mathbb{N}$ 和 $s \in \mathbb{Z}_m$，$\mathscr{L}_k$ 是标准勒让德多项式，$L > 0$ 是一个长度缩放常数，其取值取决于原子分布所在区域的大小。最后，$\chi_s(\sigma) = \delta_{s\sigma}$ 是一个“独热嵌入”，也就是说，$\{\chi_s\}_{s\in\mathbb{Z}_m}$ 是离散集 $\mathbb{Z}_m$上的完备基。坐标变换 $r \mapsto \frac{2}{\pi}\arctan(r/L)$将空间 $\mathbb{R}$ 变换到勒让德多项式的定义域 $[-1, 1]$。由于这一变换，基的分辨率在原点附近最高；原子核将位于原点附近，因此该处需要较高的分辨率。

我们随后构造的多体波函数也将乘以形如 $\exp(-\theta\cdot\sqrt{1+r^2})$ 的包络函数，其中 θ 为一个参数，以确保其在无穷远处衰减。为展示变换和包络函数的效果，我们绘制了图 3.1中的前六个基函数。我们直观地确认，这种构造生成的基函数在“核心区域”振荡，并随着 $r \to \infty$平滑衰减，从而模拟电子结构计算中使用的原子轨道 [39]。

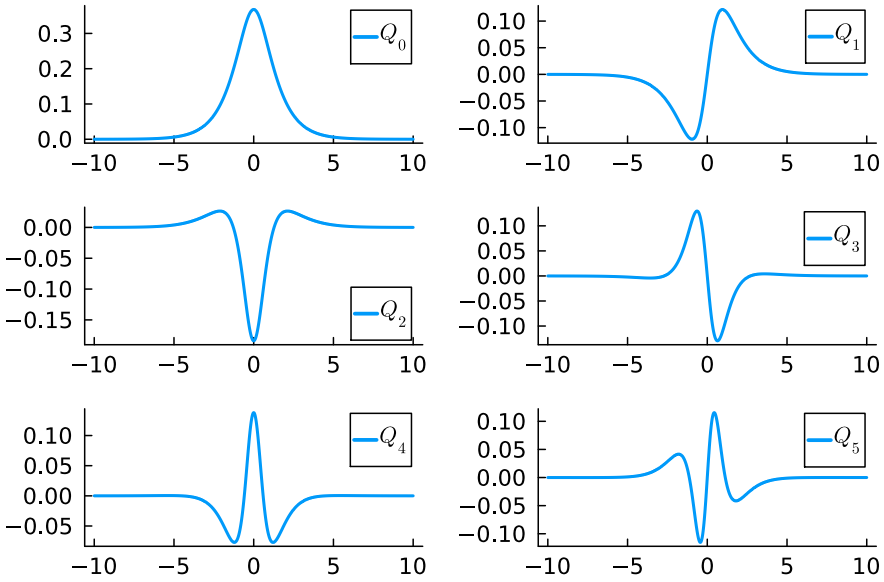

FIG. 3.1 *The first six position basis functions $Q_k = \mathscr{L}_k(2 \cdot /\pi) \cdot -\sqrt{1+r^2}$ il- .: 反正 lustrating the effect of the coordinate transform and envelope. The envelope is not part of the one-particle basis but is included here for illustration purposes.*

切0 exp( )

实际上，有时应将全局基函数（3.1）放置在每个原子位置 $R_I (I = 1, \cdots, M)$，从而得到一组新的单粒子基函数

$$\phi_{(I,k,s)}(x) := \mathscr{L}_k\left(\frac{2}{\pi}\arctan\left(\frac{r - R_I}{L}\right)\right) \cdot \chi_s(\sigma) \quad \text{for } x = (r, \sigma) \in \mathbb{R} \times \mathbb{Z}_m. \tag{3.2}$$

当原子分布在较大的区域内时，这种构造尤其有效；参见第 5 节中原子间距较大的氢链示例。为简化表述，下面我们将重点讨论单粒子基组（3.1），但所有讨论都可轻易推广至（3.2）。

“自旋”坐标对应的单粒子基函数将在其中发挥重要作用

，在构造反对称波函数时将发挥重要作用。我们取 $m = 2$，其中 $\mathbb{Z}_2 = \{\uparrow, \downarrow\}$ 表示自旋坐标（参见第 3.3和3.4节）。

最后，我们为每个单粒子基函数指定一个次数，这将为截断基提供一个自然的近似参数。由于每个 χ_s 都起着"等价"的作用，因此自然可以将基函数 ϕ_ν 与其空间分量的次数关联起来，即定义

$$\deg(\phi_\nu) = \deg\left(\phi_{(k,s)}\right) := k. \tag{3.3}$$

3.2. 稀疏对称多项式。接下来，我们将对 Ω^N 上的对称函数进行参数化，即函数 $f(x_1, \cdots, x_N)$，$x_j \in \Omega$，这些函数在任意置换 $\pi \in S(N)$ 下保持不变，

$$f(x_{\pi(1)}, \cdots, x_{\pi(N)}) = f(x_1, \cdots, x_N). \tag{3.4}$$

若不利用任何已有结构，我们可以直接将 f 展开为单粒子基的张量积之和，即

$$f \sim \sum_{\boldsymbol{\nu} \in (\mathbb{N}\times\mathbb{Z}_m)^N} c_{\boldsymbol{\nu}} \bigotimes_{t=1}^{N} \phi_{\nu_t}$$

我们将通过以下四个步骤，使这种方法即使在粒子数 N 很大时也能实现计算高效：(i) 利用对称性；(ii) 将基转换为 ACE 格式；(iii) 限制关联阶数；(iv) 总次数稀疏化。

(i) 对称基： 直观地，人们可能会通过对张量积进行对称化来为这一类构造基，

$$\phi^{\rm sym}_{\boldsymbol{\nu}} := \sum_{\pi \in S(N)} \bigotimes_{t=1}^{N} \phi_{\nu_{\pi(t)}} \qquad \text{for } \boldsymbol{\nu} \in (\mathbb{N} \times \mathbb{Z}_m)^N_{\rm ord}.$$

将我们的参数化限制为对称多项式后，基函数的数量会显著减少，因为现在只需考虑有序元组 $\boldsymbol{\nu}$ 。然而，由于对称化的 $O(N!)$ 成本，计算成本仍保持不变。

(ii)ACE 形式体系：ACE形式体系构造了一个与函数 $\phi^{\rm sym}_{\boldsymbol{\nu}}$ 张成相同空间的基：对于 $\boldsymbol{\nu} = (\nu_1, \ldots, \nu_N) \in (\mathbb{N} \times \mathbb{Z}_m)^N_{\rm ord}$ 我们定义

$$\begin{aligned} \boldsymbol{A}_{\boldsymbol{\nu}}(\boldsymbol{x}) = \boldsymbol{A}_{\nu_1,\ldots,\nu_N}(x_1, \cdots, x_N) &:= \prod_{t=1}^{N} A_{\nu_t}(x_1, \cdots, x_N), \qquad \text{where} \\ A_\nu(x_1, \cdots, x_N) &:= \sum_{n=1}^{N} \phi_\nu(x_n). \end{aligned} \tag{3.5}$$

A_ν最好被理解为幂和多项式的推广形式，幂和多项式是不变理论中的经典概念。这一构造将朴素对称化基的 $O(N!)$缩放成本替换为池化操作的 $O(N)$ 成本，以及通过递归求值算法 [2, 14, 32]对每个基函数的 $O(1)$ 成本。 [2,引理 2.1] 表明，该集合

$$\left\{ \boldsymbol{A}_{\boldsymbol{\nu}} : \boldsymbol{\nu} \in (\mathbb{N} \times \mathbb{Z}_m)^N_{\rm ord} \right\}$$

构成对称多项式与离散空间张量积空间的完备基。

(iii) 截断簇展开：我们最重要的稀疏化步骤具有物理动机，源于对簇展开进行截断。该步骤类似于经典的方差分析展开 [16, 25]。

为此，我们令 $0 < \mathcal{B} \leq N$，并通过类似的表达式，将 $\boldsymbol{A}_{\boldsymbol{\nu}}$ 的定义展开为索引元组$\boldsymbol{\nu} = (\nu_1, \ldots, \nu_{\mathcal{B}}) \in (\mathbb{N} \times \mathbb{Z}_m)^{\mathcal{B}}_{\rm ord}$，

$$\boldsymbol{A}_{\nu_1,\ldots,\nu_{\mathcal{B}}}(x_1, \cdots, x_N) := \prod_{t=1}^{\mathcal{B}} A_{\nu_t}(x_1, \cdots, x_N). \tag{3.6}$$

这些新的基函数仍是 $x_1, \ldots, x_N$的函数，但每次只包含至多涉及 $\mathcal{B}$ 个粒子的乘积。因此，我们称它们为 $\mathcal{B}$-关联。对称目标函数 f 的截断簇展开可形式化地写为

$$f \sim \sum_{\boldsymbol{\nu} \in (\mathbb{N}\times\mathbb{Z}_m)^{\mathcal{B}}_{\rm ord}} c_{\boldsymbol{\nu}} \boldsymbol{A}_{\boldsymbol{\nu}}. \tag{3.7}$$

将关联阶数上限设为 $\mathcal{B}$ ，可显著降低 -模型的维数。

为便于后续引用，我们将基函数$\boldsymbol{A}_{\boldsymbol{\nu}}$ 的空间关联阶数记为

$$\|\boldsymbol{\nu}\|_0 := \#\Big\{t \Big| \deg(\phi_{\nu_t}) \neq 0\Big\} \qquad \text{for } \boldsymbol{\nu} \in (\mathbb{N} \times \mathbb{Z}_m)^{\mathcal{B}}_{\rm ord}.$$

注释 3.1（簇展开与线性相关）。与（3.7）略有不同，且更符合以往关于原子簇展开的研究 [12, 14, 13]，簇展开也可以写作

$$f \sim \sum_{\ell=1}^{\mathcal{B}} \sum_{\boldsymbol{\nu} \in (\mathbb{N}\times\mathbb{Z}_m)^{\ell}_{\rm ord}} c_{\boldsymbol{\nu}} \boldsymbol{A}_{\boldsymbol{\nu}} \qquad \text{with } \boldsymbol{A}_{\boldsymbol{\nu}}(\boldsymbol{x}) = \boldsymbol{A}_{\nu_1,\ldots,\nu_\ell}(\boldsymbol{x}) := \prod_{t=1}^{\ell} A_{\nu_t}(\boldsymbol{x}) \tag{3.8}$$

其中基函数是 ℓ阶关联项，与 $\ell \leq \mathcal{B}$。需要注意的是，（3.6）和（3.8）都不是原始对称基（3.5）的子集（即N阶关联项）。不过，它们属于同一个对称多项式空间。其次，（ ℓ阶关联项中，不同阶数 ℓ 的（3.8）并不线性无关，因为

$$1 \equiv (A_{0,\uparrow} + A_{0,\downarrow})/N. \tag{3.9}$$

利用这一事实，我们可以将所有 ℓ阶关联项表示为 n阶关联项的短线性组合，其中 $n > \ell$。尽管它不再是一个基，但很容易证明（ ℓ）中的3.8阶关联项满足框架性质，数值测试表明，我们所有的结果都可以用（3.8）重现。为使表述清晰，从现在起我们将只关注（3.7）中给出的公式。

（*iv*）总次数稀疏化：基函数的主导项$\boldsymbol{A}_{\boldsymbol{\nu}}$ 在（3.6）中恰好是 $\otimes_{t=1}^{\mathcal{B}} \phi_{\nu_t}$的对称化，因此我们为其指定总次数

$$\deg(\boldsymbol{A}_{\boldsymbol{\nu}}) := \|\boldsymbol{\nu}\|_1 := \sum_{t=1}^{\mathcal{B}} \deg(\phi_{\nu_t}) \qquad \text{for } \boldsymbol{\nu} \in (\mathbb{N} \times \mathbb{Z}_m)^{\mathcal{B}}_{\rm ord}, \tag{3.10}$$

其中，单粒子基函数的次数$\deg(\phi_\nu)$ 由（3.3）给出。通过指定最大总次数 $\boldsymbol{A_\nu}$ 截断多体基 $\mathcal{D} \in \mathbb{N}_+$

$$\|\boldsymbol{\nu}\|_1 \leq \mathcal{D},$$

从而实现适用于解析目标函数的基的大幅稀疏化 [24, 45]。文献 [2] （另见（3.13））表明，与对称化结合使用时，该方法尤其有效，能够显著缓解采用全张量积网格进行传统离散化所带来的维数灾难。

通过使最大总次数取决于基函数的空间关联阶数$\|\boldsymbol{\nu}\|_0$，我们可以针对不同的空间关联阶数调整所需的分辨率。这很方便，因为自然可以预期，更高关联阶数的项对总能量的贡献较小，因此所需的分辨率也较低。为此，我们指定一个元组 $\boldsymbol{\mathcal{D}} = (\mathcal{D}_\ell)_{\ell=1}^{\mathcal{B}} \in \mathbb{N}_+^{\mathcal{B}}$。然后，相应的基函数指标集合为

$$\mathcal{I}(\mathcal{B}, \boldsymbol{\mathcal{D}}) := \Big\{ \boldsymbol{\nu} \in (\mathbb{N} \times \mathbb{Z}_m)_{\rm ord}^{\mathcal{B}} \ : \ \|\boldsymbol{\nu}\|_1 \leq \mathcal{D}_\ell \text{ if } \|\boldsymbol{\nu}\|_0 = \ell \Big\}, \tag{3.11}$$

由此得到一个 N-变量对称函数 f 的参数化形式，方法为

$$\begin{aligned} f(\boldsymbol{x}) \approx f_{\mathcal{B},\boldsymbol{\mathcal{D}}}(\boldsymbol{x}) \in \ V_{\mathcal{B},\boldsymbol{\mathcal{D}}}^N \ &:= \operatorname{span}\Big\{ \boldsymbol{A_\nu} : \boldsymbol{\nu} \in \mathcal{I}(\mathcal{B}, \boldsymbol{\mathcal{D}}) \Big\}, \\ f_{\mathcal{B},\boldsymbol{\mathcal{D}}}(\boldsymbol{x}) &= \sum_{\boldsymbol{\nu} \in \mathcal{I}(\mathcal{B},\boldsymbol{\mathcal{D}})} c_{\boldsymbol{\nu}} \boldsymbol{A_\nu}, \end{aligned} \tag{3.12}$$

其中 $\boldsymbol{c} := \{c_{\boldsymbol{\nu}}\}_{\boldsymbol{\nu} \in \mathcal{I}(\mathcal{B},\boldsymbol{\mathcal{D}})}$ 为展开的系数。

.....的主要结果 [2] 估计表示（3.12）的参数数量

$$\dim\big(V_{\mathcal{B},\boldsymbol{\mathcal{D}}}^N\big) \lesssim \begin{cases} m^{\mathcal{B}} \sum_{\ell=1}^{\mathcal{B}} D_\ell^\ell / (\ell!)^2 & \text{as } \min_\ell D_\ell \to \infty, \\ m^{\mathcal{B}} \sum_{\ell=1}^{\mathcal{B}} \exp\big(\pi \sqrt{\tfrac{4}{3} D_\ell}\big) & \text{uniformly in } N, \mathcal{B}, \boldsymbol{\mathcal{D}}. \end{cases} \tag{3.13}$$

一个并不令人意外但十分重要的观察是，参数数量仅取决于关联阶数 $\mathcal{B}$，而不取决于粒子数量 N。这意味着，只有在需要较高关联阶数来解析目标函数时，才会出现维数灾难。即便如此，置换不变性也显著缓解了这一影响：在我们的设定中，我们有 $m = 2$，因此第二个界仅以非常微弱的方式依赖于关联阶数 $\mathcal{B}$，这表明即使是较高的关联阶数，在我们的框架内也很容易处理。最后，我们注意到，通常可以预期，较高关联阶数项所需的分辨率较低，也就是说，我们可以令 D_ℓ 严格递减，这还能进一步降低计算成本。例如，在计算波函数时，我们预计对应于 $\mathcal{B} = 1$ 的 Hartree–Fock模型已经解析了精确波函数中的重要贡献，而后续修正相对而言是“较小的”。

严格地阐释这些启发式思想远远超出了本文的范围，但在粗粒化电子结构模型的背景下，我们可以参见 [51] 中的一个具体且严谨的示例。在那里，可以平衡所有近似参数，从而获得显式的超代数最佳逼近率 [2]。

3.3. 通过范德蒙德行列式实现反对称性。在确立了参数化高维对称函数的方法后，我们现在回到参数化反对称多电子波函数这一最初任务$\Psi(x_1, \cdots, x_N)$，其中包含空间和自旋坐标。我们考虑的第一种方法称为范德蒙德试探形式，其基于这样一个众所周知的事实：每个反对称多项式都可被范德蒙德乘积整除，从而得到一个对称多项式 作为结果 [7]。

我们将一般波函数表示为

$$\Psi_{\mathrm{V}}(\boldsymbol{x}) := \Phi(\boldsymbol{x}) \cdot \mathcal{V}(\boldsymbol{x}), \tag{3.14}$$

其中， Φ 关于位置和自旋对的交换是对称的，而 $\mathcal{V}$ 是一个行列式，其“单项式” v_i ($1 \le i \le N$)取决于电子构型$\boldsymbol{x}$，具体由自旋向上粒子的数量 $N_{\boldsymbol{x}\uparrow}$ 给出：$= \#\{\sigma_i : \sigma_i = \uparrow\}$,

$$\mathcal{V}(\boldsymbol{x}) := \det \begin{pmatrix} v_1(x_1) & \cdots & v_N(x_1) \\ \vdots & \ddots & \vdots \\ v_1(x_N) & \cdots & v_N(x_N) \end{pmatrix} \tag{3.15}$$

$$v_i(x_j) := v_i(r_j, \sigma_j; \boldsymbol{x}) := \begin{cases} r_j^{i-1} \cdot \delta_{\sigma_j,\uparrow}, & i \le N_{\boldsymbol{x}\uparrow} \\ r_j^{i-1-N_{\boldsymbol{x}\uparrow}} \cdot \delta_{\sigma_j,\downarrow}, & N_{\boldsymbol{x}\uparrow}+1 \le i \le N \end{cases}.$$

对称前因子 Φ 可以用第3.2小节中讨论的ACE模型进行参数化。对于给定的关联阶数 $\mathcal{B} \in \mathbb{N}$ 和次数 $\boldsymbol{\mathcal{D}} \in \mathbb{N}^{\mathcal{B}}$，我们将其参数化为 $\Phi(\boldsymbol{x}) = f_{\mathcal{B},\,\boldsymbol{\mathcal{D}}}(\boldsymbol{x})$ ，其中和$f_{\mathcal{B},\,\boldsymbol{\mathcal{D}}}$ 由（3.12）给出。因此，ACE-范德蒙德假设的最终形式为

$$\Psi_{\mathrm{V}}(\boldsymbol{x}) := f_{\mathcal{B},\boldsymbol{\mathcal{D}}}(\boldsymbol{x}) \cdot \mathcal{V}(\boldsymbol{x}). \tag{3.16}$$

注释 3.2（传统范德蒙德行列式）。如果我们对电子构型进行置换，并令前 $N_{\boldsymbol{x}\uparrow}$ 个粒子自旋向上，

$$\boldsymbol{x} = (r_1, \uparrow; \cdots; r_{N_{\boldsymbol{x}\uparrow}}, \uparrow; r_{N_{\boldsymbol{x}\uparrow}+1}, \downarrow; \cdots; r_N, \downarrow),$$

于是（3.15）可化为一种众所周知的范德蒙行列式形式

$$\mathcal{V}(\boldsymbol{x}) = \prod_{\substack{1 \le i < j \le N \\ \sigma_i = \sigma_j = \uparrow}} (r_i - r_j) \cdot \prod_{\substack{1 \le i < j \le N \\ \sigma_i = \sigma_j = \downarrow}} (r_i - r_j). \tag{3.17}$$

尽管在实践中使用起来更方便，但直接将（3.17）代入（3.14不会得到反对称波函数，详见第 3.5 小节中的进一步讨论。

注 3.3（完备性）。利用与 [7, 27] 中类似的论证，可以证明范德蒙德试探形式（3.14）是“完备的”。这里所说的是，任意反对称的 N变量多项式都可以写成形式（3.14）和（3.16）。此外还可以证明，如果一个反对称函数足够光滑，那么它可以严格写成形式（3.14），其中对称前因子 Φ 仍然光滑，因此可以用ACE多项式进行良好逼近。我们在定理A.1 的附录A中，在 [31, 定理3和 4]的分析基础上，对这些结论作出严格表述。

对于维数大于一的粒子，朴素的范德蒙德试探形式（3.14）（其中对范德蒙德行列式的概念作适当推广）并不完备，而是需要级数展开。有关更多细节，请参阅 [27] 。

注记 3.4（数值稳定性）。在我们使用范德蒙德试探形式进行数值测试时，观察到数值不稳定性，尤其是在电子数量较多时。这并不令人意外，因为范德蒙德行列式使用单项式作为矩阵元素。为了解决这一问题，可以用 Slater 行列式替代范德蒙德行列式，这将导出Hartree–Fock模型的推广；或者采用回流变换的一个特殊情形，我们将在下一小节中对此进行探讨。另一种方法是精心设计优化算法，同样可以克服这些不稳定性；在此我们指的是我们在第4.2节中开发的级联多层求解器。

3.4. 通过回流变换实现反对称性。回流变换最初由 Feynman 和 Cohen [18] 提出。近年来，将其与深度神经网络架构 [29, 42, 43]相结合取得了巨大成功。回流试探形式可视为 Slater 行列式的推广，

$$\Psi(x_1,\cdots,x_N) \approx \det \begin{pmatrix} \varphi_1(x_1) & \cdots & \varphi_N(x_1) \\ \vdots & \ddots & \vdots \\ \varphi_1(x_N) & \cdots & \varphi_N(x_N) \end{pmatrix}, \tag{3.18}$$

其中，φ_k $(1 \le k \le N)$ 是正交归一的单粒子轨道。这是近似多电子波函数最著名、最常见的选择，并由此产生了著名的 Hartree–Fock 方法 [28]。回流试探形式的思想是用多变量函数替代单粒子轨道 φ_k；即，

$$\Psi_{\rm BF}(\boldsymbol{x}) := \det \begin{pmatrix} \varphi_1(x_1;\boldsymbol{x}_{\neq 1}) & \cdots & \varphi_N(x_1;\boldsymbol{x}_{\neq 1}) \\ \vdots & \ddots & \vdots \\ \varphi_1(x_N;\boldsymbol{x}_{\neq N}) & \cdots & \varphi_N(x_N;\boldsymbol{x}_{\neq N}) \end{pmatrix} \tag{3.19}$$

其中 $\boldsymbol{x}_{\neq i} := \{x_k\}_{k\neq i}$。记号 $\varphi_j(x_i;\boldsymbol{x}_{\neq i})$ 表示这些广义轨道满足“部分”对称性

$$\varphi_j\big(x_i;\{x_{\pi(k)}\}_{k\neq i}\big) = \varphi_j\big(x_i;\boldsymbol{x}_{\neq i}\big) \qquad \forall \pi \in S(N-1). \tag{3.20}$$

等价的表述方式是说，向量 $(\varphi_1, \ldots\ldots, \varphi_N)$ 是置换-协变的。很容易看出，函数形式 (3.19) 再次满足反对称性 (2.1)。

为补全该试探形式（3.19），我们必须说明如何用原子簇展开形式体系来参数化 N 粒子轨道 φ_j。为此，我们修改基函数（3.5），将“突出显示的”粒子 x_i 分离出来，

$$\boldsymbol{A}^{\rm BF}_{\boldsymbol{\nu}}(x_i;\{x_k\}_{k\neq i}) := \phi_{\nu_1}(x_i) \prod_{t=2}^{N} A_{\nu_t}(\{x_k\}_{k\neq i}) \qquad \text{for } x_i, x_k \in \mathbb{R}\times\mathbb{Z}_2,$$

其中 $\boldsymbol{\nu} = (\nu_1;\ \nu_2,\ \cdots,\ \nu_N) \in (\mathbb{N}\times\mathbb{Z}_2) \times (\mathbb{N}\times\mathbb{Z}_2)^{N-1}$ 或且 ϕ_ν 是单粒子基函数（参见（3.1））。作为张量积，可以立即得到$\boldsymbol{A}^{\rm BF}_{\boldsymbol{\nu}}$构成满足部分对称性（3.20）的完备函数基。

现在，我们选取最大的关联阶数 $\mathcal{B} \geq 1$ 和次数 $\boldsymbol{\mathcal{D}} = (\mathcal{D}_\ell)_{\ell=1}^{\mathcal{B}}$，以及用于回流轨道的原子簇展开基，类似于 (3.6)，如下

$$\boldsymbol{A}_{\boldsymbol{\nu}}^{\mathrm{BF}}\big(x_i; \{x_k\}_{k\neq i}\big) := \begin{cases} \phi_{\nu_1}(x_i) & \text{if } \mathcal{B} = 1 \\ \phi_{\nu_1}(x_i) \prod_{t=2}^{\mathcal{B}} A_{\nu_t}(\{x_k\}_{k\neq i}) & \text{if } \mathcal{B} \geq 2 \end{cases} \tag{3.21}$$

对于 $x_i, x_k \in \mathbb{R} \times \mathbb{Z}_2$，其中 $\boldsymbol{\nu}$ 属于基索引集

$$\mathcal{I}^{\mathrm{BF}}(\mathcal{B}, \boldsymbol{\mathcal{D}}) := \Big\{ \boldsymbol{\nu} \in (\mathbb{N} \times \mathbb{Z}_2) \times (\mathbb{N} \times \mathbb{Z}_2)^{\mathcal{B}-1}_{\mathrm{ord}} \ : \ \|\boldsymbol{\nu}\|_1 \leq \mathcal{D}_\ell \text{ if } \|\boldsymbol{\nu}\|_0 = \ell \Big\}.$$

现在我们可以使用式（ φ_j 中的基函数（3.21）进行参数化，

$$\varphi_j(x_i; \boldsymbol{x}_{\neq i}) = f^{(j)}_{\mathcal{B},\boldsymbol{\mathcal{D}}}\big(x_i; \boldsymbol{x}_{\neq i}\big) := \sum_{\boldsymbol{\nu} \in \mathcal{I}^{\mathrm{BF}}(\mathcal{B},\boldsymbol{\mathcal{D}})} c^{(j)}_{\boldsymbol{\nu}} \boldsymbol{A}^{\mathrm{BF}}_{\boldsymbol{\nu}}(x_i; \boldsymbol{x}_{\neq i}). \tag{3.22}$$

因此，原子簇展开-回流试探形式的最终形式为

$$\Psi_{\mathrm{BF}}(\boldsymbol{x}) := \det \begin{pmatrix} f^{(1)}_{\mathcal{B},\boldsymbol{\mathcal{D}}}(x_1; \boldsymbol{x}_{\neq 1}) & \cdots & f^{(N)}_{\mathcal{B},\boldsymbol{\mathcal{D}}}(x_1; \boldsymbol{x}_{\neq 1}) \\ \vdots & \ddots & \vdots \\ f^{(1)}_{\mathcal{B},\boldsymbol{\mathcal{D}}}(x_N; \boldsymbol{x}_{\neq N}) & \cdots & f^{(N)}_{\mathcal{B},\boldsymbol{\mathcal{D}}}(x_N; \boldsymbol{x}_{\neq N}) \end{pmatrix}. \tag{3.23}$$

至此，我们完成了对原子簇展开-回流波函数试探形式的定义。

注释 3.5（完备性）。 [31, 定理 2 和 7] 指出，回流试探形式通过将对称部分吸收到行列式中，严格推广了范德蒙德行列式（另见 定理 A.3 ，见附录A了解详情）。因此，对于一维系统，回流试探形式也给出了反对称多项式的一种“完备”表示。然而，在更高维空间中，这一结论可能不成立。文 [30]中指出，在多项式这一类别中，完全反对称多项式无法由回流试探形式高效表示，而是需要对这种形式的指数多个行列式求和。不过请注意，这并不妨碍回流试探形式对高维粒子也具有出色的近似性质。这仅仅表明，严格理解其近似性质可能颇具挑战性。

备注3.6（实现说明）。一般的原子簇展开框架可实现便捷的回流基 $\boldsymbol{A}^{\mathrm{BF}}_{\boldsymbol{\nu}}$。为了“标记”交换 x_i， x_j 并非一种对称操作，只需引入一个额外的人工自旋类别 $\emptyset$ ，并如下修改粒子：

$$x_i' := (r_i, \emptyset), \quad x_j' := x_j \quad \text{for } j \neq i.$$

令 $\boldsymbol{A}'_{\boldsymbol{\nu}}$ 为标准的原子簇展开基，但此处用于粒子空间 $\mathbb{R} \times \mathbb{Z}_3$，其中 $\mathbb{Z}_3$ 表示类别 $\uparrow, \downarrow, \emptyset$，现在只需取 $\boldsymbol{A}^{\mathrm{BF}}_{\boldsymbol{\nu}}(x_i, \boldsymbol{x}_{\neq i}) := \boldsymbol{A}'_{\boldsymbol{\nu}}(\boldsymbol{x}')$。采用这一策略，即可使用标准的原子簇展开实现来处理部分对称性。

3.5. 自旋指定的波函数。由于哈密顿算符（2.2）是一个与自旋无关的算符，使用自旋指定的波函数进行处理较为方便[19, 40]。更准确地说，我们考虑一个具有 $N_\uparrow$ 个自旋向上的电子和 $N_\downarrow$个自旋向下的电子的系统，其中 $N = N_\uparrow + N_\downarrow$ 给出电子总数，并且

$S_z = \frac{1}{2}(N_\uparrow - N_\downarrow)$自旋极化。随后，我们可以仅用空间坐标来改写这种自旋指定系统的一般波函数，

$$\Psi(x_1, \cdots, x_N) = \Psi(r_1, \uparrow, \cdots, r_{N_\uparrow}, \uparrow, r_{N_\uparrow+1}, \downarrow, \cdots, r_N, \downarrow) =: \Psi_{\rm s}(\mathbf{r}^\uparrow, \mathbf{r}^\downarrow), \tag{3.24}$$

其中，我们重新标记了电子指标，使得前 $N_\uparrow$ 个电子为自旋向上，其位置为 $\mathbf{r}^\uparrow = (r_1, \cdots, r_{N_\uparrow})$，其余 $N_\downarrow$ 个电子为自旋向下，其位置为$\mathbf{r}^\downarrow = (r_{N_\uparrow+1}, \cdots, r_N)$。

然后，我们可以近似表示自旋指定的波函数为

$$\Psi_{\rm s}(\mathbf{r}^\uparrow, \mathbf{r}^\downarrow) = \Psi_{\rm V}\big(x_1, \ldots, x_N\big) \quad \text{or} \quad \Psi_{\rm BF}\big(x_1, \ldots, x_N\big), \qquad \text{where} \tag{3.25}$$

$$x_i = \begin{cases} (r_i, \uparrow), & i = 1, \ldots, N_\uparrow, \\ (r_i, \downarrow), & i = 1, \ldots, N_\downarrow, \end{cases}$$

和 $\Psi_{\rm V}$， $\Psi_{\rm BF}$ 是在 第 3.3 小节、第 3.4 小节 中分别构建的。

在试探波函数的拟设中固定自旋的一个重要优势（我们将在下一节立即看到）是，在变分蒙特卡罗方法中，我们只需对空间坐标（r_1，$\cdots$，r_N）的构型进行采样 $\in \mathbb{R}^N$。

注意，的典范扩展 $\Psi_{\rm s}$ 到 x_i 通过（3.25）定义的坐标已经完全反对称，无需额外的反对称化步骤。

4. 级联多层优化。使用原子簇展开拟设（3.16）或（3.23），多电子波函数被参数化为 $\Psi \approx \Psi_{\boldsymbol{c}}$，其中 $\boldsymbol{c} \in \mathbb{R}^{\mathcal{M}}$ 是所有原子簇展开参数的集合，其中 $\mathcal{M}$ 为参数总数。基态能量（2.5）可以通过关于波函数参数最小化一个“损失”（即瑞利商）来近似，

$$\min_{\boldsymbol{c}\in\mathbb{R}^{\mathcal{M}}} \mathcal{L}(\boldsymbol{c}) \qquad \text{with} \qquad \mathcal{L}(\boldsymbol{c}) := E(\Psi_{\boldsymbol{c}}) = \frac{\langle \Psi_{\boldsymbol{c}} | \mathcal{H} | \Psi_{\boldsymbol{c}} \rangle}{\langle \Psi_{\boldsymbol{c}} | \Psi_{\boldsymbol{c}} \rangle}. \tag{4.1}$$

我们将讨论使 $\mathcal{L}(\boldsymbol{c})$最小化的优化策略，并设计一种级联多层方法，其灵感来自 [6, 11, 17, 26, 33, 44], ，适用于我们基于原子簇展开的分层波函数参数化。原则上，我们的多层方法可应用于任何具有分层结构的波函数参数化，例如参见 [23]。

4.1.变分蒙特卡洛方法。损失函数 $\mathcal{L}(\boldsymbol{c})$ 在（4.1）中的评估涉及一个潜在的高维积分。在变分蒙特卡洛框架下，通常通过从分布 $P(\cdot、\ \Psi_{\boldsymbol{c}})$（$P$以及局部能量 E_L 定义于（2.7））

$$\begin{aligned} \mathcal{L}(\boldsymbol{c}) := E(\Psi_{\boldsymbol{c}}) &= \mathbb{E}_{\boldsymbol{x}\sim P(\cdot;\Psi_{\boldsymbol{c}})}\big[E_{\rm L}(\boldsymbol{x};\Psi_{\boldsymbol{c}})\big] \\ &\approx \frac{1}{n}\sum_{\boldsymbol{x}\in S_n} E_{\rm L}(\boldsymbol{x};\Psi_{\boldsymbol{c}}) =: \mathcal{L}_n(\boldsymbol{c}). \end{aligned} \tag{4.2}$$

这里，电子构型集合 S_n 是由 n 个样本 $\boldsymbol{x} \in S_n, \boldsymbol{x} \sim P(\cdot;\ \Psi_{\boldsymbol{c}})$ 组成的。这些样本通过马尔可夫链蒙特卡罗方法生成，其中 n 表示采样器的数量。由于在实际模拟中，我们使用的是赋予自旋的波函数（3.25），因此只需对具有空间坐标的电子构型进行采样，因为自旋坐标已固定。 $\mathbb{R}^n$

在实践中，我们使用标准的 Metropolis-Hastings 算法 [19]。我们还将其与标准的 Langevin 蒙特卡罗方法 [40] 进行了比较，发现两者的采样性能相近。

注4.1（实施说明）。在实践中，我们使用 $n = 2000$ 个样本，并在每次参数更新后运行10步梅特罗波利斯-黑斯廷斯采样。所提出的移动服从高斯分布，其各向同性协方差在运行过程中动态选择，使接受率保持在目标范围 [45%至55%]内。我们通过一种简单的方案维持这一比例：如果接受率远高于目标值，则将步长小幅增加；如果接受率远低于目标值，则将步长小幅减小。

最小化 $\mathcal{L}(\boldsymbol{c})$可通过随机梯度下降（SGD）迭代实现

$$\boldsymbol{c}_{k+1} = \boldsymbol{c}_k - \eta_k \cdot \eth_n(\boldsymbol{c}_k), \tag{4.3}$$

其中 $\eta_k > 0$ 是学习率，$\eth_n(\boldsymbol{c})$ 是通过使用来自分布 $P(\cdot, \Psi_{\boldsymbol{c}})$的一组样本 S_n 估计得到的梯度 $\nabla_{\boldsymbol{c}}\mathcal{L}(\boldsymbol{c})$ 的近似值

$$\begin{aligned} \nabla_{\boldsymbol{c}}\mathcal{L}(\boldsymbol{c}) &= 2\,\mathbb{E}_{\boldsymbol{x}\sim P(\cdot;\Psi_{\boldsymbol{c}})}\big[\big(\partial_{\boldsymbol{c}}\log|\Psi_{\boldsymbol{c}}(\boldsymbol{x})|\big)\big(E_{\mathrm{L}}(\boldsymbol{x};\Psi_{\boldsymbol{c}}) - \mathcal{L}(\boldsymbol{c})\big)\big] \\ &\approx \frac{2}{n}\sum_{\boldsymbol{x}\in S_n}\big(\partial_{\boldsymbol{c}}\log|\Psi_{\boldsymbol{c}}(\boldsymbol{x})|\big)\big(E_{\mathrm{L}}(\boldsymbol{x};\Psi_{\boldsymbol{c}}) - \mathcal{L}_n(\boldsymbol{c})\big) =: \eth_n(\boldsymbol{c}). \end{aligned} \tag{4.4}$$

该表述给出了损失梯度的无偏估计。梯度表达式的推导见 [35]。关于 SGD 迭代收敛性分析的两篇近期研究，请参阅 [1, 34] （4.3）。

在算法 4.1 中，我们回顾用于求解（4.1）的标准变分蒙特卡洛算法。

算法 4.1 标准变分蒙特卡洛算法

输入： 波函数参数化 $\Psi_{\boldsymbol{c}}$；初始猜测 $\boldsymbol{c}_0 \in \mathbb{R}^{\mathcal{M}}$；迭代次数 $T \in \mathbb{Z}_+$。1: **对于** $k = 0$, 1，……，T**执行**2: 生成一组 S_n 个 n 样本 $\boldsymbol{x} \in S_n$，$\boldsymbol{x} \sim P$（·，$\Psi_{\boldsymbol{c}_k}$）（例如，通过 MCMC）。3: 使用（4.4）近似计算梯度。4: 使用（4.3）更新参数 $\boldsymbol{c}_k \mapsto \boldsymbol{c}_{k+1}$。5: **结束循环 输出：** $\boldsymbol{c} = \boldsymbol{c}_T$ 及对应的能量 $\mathcal{L}$（$\boldsymbol{c}$）。

人们已经开发出不同的方法来加快优化收敛。例如，所谓的随机重构方法（或等价地，自然梯度方法） [46, 47]通过费 Fisher 信息矩阵的逆对随机梯度进行预处理；线性方法 [52] 涉及损失函数的二阶导数；而克罗内克因子化近似曲率（KFAC）方法 [38] 使用 Fisher 信息矩阵的克罗内克因子化近似。本文不探讨这些技术。相反，我们采用加权 Adam（AdamW）方案 [36] 并开发一种级联多层方法（见下一小节），以加快算法 4.1的收敛。我们采用衰减学习率 $\eta_k = \alpha/(1 + k/\beta)$，参数 α 和 β保持固定。

4.2. 级联多层方法。 当采用基于ACE的波函数参数化（3.16）和（3.23）且相关性相对较大时，

次数和多项式次数，标准的变分蒙特卡洛算法（算法 4.1）由于损失函数中的病态性，收敛到基态的速度较慢。我们提出了一种级联多层算法，该算法利用ACE模型的内在层次结构，显著加速优化过程。级联多层方法最初由 [5]针对二阶椭圆问题提出，但将这一思想应用于许多其他层次化参数化形式也是自然的。关于这种多层方法的更多讨论和应用，请参阅 [6, 11, 44] 。这一思想已被用于训练神经网络：在“较粗尺度”上使用更多次迭代，以便为“较细尺度”获得良好的初始猜测，并避免陷入局部极小值 [17, 26, 33]。类似的思想也被应用于优化量子多体问题中的高斯过程态波函数 [23], ，这些波函数通过“加性”核的形式，在隐式参数化中导出所有可能的多位点相关特征的线性组合，并以层次方式包含所有较低秩的特征。

我们要指出的是，我们的方法不应与“多层蒙特卡洛”（MLMC）方法 [21]混淆。MLMC方法的理念是，在随机模拟中计算期望值时，以相对较低的成本进行低精度的大量采样，而仅以相应较高的成本进行高精度的极少量采样。MLMC方法同样利用模型层次结构，但其目标是加速采样，而我们的多层方法则是在每一步使用朴素采样器来加速参数优化。由于我们在每个层级抽取样本的分布都不同，因此无法轻易将MLMC 方法纳入我们的变分蒙特卡洛算法。

对于给定的电子数 N，令

$$X_0 \subsetneq X_1 \subsetneq \cdots \subsetneq X_j \subsetneq \cdots \subsetneq X_L \tag{4.5}$$

是多电子波函数的一个（严格）嵌套的多层函数类序列。在每一层级， X_j中的一个元素写作 $\Psi_j(\cdot;\boldsymbol{c}^{(j)})$，其中参数 $\boldsymbol{c}^{(j)}$在向量空间 $\mathbb{R}^{\mathcal{M}_j}$中不受约束。

参数化的嵌套性（4.5）意味着存在一个延拓算子

$$\Pi_{j,j+1} : \mathbb{R}^{\mathcal{M}_j} \to \mathbb{R}^{\mathcal{M}_{j+1}}, \qquad \boldsymbol{c}_0^{(j+1)} := \Pi_{j,j+1}[\boldsymbol{c}^{(j)}] \tag{4.6}$$

使得

$$\Psi_{j+1}\big(\cdot;\boldsymbol{c}_0^{(j+1)}\big) \equiv \Psi_j\big(\cdot;\boldsymbol{c}^{(j)}\big). \tag{4.7}$$

级联多层方法的思路是，首先在计算成本较低的粗层级 X_0 应用变分蒙特卡洛算法，得到参数 $\boldsymbol{c}^{(0)}$，然后使用延拓 $\Pi_{0,1}\boldsymbol{c}^{(0)}$ 作为层级 X_1上另一次变分蒙特卡洛迭代的初始猜测，再重复这一过程。该过程可用以下算法表示：

算法 4.2 级联多层变分蒙特卡洛算法

输入： 层级 $\{X_j\}_{j=0}^{L}$；迭代步骤 $\{M_j\}_{j=0}^{L}$；初始猜测 $\boldsymbol{c}_0^{(0)}$。1: **对于** $j = 0,\ 1,\ \ldots,\ L$ **执行** 2: $\boldsymbol{c}_{M_j}^{(j)} \leftarrow$ 算法 4.1 ，采用参数化 X_j，初始猜测 $\boldsymbol{c}_0^{(j)}$， M_j 步骤。3: 设置 $\boldsymbol{c}_0^{(j+1)} = \Pi_{j,\ j+1}\boldsymbol{c}_{M_j}^{(j)}$。4: **结束循环 输出：** $\boldsymbol{c} = \boldsymbol{c}_{M_L}^{(L)}$ 以及相应的能量 $\mathcal{L}(\boldsymbol{c})$。

备注 4.2. 多层方法本质上在不同步骤中分别处理目标波函数的低频和高频分量，因此可以实现与经典多层方法相同的加速效果。级联多层方法的一个显著特点是，粗函数类一旦得到细化便可将其舍弃，这在算法上大大简化了该方法。

4.3. 原子簇展开波函数的层次结构与延拓。现在需要为我们在前几节中介绍的原子簇展开-Vandermonde和原子簇展开-Backflow 波函数参数化，具体说明函数类层次结构和延拓算子。

设 $\{\mathcal{B}^{(j)}\}$ 和 $\{\boldsymbol{\mathcal{D}}^{(j)}\}$ 为满足条件的“递增”截断参数。

$$\mathcal{B}^{(j)} \le \mathcal{B}^{(j+1)} \qquad \text{and} \qquad \mathcal{D}_\ell^{(j)} \le \mathcal{D}_\ell^{(j+1)} \quad \text{for } \ell = 1, \cdots, \mathcal{B}^{(j)},$$

其中，上述“≤”中至少有一个应为“<”。然后我们可以将多层函数类 X_j 指定为

$$X_j := \Big\{ \Psi = \Psi_{\rm V} : \ \Psi_{\rm V} \text{ can be written as (3.16) with } \mathcal{B} = \mathcal{B}^{(j)} \text{ and } \boldsymbol{\mathcal{D}} = \boldsymbol{\mathcal{D}}^{(j)} \Big\}$$

如果波函数采用 ACE-范德蒙德假设进行参数化，且

$$X_j := \Big\{ \Psi = \Psi_{\rm BF} : \ \Psi_{\rm BF} \text{ can be written as (3.23) with } \mathcal{B} = \mathcal{B}^{(j)} \text{ and } \boldsymbol{\mathcal{D}} = \boldsymbol{\mathcal{D}}^{(j)} \Big\}$$

如果波函数通过原子簇展开-回流试探形式进行参数化。

当从低层级 X_j 提升到更高层级 X_{j+1}时，我们纳入具有更高多项式次数或更高关联阶数的基函数，从而将一些更高分辨率的分量（更高频率或更高关联）添加到基组中。相应的延拓算子$\mathit{\Pi}_{j,j+1}$ 可在下文中明确给出。

设 $\boldsymbol{c}^{(j)}$ 为第 j 层上的参数，那么要将其嵌入更细的层级并表示为 $\boldsymbol{c}^{(j+1)}$，我们应定义从第 j 层到第（$j+1$）层的延拓$\mathit{\Pi}_{j,\ j+1}$ 。延拓的定义分为两种情况：(1) $\mathcal{B}^{j+1} = \mathcal{B}^j$；以及 (2)$\mathcal{B}^{j+1} = \mathcal{B}^j + 1$。注意，情况 $\mathcal{B}^{j+1} - \mathcal{B}^j > 1$ 可以通过情况 (2) 的复合得到。

对于第一种情况 $\mathcal{B}^{j+1} = \mathcal{B}^j$，我们将 $\mathit{\Pi}_{j,j+1}$ 定义为

$$\boldsymbol{c}^{(j+1)} = \mathit{\Pi}_{j,j+1}\boldsymbol{c}^{(j)} \qquad \text{with} \tag{4.8}$$

$$\Big(\mathit{\Pi}_{j,j+1}\boldsymbol{c}^{(j)}\Big)_{\boldsymbol{\nu}} := \begin{cases} \boldsymbol{c}_{\boldsymbol{\nu}}^{(j)} & \text{if } \boldsymbol{\nu} \in \mathcal{I}(\mathcal{B}^{(j)}, \boldsymbol{\mathcal{D}}^{(j)}) \\ 0 & \text{if } \boldsymbol{\nu} \in \mathcal{I}(\mathcal{B}^{(j+1)}, \boldsymbol{\mathcal{D}}^{(j+1)}) \backslash \mathcal{I}(\mathcal{B}^{(j)}, \boldsymbol{\mathcal{D}}^{(j)}) \end{cases}.$$

对于第二种情况 $\mathcal{B}^{j+1} = \mathcal{B}^j + 1$，由于索引集 $\mathcal{I}(\mathcal{B}^{(j)},\ \boldsymbol{\mathcal{D}}^{(j)})$ 不是 $\mathcal{I}(\mathcal{B}^{(j+1)},\ \boldsymbol{\mathcal{D}}^{(j+1)})$ 的子集，我们需要先为这些索引引入两个嵌入

$$\mathcal{T}_\uparrow : \mathcal{I}(\mathcal{B}, \boldsymbol{\mathcal{D}}) \to \mathcal{I}(\mathcal{B}+1, \boldsymbol{\mathcal{D}}), \qquad \mathcal{T}_\uparrow(\nu_1, \cdots, \nu_{\mathcal{B}}) = \pi\Big((0,\uparrow), \nu_1, \cdots, \nu_{\mathcal{B}}\Big) \qquad \text{and}$$

$$\mathcal{T}_\downarrow : \mathcal{I}(\mathcal{B}, \boldsymbol{\mathcal{D}}) \to \mathcal{I}(\mathcal{B}+1, \boldsymbol{\mathcal{D}}), \qquad \mathcal{T}_\downarrow(\nu_1, \cdots, \nu_{\mathcal{B}}) = \pi\Big((0,\downarrow), \nu_1, \cdots, \nu_{\mathcal{B}}\Big),$$

其中，在定义中需要一个额外的置换 π，使得 $\mathcal{T}_\uparrow(\boldsymbol{\nu})$ $\mathcal{T}_\downarrow(\boldsymbol{\nu})$ $\in$

$(\mathbb{N}\times\mathbb{Z}_m)^{\mathcal{B}+1}$ 阶。然后，延拓 $\Pi_{j,j+1}$ 定义为

$$\boldsymbol{c}^{(j+1)} = \Pi_{j,j+1}\boldsymbol{c}^{(j)} \qquad \text{with} \tag{4.9}$$

$$\left(\Pi_{j,j+1}\boldsymbol{c}^{(j)}\right)_{\boldsymbol{\nu}} := \begin{cases} \dfrac{1}{N}\displaystyle\sum_{\boldsymbol{\nu}^{(j)}\in\mathcal{I}_{\boldsymbol{\nu}}} \boldsymbol{c}^{(j)}_{\boldsymbol{\nu}^{(j)}} & \text{if } \mathcal{I}_{\boldsymbol{\nu}} \neq \emptyset \\ 0 & \text{otherwise} \end{cases}$$

对于 $\boldsymbol{\nu}\in\mathcal{I}\big(\mathcal{B}^{(j+1)},\boldsymbol{\mathcal{D}}^{(j+1)}\big)$，其中

$$\mathcal{I}_{\boldsymbol{\nu}} := \left\{\boldsymbol{\nu}^{(j)}\in\mathcal{I}\big(\mathcal{B}^{(j)},\boldsymbol{\mathcal{D}}^{(j)}\big) \;:\; \mathcal{T}_{\uparrow}\big(\boldsymbol{\nu}^{(j)}\big)=\boldsymbol{\nu} \text{ or } \mathcal{T}_{\downarrow}\big(\boldsymbol{\nu}^{(j)}\big)=\boldsymbol{\nu}\right\}.$$

利用（3.9），我们看到延拓 $\Pi_{j,\ j+1}\boldsymbol{c}^{(j)}$ 给出

$$\begin{aligned}\Psi_{\boldsymbol{c}^{(j+1)}} &= \sum_{\substack{\boldsymbol{\nu}\in\mathcal{I}(\mathcal{B}^{(j+1)},\boldsymbol{\mathcal{D}}^{(j+1)})\\ \mathcal{I}_{\boldsymbol{\nu}}\neq\emptyset}} \frac{1}{N}\sum_{\boldsymbol{\nu}^{(j)}\in\mathcal{I}_{\boldsymbol{\nu}}} \boldsymbol{c}^{(j)}_{\boldsymbol{\nu}^{(j)}}\boldsymbol{A}_{\boldsymbol{\nu}} \\ &= \sum_{\boldsymbol{\nu}^{(j)}\in\mathcal{I}(\mathcal{B}^{(j)},\boldsymbol{\mathcal{D}}^{(j)})} \boldsymbol{c}^{(j)}_{\boldsymbol{\nu}^{(j)}}\boldsymbol{A}_{\boldsymbol{\nu}^{(j)}}\frac{1}{N}(A_{0,\uparrow}+A_{0,\downarrow}) = \sum_{\boldsymbol{\nu}^{(j)}\in\mathcal{I}(\mathcal{B}^{(j)},\boldsymbol{\mathcal{D}}^{(j)})} \boldsymbol{c}^{(j)}_{\boldsymbol{\nu}^{(j)}}\boldsymbol{A}_{\boldsymbol{\nu}^{(j)}} = \Psi_{\boldsymbol{c}^{(j)}},\end{aligned}$$

这意味着波函数在 th 之前和之后完全相同 e
延拓。

我们看到，由4.8）和（4.9）定义的延拓在实践中都可以轻松且高效地实现。

5. 数值实验。本节通过模拟若干具有代表性的示例体系，展示我们算法的效率。所有模拟结果均以原子单位（原子单位）给出。所有模拟均在一台配备 16 个 Intel Xeon W-3275M 处理器和 1T RAM 的工作站上，使用 Julia [4] 软件包 ACESchrodinger.jl [41]完成。

我们将对第二周期原子的一维版本、锂氢分子（LiH）以及氢链（H_n）进行模拟。为说明结果的准确性，我们将其与其他标准量子化学方法的结果进行比较：（i）限制性 Hartree–Fock（RHF）、（ii）非限制性 Hartree–Fock（UHF）以及（iii）Jastrow-Slater（JS）。在 Jastrow-Slater 计算中，波函数被参数化为一个 Slater 行列式（形式见（3.18））与一个 Jastrow 因子的乘积。我们使用的基准 Jastrow 因子具有以下形式：

$$J(r_1,\sigma_1,\cdots,r_N,\sigma_N) = \sum_{i=1}^{N}\chi(r_i) + \sum_{i<j}\omega^{\sigma_i\sigma_j}\big(|r_i-r_j|\big) \qquad \text{with}$$

$$\chi(r) = \sum_{I=1}^{M}\xi_I\sqrt{1+|r-R_I|^2} \quad \text{and} \quad \omega^{\sigma\zeta}(u) = \frac{a^{\sigma\zeta}}{\sqrt{1+u^2}}\big(1-\exp(-b^{\sigma\zeta}\sqrt{1+u^2})\big),$$

其中，$\xi_I (I=1,\cdots,M)$ 和 $a^{\sigma\zeta},\ b^{\sigma\zeta} (\sigma,\zeta\in\{\uparrow,\downarrow\})$ 是可调参数。

为直观展示基态解，我们展示由多电子波函数定义的单电子密度 $\rho(r)$和双电子密度 $\rho_2(r, r')$。

函数乘以

$$\rho(r) = N \sum_{\sigma_1,\cdots,\sigma_N \in \mathbb{Z}_2} \int \left|\Psi(r,\sigma_1,r_2,\sigma_2,\ldots,r_N,\sigma_N)\right|^2 \mathrm{d}r_2 \ldots \mathrm{d}r_N, \quad \text{and}$$

$$\rho_2(r,r') = \binom{N}{2} \sum_{\sigma_1,\cdots,\sigma_N \in \mathbb{Z}_2} \int \left|\Psi(r,\sigma_1,r',\sigma_2,r_3,\sigma_3,\ldots,r_N,\sigma_N)\right|^2 \mathrm{d}r_3 \ldots \mathrm{d}r_N.$$

上述定义中的积分通过采用 Metropolis-Hastings 抽样的蒙特卡洛方法计算。

例 1：原子。我们对元素周期表第二周期中的一维原子进行模拟。首先采用回流模型，并在图 5.1中比较级联多层算法与纯 AdamW 方法的收敛情况。取最终关联阶数 $\mathcal{B} = 2$ 和多项式次数 $\boldsymbol{\mathcal{D}} = (16,16)$，可以看出，多层方法的收敛速度显著快于直接使用 AdamW 的方法。对于铍原子（$N = 4$），多层方法经过 700 步即可达到 -6.784 原子单位的能量（精度为 0.001 原子单位），而纯 AdamW 方法约需 1000 步；对于氧原子（N =8），多层方法约需 600 步即可达到 -21.692 原子单位的能量（精度为 0.005 原子单位），而纯 AdamW 方法则需 1400 多步才能达到相同精度。我们发现，随着参数化自由度数量的增加，多层算法的改进效果愈发显著。例如，若采用范德蒙德模型模拟氖原子，且 $\mathcal{B} = 3$，直接使用 AdamW 的算法大多数情况下会无法收敛，而多层方法则始终能很好地收敛至基态；参见表 5.1。

此外，我们强调，多层方法不仅需要更少的迭代步骤即可达到所需精度，而且在早期迭代阶段，波函数的计算成本低得多，因此还能节省计算成本。

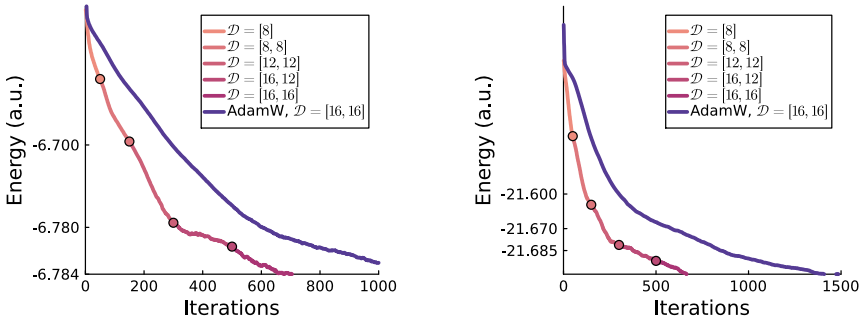

FIG. 5.1. *Convergence of the multilevel VMC algorithms with backflow model. Left: Be; Right: O.*

接下来，我们比较不同参数化的准确性。在 图 5.2中，我们展示了基态能量随自由度数量的变化。我们观察到，在达到相同的低能量时，回流模型所需的参数多于范德蒙德模型。另一方面，回流试探形式只需较低的关联阶数和显著更小的多项式次数即可达到可比的准确性。此外，我们观察到，回流模型达到了

TABLE 5.1
Decay of the energy errors of the multilevel and pure AdamW methods with the Vandermonde model of Neon. The converged ground state energy is taken as the reference to calculate the errors.

迭代	$\mathcal{D} = (32,16,8)$		$\mathcal{D} = (16,8,4)$	
	AdamW（原子单位）	多层级（原子单位）	AdamW（原子单位）	多层级（原子单位）
10	46.9158	23.0249	151.4131	17.3581
20	1173.2703	21.3572	2368.7069	6.3844
40	5172.9311	16.7712	6136.1001	5.0375
80	4426.9050	9.9026	1542.2801	1.2489
160	1351.9313	2.0020	3068.2906	0.1859
320	38.2721	0.1626	12.1771	0.0395

在关联阶数为1时具有与 UHF 模型相同的准确度，并且在关联阶数为 2 时优于 Jastrow-Slater 模型。

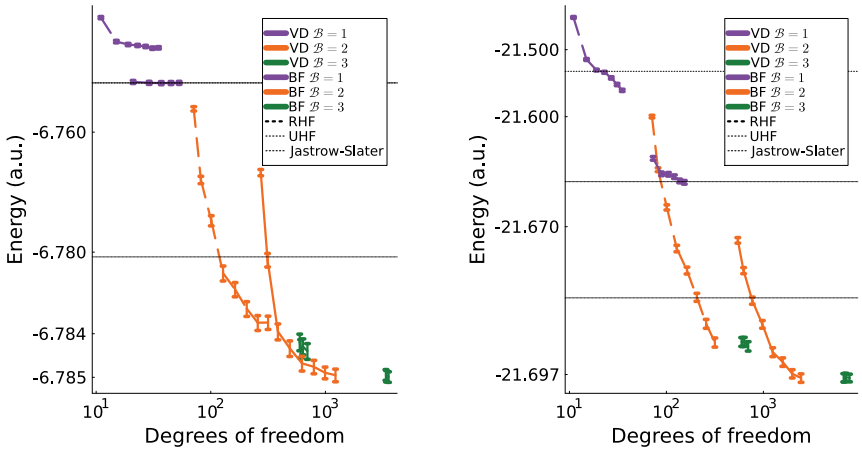

FIG. 5.2 *Convergence of the ground state energy with respect to the ACE degrees of freedom.* 。, *In the legends*
"BF" stands for the backflow model and "VD" stands for the Vandermonde model. The error bars show the standard deviations in the reported energy. Left: Be; Right: O.

注 5.1. 对原型代码的性能进行基准测试通常并无太大参考价值，因此我们不会过多关注这一方面。不过，为了佐证前述评论，我们在 表 5.2 中给出了计算波函数 $\Psi_{\boldsymbol{c}}(\boldsymbol{x})$ 以及哈密顿算符 $(\mathcal{H}\Psi_{\boldsymbol{c}})(\boldsymbol{x})$ 在给定构型 $\boldsymbol{x}$ 下针对氧原子的运行时间。这项性能测试是在搭载 M2 芯片的 MacBook Air 上进行的。我们再简要比较一下这两个模型：对于回流试探形式，使用 $\mathcal{D} = (16, 8)$（含 977 个参数）；对于范德蒙德试探形式，使用 $\mathcal{D} = (16, 14)$（含 257 个参数），两种方法达到了相近的精度。每个优化轮次中，回流试探形式所需时间是范德蒙德试探形式的 1。13 倍。

我们进一步展示了图 5.3 基态的单电子和双电子密度。我们可以看到，对于这类原子系统，回流模型得到的密度在定性上与 RHF 模型得到的密度相似。不过，两者之间也存在明显的

TABLE 5.2

Timing results (μs, serial) for evaluating $\Psi_{\boldsymbol{c}}\ \boldsymbol{x}$ *and* $(\mathcal{H}\Psi_{\boldsymbol{c}})\ \boldsymbol{x}$ *at a given configuration* $\boldsymbol{x}$ *of* 0 0 *Oxygen. The corresponding degrees of freedom (dof) for each parameterisation are also given.*

次数 $\mathcal{D}$	Vandermonde-ACE			Backflow-ACE		
	自由度	$\Psi_{\boldsymbol{c}}(\boldsymbol{x})$	$(\mathcal{H}\Psi_{\boldsymbol{c}})(\boldsymbol{x})$	# 自由度	$\Psi_{\boldsymbol{c}}(\boldsymbol{x})$	$(\mathcal{H}\Psi_{\boldsymbol{c}})(\boldsymbol{x})$
(32)	67	2.53	85.0	265	7.41	81.5
(32,16)	380	2.88	129	2961	16.9	152
(32,16,8)	793	3.69	225	8633	28.6	276
(32,16,8,4)	1211	5.09	358	15137	41.7	411

差异，尤其是氧和氖原子的单电子密度，因为回流模型能够捕捉到 Hartree–Fock 模型无法呈现的特征。

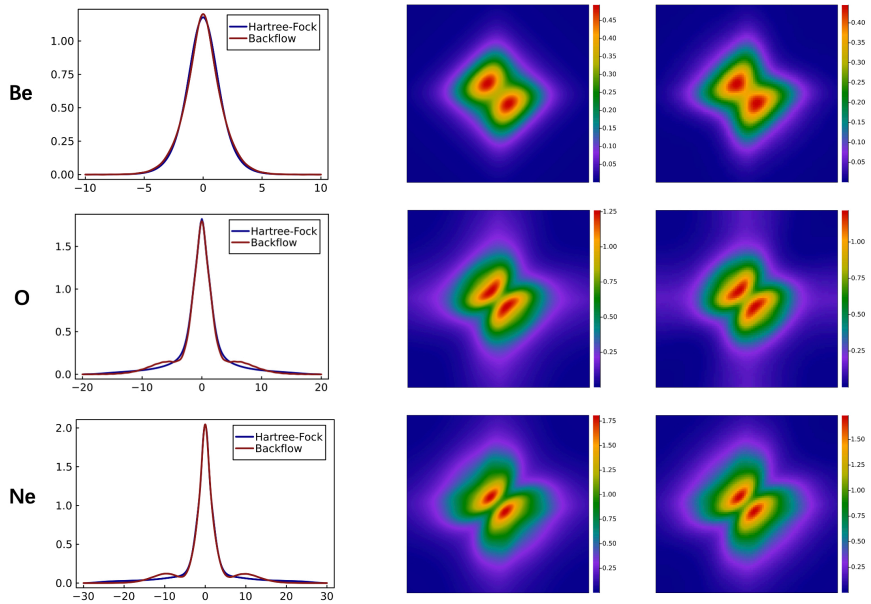

FIG. 5.3 *The ground state single- and pair- densities for the 1d atoms. Left: single-electron . density; Middle: pair-density by RHF; Right: pair-density by backflow.*

示例 2：LiH。LiH 分子有 $M = 2$ 个原子（核电荷分别为 3 和 1）以及 $N = 4$ 个电子。这两个原子位于 $R_1 = -d/2$ 和$R_2 = d/2$ ，其中 $d = 2.8$ 原子单位为锂原子与氢原子之间的距离。我们在 图 5.4 中展示了基态单电子密度和双电子密度。从双电子密度图中可以清楚地看到左右关联：当一个电子位于原点左侧时，找到另一个电子的概率更倾向于右侧区域，反之亦然；而两个电子同时位于原点右侧的概率非常低。

示例3：氢链。最后，我们考虑氢链H$_n$ ，其中含有 n个原子和 n 个电子。氢原子位于 $R_k = -(n-1)d/2 -$

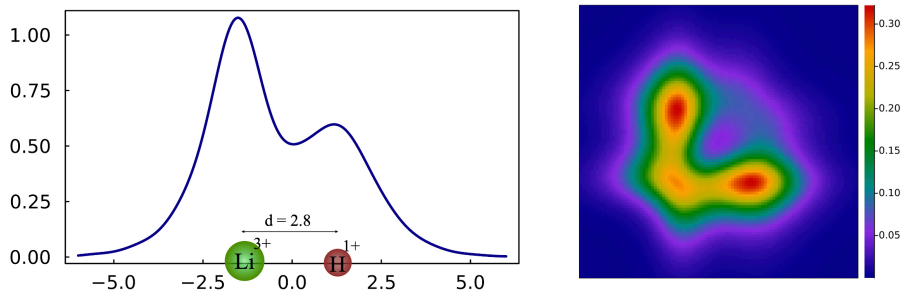

FIG. 5.4.

Left: The atomic configuration of (1d) LiH and the ground state single-electron density. Right: The ground state pair-electron density.

$(k-1)d\,(k=1,\ 2,\ \cdots,\ n)$，其中 d 表示原子之间的原子间距（原子单位）。我们关注链的解离极限，即当原子间距 d 变得很大时。在该极限下，体系的基态预计会表现得类似于 n 孤立的氢原子。因此，我们使用单个氢原子的基态能量 $E_1 = -0.669777$ 原子单位（通过在精细网格上求解微分方程获得）作为参考值，以检验模拟的准确性。需要指出的是，这是一个典型的强关联体系，对于这种体系，许多方法都无法捕捉到正确的行为 [10, 20, 54]。从量子化学的角度来看，对相关问题的通常理解是静态关联，这对应于本质上的多行列式情形。当然，也可以通过使用非限制性Hartree–Fock 模型等方式打破自旋对称性；在无限原子间距下这或许是允许的，但对于任何其他原子间距，该解都无法给出正确的基态。

在本例中，我们将仅采用回流模型。我们绘制了 H_8 在图5.5中的解离曲线，分别针对 RHF 和回流模型。我们观察到，通过与参考 E_1进行比较，回流模型在解离极限处给出了准确的基态，而随着 d 增大，RHF 模型则表现出定性上不正确的行为。图中还对基态电子密度进行了比较，由此可以再次观察到，在较大的原子间距处存在明显的定性偏差。我们在表 5.3中展示了不同原子数 n 和原子间距 d下的更多模拟结果，并观察到，即使在较大的原子间距 d下，ACE 回流模型也能达到足够的精度。需要说明的是，由于原子间距相当大，这些模拟使用了以原子为中心的单粒子基函数（3.2）。

TABLE 5.3
Ground state energy (per atom) for Hydrogen chains with n atoms.

间距（原子单位）	H_2 (原子单位)	H_4 (原子单位)	H_6 (原子单位)	H_8 (原子单位)	H_{10} (原子单位)
12.0	-0.6697(6)	-0.6697(5)	-0.6697(3)	-0.6697(1)	-0.6697(1)
20.0	-0.6697(7)	-0.6697(6)	-0.6697(7)	-0.6697(7)	-0.6697(7)

6. 结论。我们开发了一种基于反对称 ACE 架构的级联多层变分蒙特卡洛方法，用于求解多电子薛定谔方程。这些架构具有天然的多层结构。我们通过数值计算证明，在典型的

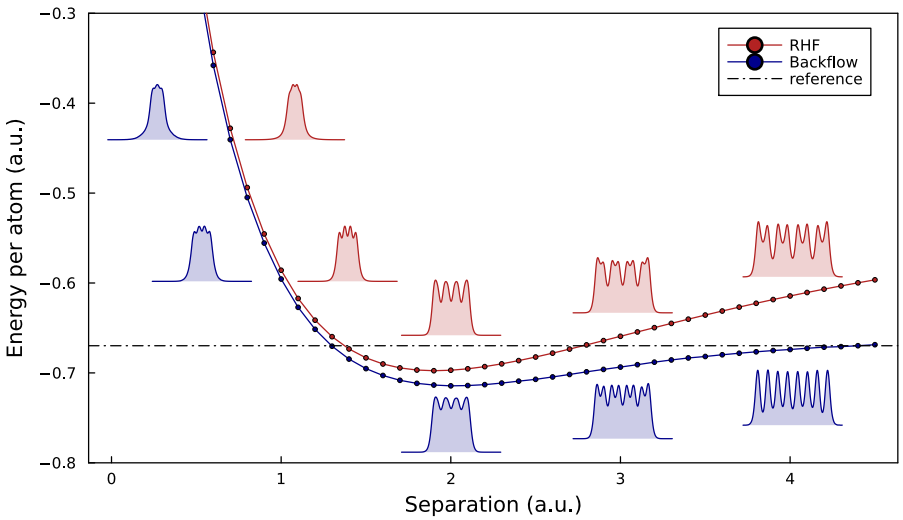

FIG. 5.5 *The dissociation curves of H_8. The reference is given by the ground state energy of . a single Hydrogen atom.*

一维电子系统中，我们的方法表现良好，且精度可系统地提升。接下来还需证明，该框架在具有标准库仑相互作用的实际三维系统中同样表现出色；这需要更精心地构造 ACE 波函数架构中所用的单粒子基。尤其值得探究的是，在强关联系统中，如何保持 ACE 参数化的性能和精度。

附录A. 参数化的完备性。在本附录中，我们给出解析结果，以支持反对称函数参数化的"完备性"。证明还将指出，回流试探形式比范德蒙德试探形式通用得多。我们的分析在概念上遵循 [7, 27, 31]。为简单起见，我们只讨论自旋指定波函数 Ψ_{s} 在（3.25）中的情形。推广到含有自旋变量的情形也很直接。我们强调，自旋指定波函数 Ψ_{s} : $\mathbb{R}^N \to \mathbb{R}$ 有不同的反对称性要求，因为它只是空间变量的函数。完整波函数的反对称约束 $\Psi(\boldsymbol{x})$ 可化简为以下关于 $\Psi_{\mathrm{s}}(\mathbf{r})$

$$\Psi_{\mathrm{s}}(\mathbf{r}^{\uparrow},\mathbf{r}^{\downarrow}) = (-1)^{\pi^{\uparrow}}(-1)^{\pi^{\downarrow}}\Psi_{\mathrm{s}}\big(\pi^{\uparrow}(\mathbf{r}^{\uparrow}),\pi^{\downarrow}(\mathbf{r}^{\downarrow})\big) \qquad \forall\ \pi^{\uparrow},\pi^{\uparrow}, \tag{A.1}$$

其中， $\pi^{\uparrow}$ 和 $\pi^{\downarrow}$ 分别表示$\mathbf{r}^{\uparrow}$ 和$\mathbf{r}^{\downarrow}$的置换。

THEOREM A.1. 若 Ψ_{s} : $\mathbb{R}^N \to \mathbb{R}$ 满足(A.1)，则它可以表示为

$$\Psi_{\mathrm{s}}(\mathbf{r}^{\uparrow},\mathbf{r}^{\downarrow}) = \Phi_{\mathrm{s}}(\mathbf{r}^{\uparrow},\mathbf{r}^{\downarrow})\cdot\mathcal{V}_{\mathrm{s}}(\mathbf{r}^{\uparrow},\mathbf{r}^{\downarrow}), \tag{A.2}$$

其中 $\mathcal{V}_{\mathrm{s}}(\mathbf{r}^{\uparrow},\mathbf{r}^{\downarrow}) = \prod_{1\le i<j\le N_{\uparrow}}(r_i-r_j)\cdot\prod_{N_{\uparrow}+1\le i<j\le N}(r_i-r_j)$ 且 Φ_{s} 满足

$$\Phi_{\mathrm{s}}(\mathbf{r}^{\uparrow},\mathbf{r}^{\downarrow}) = \Phi_{\mathrm{s}}\big(\pi^{\uparrow}(\mathbf{r}^{\uparrow}),\pi^{\downarrow}(\mathbf{r}^{\downarrow})\big) \qquad \forall\ \pi^{\uparrow},\pi^{\downarrow}. \tag{A.3}$$

此外，如果 Ψ_{s} 是多项式或解析函数，那么 Φ_{s} 也是如此。

证明。显然， $\mathcal{V}_{\rm s}$ 满足与 $\Psi_{\rm s}$相同的反对称条件。因此，作为两个反对称函数的商， $\Phi_{\rm s}$ 满足（A.3）。

此外，如果 $\Psi_{\rm s}$ 是满足 (A.1) 的多项式，我们可以利用因式定理证明 $r_i^\sigma - r_j^\sigma$($\sigma \in \uparrow, \downarrow$) 整除 $\Psi_{\rm s}(\mathbf{r}^\uparrow, \mathbf{r}^\downarrow)$。根据多元多项式的唯一分解性质，我们可以将 $\Psi_{\rm s}$ 写成 (A.2)，其中 $\Phi_{\rm s}(\mathbf{r}^\uparrow, \mathbf{r}^\downarrow)$ 是具有“部分”对称性 (A.3) 的多项式。此外，我们可以通过观察所有 $r_i^\sigma - r_j^\sigma (\sigma \in \uparrow, \downarrow)$ 的乘积的次数为 $N_\sigma(N_\sigma - 1)/2$，估算 $\Phi_{\rm s}$ 的次数。因此， $\Phi_{\rm s}(\mathbf{r}^\uparrow, \mathbf{r}^\downarrow)$ 的次数不超过 $\deg(\Psi_{\rm s}) - N_\uparrow(N_\uparrow - 1)/2 - N_\downarrow(N_\downarrow - 1)/2$。

接下来我们证明，若 $\Psi_{\rm s}$ 是解析的，那么 Φ_s 也解析：假设 $\Psi_{\rm s}$ 是解析的，我们可以将其写成如下多元泰勒级数展开：

$$\Psi_{\rm s}(\mathbf{r}) = \sum_{\mathbf{k}} c_{\mathbf{k}} \mathbf{r}^{\mathbf{k}},$$

其中$\mathbf{k} = (k_1, \ldots, k_N) \in \mathbb{N}_0^N$， $\mathbf{r}^{\mathbf{k}} := r_1^{k_1} \cdots r_N^{k_N}$，而 $c_{\mathbf{k}} \in \mathbb{R}$ 是展开的系数。使用反对称化算子 AS对等式两边进行反对称化，

$$AS[\mathbf{r}^{\mathbf{k}}] := \sum_{\pi^\uparrow} \sum_{\pi^\downarrow} (-1)^{\epsilon_{\pi^\uparrow}} (-1)^{\epsilon_{\pi^\downarrow}} \pi^\uparrow(r_1)^{k_1} \cdots \pi^\uparrow(r_{N_\uparrow})^{k_{N_\uparrow}} \pi^\downarrow(r_{N_\uparrow+1})^{k_{N_\uparrow}+1} \cdots \pi^\downarrow(r_N)^N,$$

我们得到 $AS[\Psi_{\rm s}(\mathbf{r})] = \sum_{\mathbf{k}} c_{\mathbf{k}} AS[\mathbf{r}^{\mathbf{k}}]$。根据上述结果，我们知道 $AS[\mathbf{r}^{\mathbf{k}}]$ 可以表示为

$$AS[\mathbf{r}^{\mathbf{k}}] = U_{\mathbf{k}}(\mathbf{r}) \cdot \mathcal{V}_{\rm s}(\mathbf{r}^\uparrow, \mathbf{r}^\downarrow),$$

其中 $U_{\mathbf{k}}(\mathbf{r})$是一个“部分”对称多项式。将该表达式代入反对称化后的泰勒级数展开中，我们得到

$$AS[\Psi_{\rm s}(\mathbf{r})] = \sum_{\mathbf{k}} c_{\mathbf{k}} U_{\mathbf{k}}(\mathbf{r}) \cdot \mathcal{V}_{\rm s}(\mathbf{r}^\uparrow, \mathbf{r}^\downarrow).$$

这具有式（A.2）所给出的形式，其中 $\Phi(\mathbf{r}) := \sum_{\mathbf{k}} c_{\mathbf{k}} U_{\mathbf{k}}(\mathbf{r})$是解析的，因为$U_{\mathbf{k}}(\mathbf{r})$是一个“部分”对称多项式。 □

注记 A.2（可微函数）。如果 $\Psi_{\rm s} : \mathbb{R}^N \to \mathbb{R}$ 是 p次可微的，即 $\Psi_{\rm s} \in C^p(\mathbb{R}^N)$，那么将 $\Psi_{\rm s}$除以因子 $r_i^\uparrow - r_j^\uparrow (1 \le i < j \le N_\uparrow)$，便可得到一个关于 C^{p-1}的 N变量函数。我们可以递归地除去 $r_\ell - r_k$，从而得到一个关于 $C^{P - N_\uparrow(N_\uparrow-1)/2}(\mathbb{R}^N)$的函数。对自旋向下部分重复相同的论证，可得出结论： $\Phi \in C^k(\mathbb{R}^N)$，其中 $k = P - N_\uparrow(N_\uparrow - 1)/2 - N_\downarrow(N_\downarrow - 1)/2$。

THEOREMA.3. 设 $\Psi_{\rm s} : \mathbb{R}^N \to \mathbb{R}$ 满足 (A.1)。那么存在轨道 $\varphi_s^i : \mathbb{R}^N \to \mathbb{R}$ ($1 \le i \le N$)，满足

$$\varphi_{\rm s}^i(r_j; \mathbf{r}_{\neq j}^\uparrow, \mathbf{r}_{\neq j}^\downarrow) = \varphi_{\rm s}^i\Big(r_j; \pi^\uparrow(\mathbf{r}_{\neq j}^\uparrow), \pi^\downarrow(\mathbf{r}_{\neq j}^\downarrow)\Big) \qquad \forall\ \pi^\uparrow, \pi^\uparrow, \tag{A.4}$$

使得 $\Psi_{\rm s}$ *and* 可以表示为形式的

$$\Psi_{\rm s}(\mathbf{r}) = \det\left(\Phi^\uparrow(\mathbf{r})\right) \cdot \det\left(\Phi^\downarrow(\mathbf{r})\right) \qquad \text{with} \tag{A.5}$$

$$\Phi^\sigma(\mathbf{r}) = \begin{pmatrix} \varphi_{\rm s}^{1+\delta}(r_{1+\delta}; \mathbf{r}_{\neq 1+\delta}) & \cdots & \varphi_{\rm s}^{N_\sigma+\delta}(r_{1+\delta}; \mathbf{r}_{\neq 1+\delta}) \\ \vdots & \ddots & \vdots \\ \varphi_{\rm s}^{1+\delta}(r_{N_\sigma+\delta}; \mathbf{r}_{\neq N_\sigma+\delta}) & \cdots & \varphi_{\rm s}^{N_\sigma+\delta}(r_{N_\sigma+\delta}; \mathbf{r}_{\neq N_\sigma+\delta}) \end{pmatrix}, \quad \sigma \in \{\uparrow, \downarrow\},$$

其中 $\delta = 0$ 若 $\sigma = \uparrow$，且 $\delta = N_\uparrow$ 若 $\sigma = \downarrow$。此外，如果 Ψ_s 是多项式或解析函数，那么轨道 $\varphi_\mathrm{s}^i\ (1 \le i \le N)$也同样如此。

证明。 我们将证明，范德蒙德试探形式（A.2）可以表示为回流参数化（A.5）的形式。回顾定理 A.1 给出了 Ψ_s 的如下表达式：

$$\Psi_\mathrm{s}(\mathbf{r}^\uparrow, \mathbf{r}^\downarrow) = \Phi_\mathrm{s}(\mathbf{r}^\uparrow, \mathbf{r}^\downarrow) \cdot \prod_{1 \le i < j \le N_\uparrow} (r_i - r_j) \cdot \prod_{N_\uparrow + 1 \le k < \ell \le N} (r_k - r_\ell),$$

其中 Φ_s（$\mathbf{r}^\uparrow$，$\mathbf{r}^\downarrow$）满足（A.3）。于是，我们可以将上式改写为

$$\Psi_\mathrm{s}(\mathbf{r}^\uparrow, \mathbf{r}^\downarrow) = \begin{vmatrix} 1 & \Phi_\mathrm{s}(\mathbf{r}^\uparrow, \mathbf{r}^\downarrow) \cdot r_1 & \cdots & r_1^{N_\uparrow - 1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & \Phi_\mathrm{s}(\mathbf{r}^\uparrow, \mathbf{r}^\downarrow) \cdot r_{N_\uparrow} & \cdots & r_{N_\uparrow}^{N_\uparrow - 1} \end{vmatrix} \cdot \begin{vmatrix} 1 & r_{N_\uparrow + 1} & \cdots & r_{N_\uparrow + 1}^{N_\downarrow - 1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & r_N & \cdots & r_N^{N_\downarrow - 1} \end{vmatrix}.$$

考虑函数 $\varphi_\mathrm{s}^i(r_j; \mathbf{r}_{\neq j}^\uparrow,\ \mathbf{r}_{\neq j}^\downarrow)$，其定义如下：

$$\varphi_\mathrm{s}^i(r_j; \mathbf{r}_{\neq j}^\uparrow, \mathbf{r}_{\neq j}^\downarrow) = \begin{cases} r_j \cdot \Phi(\mathbf{r}^\uparrow, \mathbf{r}^\downarrow), & i = 2 \\ r_j^{i-1}, & i \le N_\uparrow, i \neq 2 \\ r_j^{i-1-N_\uparrow}, & i > N_\uparrow \end{cases},$$

显然，函数 φ_s^i 满足一种“部分”对称性，如方程（A.4）中所定义。此外，若 Ψ_s 是一个多项式或解析函数，则 φ_s^i 也具有相应性质，例如为多项式或解析函数。 □

附录 B. 拉普拉斯算子实现。 变分蒙特卡洛算法中的一个关键算法组成部分，是高效计算拉普拉斯算子，随后对参数求导以获得损失梯度。总共涉及三次求导，因此有必要简要说明其实现方式。

为简化起见，本节重点讨论回流参数化，但这些说明通常也适用。设 $\boldsymbol{x} \in (\mathbb{R} \times \mathbb{Z}_m)^N$ 为具有空间和自旋坐标的电子。波函数的每次正向求值都可以理解为一个链式过程

$$\begin{array}{ccccccccccc} (\mathbb{R} \times \mathbb{Z}_m)^N & \to & \mathbb{R}^{N \times K} & \to & \mathbb{R}^{N \times K_1} & \to & \mathbb{R}^{N \times K_2} & \to & \mathbb{R}^{N \times N} & \to & \mathbb{R}. \\ \boldsymbol{x} & \mapsto & L_K & \mapsto & A & \mapsto & \boldsymbol{A} & \mapsto & \Phi & \mapsto & \Psi, \end{array}$$

其中，K, K_1, K_2 分别表示 L_K, A 和 $\boldsymbol{A}$中的基函数数量。 L_K、 A 和 $\boldsymbol{A}$ 对应于存储输入（$x_i; \boldsymbol{x}_{\neq i}$）的 N基函数的数组。轨道 Φ 通过矩阵乘法 $\Phi = C \cdot \boldsymbol{A}$得到，其中 C 为轨道参数矩阵；随后计算波函数为 $\log\ |\Psi| = \log\ |\det\ \Phi|$，以保证数值稳定性。

梯度 $\nabla_{\boldsymbol{x}} \Psi = \nabla \Psi$ 可以通过反向微分高效地计算，但对于评估 $\Delta \Psi$ ，我们尚不了解类似的技术，只能利用前向模式微分。直接进行计算可能需要将整个海森矩阵 $\nabla_{\boldsymbol{x}}^2 \Psi$ 作为对 $\nabla_{\boldsymbol{x}} \Psi$ 进行前向模式微分的结果计算出来，然后取其迹。相反，我们发现，只需存储和计算最少的信息，即可在单次前向传播中完成计算；在我们的实现中，这种方法似乎更加高效。然而，与直接计算完整海森矩阵的方法相比，这种效率提升仅体现在前因子上。

我们首先注意到

$$\Delta\Psi = \Phi^{-\top} : \Delta\Phi - \sum_{i=1}^{N} (\Phi^{-1}\frac{\partial\Phi}{\partial x_i})^{\top} : (\Phi^{-1}\frac{\partial\Phi}{\partial x_i}),$$

其中：表示矩阵点积。因此，为了获得 $\Delta_{\boldsymbol{x}}\Psi$，我们需要 $\Delta_{\boldsymbol{x}}\Phi$和雅可比矩阵 $\nabla_{\boldsymbol{x}}\Phi$。递归地进行推导，单次前向传播即可用于计算$\Psi$， $\nabla\Psi$， $\Delta\Psi$，从而得到以下操作链：

$$\boldsymbol{x} \mapsto (L, \partial_x L_K, \partial_x^2 L_K) \mapsto (A, \nabla A, \Delta A) \mapsto (\boldsymbol{A}, \nabla\boldsymbol{A}, \Delta\boldsymbol{A}) \mapsto (\Phi, \nabla\Phi, \Delta\Phi) \mapsto \Delta\Psi.$$

关键观察是，从未对完整的海森矩阵进行求值或存储，但这需要以正向模式计算雅可比矩阵为代价。

上述观察结果使得损失函数（4.2）能够被高效地评估。最后，为了计算其相对于参数的梯度，我们现在可以使用高效的反向模式微分。

致谢。 我们感谢 Zeno Sch¨atzle 和 Juerong Feng 就本文主题所进行的启发性讨论。

参考文献

[1] N. Abrahamsen, Z. Ding, G. Goldshlager, and L. Lin, *Convergence of stochastic gradient descent on parameterized sphere with applications to variational Monte Carlo simulation*, 2023 2303 11602, https://arxiv.org/abs/ . . [2] M. Bachmayr, G. Dusson, and C. Ortner, *Polynomial approximation of symmetric functions* 2023 2109 14771 , , https://arxiv.org/abs/. .

[3] F. Becca and S. Sorella, *Quantum Monte Carlo approaches for correlated systems*, 剑- 桥大学出版社,2017.

[4] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, *Julia: A fresh approach to numerical computing* 59 2017 65 98, 10 1137 141000671 , SIAM评论, ( ), pp. – https://doi.org/ . / . [5] F. A. Bornemann and P. Deuflhard, *The cascadic multigrid method for elliptic problems*, 数值数学, 75 (1996),pp.135–152, https://doi.org/10.1007/s002110050234. [6] D. Braess and W. Dahmen, *A cascadic multigrid algorithm for the Stokes equations*, 数- 值数学, 82 (1999),pp.179–191, https://doi.org/10.1007/s002110050416. [7] A. L. Cauchy, *Mémoire sur les fonctions qui ne peuvent obtenir que deux valeurs égales et de signes contraires par suite des transpositions opérées entre les variables quélles renferment*, 巴黎综合理工学院学报, 10 (1815),pp. 29–112. [8] K. Choo, A. Mezzacapo, and G. Carleo, *Fermionic neural-network states for ab-initio electronic structure* 11 2020 10 1038 , 自然通讯, ( ), p. 2368,https://doi.org/ . / s41467-020-15724-9. [9] J. Coe, V. V. França, and I. d'Amico, *Feasibility of approximating spatial and local entanglement in long-range interacting systems using the extended Hubbard model*, 欧洲物理快报, 93 (2011), p. 10001, https://doi.org/10.1209/0295-5075/93/10001. [10] A. J. Cohen, P. Mori-Sánchez, and W. Yang, *Challenges for density functional theory*, 化学评论, 112 (2012),pp.289–320, https://doi.org/10.1021/cr200107z. [11] P. Deuflhard,*Cascadic conjugate gradient methods for elliptic partial differential equations: algorithm and numerical results* 180 1994 29 29, , 当代数学, ( ), pp.– https: //doi.org/10.1090/conm/180/01954. [12] R. Drautz, *Atomic cluster expansion for accurate and transferable interatomic potentials*, B辑物理评论, 99 (2019), p.014104,https://doi.org/10.1103/PhysRevB.99.014104. [13] R. Drautz and C. Ortner, *Atomic cluster expansion and wave function representations*, 2022 2206 11375 , https://arxiv.org/abs/ . . [14] G. Dusson, M. Bachmayr, G. Csányi, R. Drautz, S. Etter, C. van der Oord, and C. Or-tner, *Atomic cluster expansion: Completeness, efficiency and stability*, 计算- 物理学杂志, 454 (2022), p. 110946, https://doi.org/10.1016/j.jcp.2022.110946. [15] J. H. Eberly, Q. Su, and J. Javanainen, *High-order harmonic production in multiphoton ionization* 6 1989 1289 , 美国光学学会杂志B辑, ( ), pp. –1298,https: //doi.org/10.1364/JOSAB.6.001289.

[16] B. Efron and C. Stein, *The jackknife estimate of variance*,《统计年刊》, 9 (1981), 第 586–596 页, https://doi.org/10.1214/aos/1176345462。[17] E. Eldad Haber and L. Ruthotto, *Stable architectures for deep neural networks*,《逆问题》, 34 (2018), 第 014004 页, https://doi.org/10.1088/1361-6420/aa9a90。[18] R. Feynman and M. Cohen, *Energy spectrum of the excitations in liquid helium*,《物理评论》, 102 (1956), 第 1189 页, https://doi.org/10.1103/PhysRev.102.1189。[19] W. Foulkes, L. Mitas, R. Needs, and G. Rajagopal, *Quantum Monte Carlo simulations of solids* 73 2001 33 83, 10 1103 ,《现代物理评论》, ( ), 第 – 页 https://doi.org/ 。/ RevModPhys.73.33。[20] C. Friesecke, A. Gerolin, and P. Gori-Giorgi, *The strong interaction limit of density functional theory* 2022 2202 09760 ,, https://arxiv.org/abs/ 。 。[21] M. B. Giles, *Multilevel Monte Carlo methods*,《数值学报》, 24 (2015), 第 259–328 页, https: //doi.org/10.1017/S0962492915000001X。[22] G. Giuseppe and T. Matthias, *Solving the quantum many-body problem with artificial neural networks* 355 2017 602 10 1126 2302 ,《科学》, ( ), 第 – 606 页, https://doi.org/ 。/science.aag 。[23] A. Glielmo, Y. Rath, G. Csányi, A. De Vita, and G. Booth, *Gaussian process states: Adata-driven representation of quantum many-body physics*,《物理评论 X》, 10 (2020), 第 041026 页, https://doi.org/10.1103/PhysRevX.10.041026。[24] M. Griebel and J. Hamaekers, *Sparse grids for the Schrödinger equation*,《ESAIM：数学建模与数值分析》（Mod´elisation Math´ematique et Analyse Num´erique）, 41 (2007), 第 215–247 页, https://doi.org/10.1051/m2an:2007015。[25] M. Griebel, F. Y. Kuo, and I. H. Sloan, *The smoothing effect of the ANOVA decomposition*,《复杂性杂志》, 26 (2010), 第 523–551 页, https://doi.org/https://doi.org/10.1016/j. jco.2010.04.003。[26] E. Haber, L. Ruthotto, E. Holtham, and S.-H. Jun, *Learning across scales – multiscale methods for convolution neural networks*,《AAAI 人工智能会议论文集》, 32 (2018), 第 3142–3148 页, https://doi.org/10.1609/aaai.v32i1.11680。[27] J. Han, L. Zhang, and E. Weinan, *Solving many-electron Schrödinger equation using deep neural networks* 399 2019 ,《计算物理学杂志》, ( ), 第 108929 页, https://doi。org/10.1016/j.jcp.2019.108929。[28] T. Helgaker, P. Jorgensen, and J. Olsen, *Molecular electronic-structure theory*,约翰·威立父子出版公司, 2014。[29] J. Hermann, Z. Schätzle, and F. Noé, *Deep-neural-network solution of the electronic Schrödinger equation* 12 2020 891 897, 10 1038 ,《自然化学》, ( ), 第 – 页 https://doi.org/ 。/ s41557-020-0544-y。[30] H. Huang, J. Landsberg, and J. Li, *Geometry of backflow transformation ansatz for quantum many-body fermionic wavefunctions* 2021 2111 10314 ,, https://arxiv.org/abs/ 。 。[31] M. Hutter, *On representing (anti)symmetric functions*,2020, https://arxiv.org/abs/2007. 15298。[32] I. Kaluzhnyi and C. Ortner, *Optimal evaluation of symmetry-adapted n-correlations via recursive contraction of sparse symmetric tensors* 2022 2202 04140 ,, https://arxiv.org/abs/ 。 。[33] R. Kimmel, M. Elad, D. Shaked, R. Keshet, and I. Sobel, *A variational framework for Retinex* 52 2003 7 ,《国际计算机视觉杂志》, ( ), 第 –23 页, https://doi.org/ 10.1023/A:1022314423998。[34] T. Li, F. Chen, H. Chen, and Z. Wen,*Provable convergence of variational Monte Carlo methods* 2023 2303 10599 ,, https://arxiv.org/abs/ 。 。[35] J. Lin, G. Goldshlager, and L. Lin, *Explicitly antisymmetrized neural network layersfor variational Monte Carlo simulation*,《计算物理学杂志》, 474 (2023), 第 111765 页, https://doi.org/10.1016/j.jcp.2022.111765。[36] I. Loshchilov and F. Hutter, *Decoupled weight decay regularization*,2019, https://arxiv. org/abs/1711.05101。[37] Y. Lysogorskiy, C. van der Oord A. Bochkarev S. Menon M. Rinaldi T. Hammerschmidt M. Mrovec A. Thompson G. Csányi C. Ortner, and R. Drautz, *Performant implementation of the atomic cluster expansion (PACE) and application to copper and silicon* 7 2021 97 97, 10 1038 ,《npj 计算材料》, ( ), 第 – 页 https://doi.org/ 。/ s41524-021-00559-9。[38] J. Martens and R. Grosse, *Optimizing neural networks with kronecker-factored approxi-mate curvature*,载于《第 32 届国际机器学习会议论文集》, 第 37 卷, PMLR, 2015, 第 2408–2417 页, https://proceedings.mlr.press/v37/martens15.html（访问于 2023-04-06）。[39] R. M. Martin, *Electronic Structure: Basic Theory and Practical Methods*,剑桥大学出版社, 2004。

[40] R. M. Martin, L. Reining, and D. M. Ceperley, *Interacting Electrons: Theory and Computational Approaches* 2016 , 剑桥大学出版社, 。

[41] C. Ortner and et.al., *ACESchrodinger.jl*。[42] D. Pfau, J. S. Spencer, A. G. Matthews, and W. M. C. Foulkes, *Ab initio solution of themany-electron Schrödinger equation with deep neural networks*, 《Physical Review Research》, 2（2020）, 第 033429 页, https://doi.org/10.1103/PhysRevResearch.2.033429。

[43] M. Scherbela, R. Reisenhofer, L. Gerard, P. Marquetand, and C. Philipp, *Solving the electronic Schrödinger equation for multiple nuclear geometries with weight-sharing deep neural networks* 2 2022 331 341, , 《Nature Computational Science》, （）, 第 – 页, https://doi.org/10.1038/s43588-022-00228-x。[44] V. V. Shaidurov, *Some estimates of the rate of convergence for the cascadic conjugate-gradient method* 31 1996 161 171, , 《Computers & Mathematics with Applications》, （）, 第 – 页, https://doi.org/10.1016/0898-1221(95)00228-6。

[45] S. A. Smolyak, *Quadrature and interpolation formulas for tensor products of certain classes of functions* 148, 1963, , 载于《Doklady Akademii Nauk》, 第 卷, 俄罗斯科学院, 第 1042–1045 页。[46] S. Sorella, *Green function Monte Carlo with stochastic reconfiguration*, 《Physical Review Letters》, 80（1998）, 第 4558–4561 页, https://doi.org/10.1103/PhysRevLett.80.4558。[47] S. Sorella, *Generalized lanczos algorithm for variational quantum Monte Carlo*, 《Physical Review B》, 64（2001）, 第 024512 页, https://doi.org/10.1103/PhysRevB.64.024512。[48] E. M. Stoudenmire, L. O. Wagner, S. R. White, and K. Burke, *One-dimensional continuum electronic structure with the density-matrix renormalization group and its implications for density-functional theory* 109 2012 056402, , 《Physical Review Letters》, （）, 第 页, https://doi.org/10.1103/PhysRevLett.109.056402。[49] A. Szabo and N. S. Ostlund, *Modern quantum chemistry: Introduction to advanced electronic structure theory* 2012 , Courier Corporation, 。

[50] M. Thiele, E. Gross, and S. Kümmel, *Adiabatic approximation in nonperturbative time-dependent density-functional theory* 100 2008 153004, , 《Physical Review Letters》, （）, 第 页, https: //doi.org/10.1103/PhysRevLett.100.153004。[51] J. Thomas, H. Chen, and C. Ortner, *Body-ordered approximations of atomic properties*, 《Archive for Rational Mechanics and Analysis》, 246（2022）, 第 1–60 页, https://doi.org/10. 1007/s00205-022-01809-w。

[52] J. Toulouse and C. J. Umrigar, *Optimization of quantum Monte Carlo wave functions by energy minimization* 126 2007 084102, , 《The Journal of Chemical Physics》, （）, 第 页, https: //doi.org/10.1063/1.2437215。[53] C. van der Oord, G. Dusson, G. Csányi, and C. Ortner, *Regularised atomic body-ordered permutation-invariant polynomials for the construction of interatomic potentials*, 《Machine Learning: Science and Technology》, 1（2020）, 第 015004 页, https://doi.org/10.1088/ 2632-2153/ab527c。[54] L. O. Wagner, E. Stoudenmire, K. Burke, and S. R. White, *Reference electronic structure calculations in one dimension* 14 2012 8581 , 《Physical Chemistry Chemical Physics》, （）, 第 – 8590 页, https://doi.org/10.1039/C2CP24118H。