

09-VASP 的优化与并行

格致斯创 科技

2023.03.25

VASP 软件的特点

09-VASP 的 优化与并行

VASP 软件的特色
经典数值优化算法概
要
矩阵的迭代对角化
VASP 软件的主程
序结构
VASP 软件概要

VASP 软件是维也纳大学 (Universität Wien) G. Kresse 等开发的第一原理模拟软件包

- VASP 采用 PAW (Projector Augmented-Wave) 方法^[2, 3], 平衡了赝势方法和全电子计算优点, 兼顾了计算的精度和效率
- VASP 在实空间优化投影函数 (Projector), 将主要的计算过程变换到实空间完成, 大大节省了内存的开销
- VASP 通过引入多样的优化算法, 提高了矩阵对角化和电荷密度搜索的效率
- 在 VASP 的并行计算中, 有效均衡了各节点处理 FFT 变换负载和通信, 提升了软件的并行效率

相比于其他第一原理计算软件, VASP 从物理思想与方法、优化算法和并行计算实现等多个方面都有更为出色的性能

VASP 的开发团队



09-VASP 的 优化与并行

VASP 软件的特色
经典数值优化算法概
要

矩阵的迭代对角化

VASP 软件的主程
序结构

VASP 软件概要

The VASP team



o. Univ. Prof. Dr. Georg Kresse



Dr. Doris Vogtenhuber



Dr. Martijn Marsman



Dr. Merzuk Kaltak



Dr. Ferenc Karsai



Dr. Martin Schlipf

Fig.: The development team of VASP.

09-VASP 的 优化与并行

VASP 软件的特色

经典数值优化算法概
要

矩阵的迭代对角化

VASP 软件的主程
序结构

VASP 软件概要

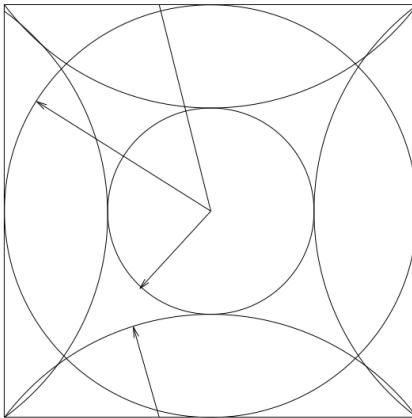
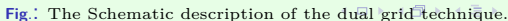


Fig.1 The small sphere contains all plane waves included in the basis set $\vec{G} < \vec{G}_{\text{cut}}$. The charge density contains components up to $2\vec{G}_{\text{cut}}$ (second sphere), and the acceleration components up to $3\vec{G}_{\text{cut}}$, which are reflected in (third sphere) because of the finite size of the FFT-mesh. Nevertheless the components a \vec{G} with $|\vec{G}| < \vec{G}_{\text{cut}}$ are correct i.e. the small sphere does not intersect with the third large sphere.



VASP 的平面波数目和内存估计

09-VASP 的 优化与并行

VASP 软件的特色
经典数值优化算法概
要

矩阵的迭代对角化

VASP 软件的主程
序结构

VASP 软件概要

总的平面波数目计算公式

$$N_{\text{PW}} = \frac{4\pi}{3} \left(\frac{\vec{G}_{\text{max}} \cdot \vec{L}}{2\pi} \right)^3 \times 1.1$$

$$= \frac{4\pi}{3} \left(\sqrt{\frac{E_{\text{cut}}}{\text{RyToeV}}} \right)^3 \times \frac{V_{\Omega}}{(\text{a.u.To}\text{\AA})^3} \times \frac{1}{8\pi^3} \times 1.1$$

其中 RyToeV 是单位换算常数 13.605826; a.u.ToÅ 是单位换算常数 0.529177249

- VASP 要求的基本 (预保留) 内存为 30M
- 当体系规模较大时, 波函数部分是内存消耗的最主要部分, 其计算公式为
平面波部分(单位 Byte)

$$16.0 \times N_{\text{PW}} \times N_{\text{Band}} \times N_{\vec{k}\text{pt}} \times N_{\text{spin}}$$

式中16是双精度浮点数的内存

局域波函数部分

$$16.0 \times N_{\text{local}} \times N_{\text{Band}} \times N_{\vec{k}\text{pt}} \times N_{\text{spin}}$$

VASP 的平面波数目和内存估计

09-VASP 的 优化与并行

VASP 软件的特色
经典数值优化算法概
要
矩阵的迭代对角化
VASP 软件的主程
序结构
VASP 软件概要

- 非局域项投影函数计算分配的内存
- FFT 变换分配内存
- 径向网格点分布分配内存
- 单中心在位项计算分配内存

并行计算时, 如果节点数为 N_{node} , 每个节点的核数 $N_{\text{core/node}}$, 则平面波数目分配估算

$$\bar{N}_{\text{PW/core}} = \frac{N_{\text{PW}}}{N_{\text{node}} \times N_{\text{core/node}}}$$

相应地, 波函数的平面波部分所需的内存估计

$$16.0 \times \min(N_{\text{PW/core}}) \times N_{\text{node}} \times N_{\text{Band}} \times N_{\vec{k}\text{pt}} \times N_{\text{spin}}$$

VASP 的优化与迭代收敛

09-VASP 的 优化与并行

VASP 软件的特色
经典数值优化算法概
要

矩阵的迭代对角化

VASP 软件的主程
序结构

VASP 软件概要

VASP 计算中, 资源消耗的主要部分是求解 Kohn-Sham 方程, 即偏微分方程 (Partial Differential Equations, PDE) 的自洽迭代, 迭代过程主要包括

- 矩阵的迭代对角化
- 电荷密度的自洽迭代

VASP 的计算高效得益于求解过程中中应用了多种经典优化算法, 保证了迭代计算的快速收敛

- 拟牛顿法 (Quasi-Newton method)
- 共轭梯度法 (Conjugate Gradients method, CG)
- 残差最小化 (RMM-DIIS) 方法

非线性方程的 Newton 法求根

09-VASP 的 优化与并行

不管哪一种数值算法，其设计原理都是将复杂转化为简单的重复，或者说，通过简单的重复生成复杂：

在算法设计和算法实现过程中，重复就是力量
迭代算法设计：“速度” vs “稳定”

VASP 软件的特色
经典数值优化算法概
要

矩阵的迭代对角化

VASP 软件的主程
序结构

VASP 软件概要

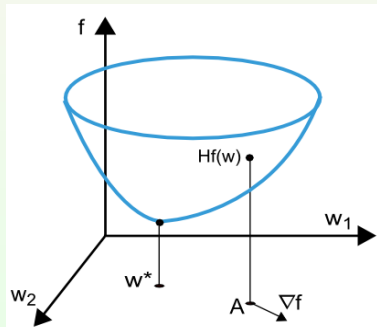
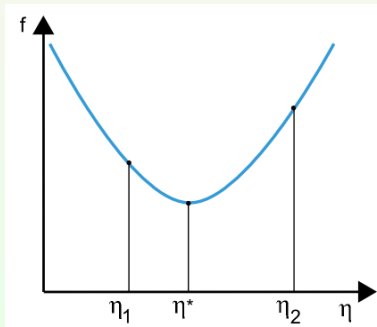
迭代优化基本思想

09-VASP 的 优化与并行

对于给定函数 f ，在极值点，函数的梯度满足

$$\nabla f = 0$$

可将函数极值问题转化成方程求根问题



Method of steepest descent

09-VASP 的 优化与并行

对于函数 $f(\mathbf{x}_0)$ 当前位置 \mathbf{x}_0 的负梯度方向 \mathbf{g}_0 满足

$$\mathbf{g}_0 = -\nabla f(\mathbf{x}_0)$$

用 \mathbf{g}_0 方向作为搜索方向,

$$\mathbf{x} = \mathbf{x}_0 + \lambda \mathbf{g}_0, \quad \lambda > 0$$

因为负的梯度方向为当前位置的最快下降方向, 所以被称为“**最陡下降法**”

对函数 f 最小化参数 λ , 可确定下一步 \mathbf{x}_1 , 可有

$$\frac{d}{d\lambda} f(\mathbf{x}_0 + \lambda \mathbf{g}_0) = \mathbf{g}_0 \cdot \nabla f(\mathbf{x}_1) = \mathbf{g}_0 \cdot \mathbf{g}_1 = 0$$

因此**最速下降法最近邻两步的梯度彼此相互垂直**

最陡下降法的收敛:

靠近极小值时收敛速度减慢, 越接近目标, 步长越小, 前进越慢

Newton-Raphson Method

09-VASP 的 优化与并行

VASP 软件的特色
经典数值优化算法概
要
矩阵的迭代对角化
VASP 软件的主程
序结构
VASP 软件概要

Newton Method 是一种在实数和复数域上近似解方程的方法。
思想: 用函数的 Taylor 级数的前几项来寻找方程 $f(x) = 0$ 的根
由 Newton 迭代公式有

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

用 Taylor 级数在 a 附近展开 $f(x)$

$$f(x) = \sum_{n=0}^{\infty} \frac{f^{(n)}(a)}{n!} (x-a)^n$$

如果只取其前两项逼近 $f(x)$, 可有

$$f(x) = f(a) + f^{(1)}(a)(x-a)$$

不难看出 $x = a - \frac{f(a)}{f^{(1)}(a)}$ 时, 有 $f(x) = 0$

Newton-Raphson Method

09-VASP 的 优化与并行

VASP 软件的特色
经典数值优化算法概
要

矩阵的迭代对角化

VASP 软件的主程
序结构

VASP 软件概要

对于函数求极值问题 (函数的导数为零), 就转换成

$$x_{n+1} = x_n - \frac{f^{(1)}(x_n)}{f^{(2)}(x_n)}$$

对高维函数, 一阶导数是梯度, 二阶导数是 **Hessian 矩阵**

$$\mathbf{H}f(\mathbf{x}) = [\frac{\partial^2 f}{\partial x_i \partial x_j}]_{n \times n}, \text{ 有}$$

$$x_{n+1} = x_n - \alpha [\mathbf{H}f(x_n)]^{-1} \nabla f(x_n) \quad n \geq 0$$

这里 α 是可调参数

- 最陡下降法是用一个平面去拟合当前位置的局部曲面
- Newton 法是用一个二次曲面拟合当前位置的局部曲面

通常情况下, 二次曲面的拟合会比平面更好, 所以牛顿法选择的路径会更符合真实的最优下降路径 (收敛更快)

Newton 法的缺点: Hessian 矩阵求逆的计算成本和复杂度较高

Quasi-Newton Method

09-VASP 的 优化与并行

VASP 软件的特色
经典数值优化算法概
要

矩阵的迭代对角化
VASP 软件的主程
序结构

VASP 软件概要

Newton 法收敛速度快, 但计算过程中需计算 Hessian 矩阵 (而且无法保证正定), 因此有了 Quasi-Newton 方法

思想: 构造可以近似 Hessian 矩阵 (或逆) 的正定对称阵

$$f(\mathbf{x}) \approx f(\mathbf{x}_{k+1}) + \nabla f(\mathbf{x}_{k+1})(\mathbf{x} - \mathbf{x}_{k+1}) + \frac{1}{2}(\mathbf{x} - \mathbf{x}_{k+1})^T \mathbf{H}_{k+1}(\mathbf{x} - \mathbf{x}_{k+1})$$

两边作用梯度算符 ∇

$$\nabla f(\mathbf{x}) \approx \nabla f(\mathbf{x}_{k+1}) + \mathbf{H}_{k+1}(\mathbf{x} - \mathbf{x}_{k+1})$$

当 $\mathbf{x} = \mathbf{x}_k$ 有

$$\mathbf{g}_{k+1} - \mathbf{g}_k \approx \mathbf{H}_{k+1}(\mathbf{x} - \mathbf{x}_k)$$

令

$$\mathbf{s}_k = \mathbf{x}_{k+1} - \mathbf{x}_k \quad \mathbf{y}_k = \mathbf{g}_{k+1} - \mathbf{g}_k$$

有

$$\mathbf{y}_k \approx \mathbf{H}_{k+1} \cdot \mathbf{s}_k \quad \text{或} \quad \mathbf{s}_k \approx \mathbf{H}_{k+1}^{-1} \mathbf{y}_k$$

Quasi-Newton 法: 靠近极小值时收敛速快; 初值选择不好, 易不收敛

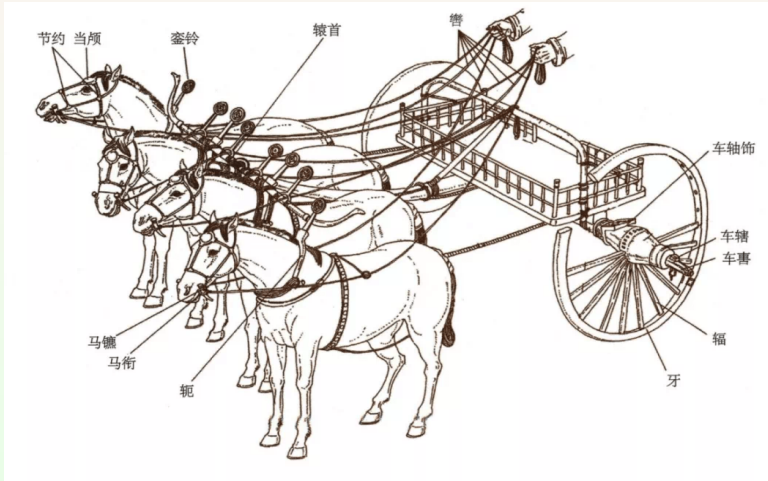
共轭梯度的“轭”

09-VASP 的 优化与并行

VASP 软件的特色
经典数值优化算法概
要

矩阵的迭代对角化
VASP 软件的主程
序结构

VASP 软件概要



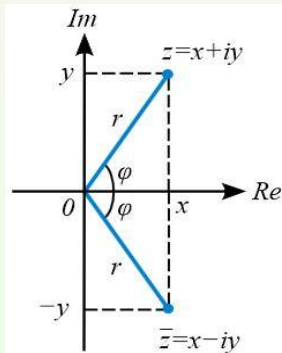
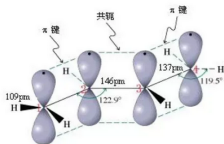
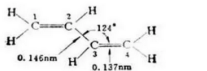
共轭的含义

09-VASP 的 优化与并行

VASP 软件的特色
经典数值优化算法概要
矩阵的迭代对角化
VASP 软件的主程序结构
VASP 软件概要



丁二烯中p-p π 键



Conjugate gradient

09-VASP 的 优化与并行

VASP 软件的特色
经典数值优化算法概
要
矩阵的迭代对角化
VASP 软件的主程
序结构
VASP 软件概要

假设函数在接近极值附件，近似有二次函数的形式

$$f(\mathbf{x}) = f_0 + \frac{1}{2} \mathbf{x} \cdot \mathbf{H} \mathbf{x} + \dots$$

其中 \mathbf{H} 是 Hessian 矩阵当 f 的偏导连续，则 \mathbf{H} 是对称矩阵，并且一般要求 \mathbf{H} 是正定的。相应的梯度表示为

$$\mathbf{g} = \nabla f(\mathbf{x}) = -\frac{\partial f}{\partial \mathbf{x}} = -\mathbf{H} \cdot \mathbf{x}$$

设从点 \mathbf{x}_i 出发沿方向 \mathbf{h}_i (不再限于梯度 \mathbf{g}_i 方向)，前进到点 \mathbf{x}_{i+1} ，根据最小化要求

$$\mathbf{h}_i \cdot \nabla f(\mathbf{x}_{i+1}) = 0$$

为确定 \mathbf{x}_{i+1} 点的继续前进方向 \mathbf{h}_{i+1} ，设 \mathbf{x}_{i+1} 可由 $\mathbf{x}_i + \lambda \mathbf{h}_{i+1}$ 得到，因此 \mathbf{x}_{i+1} 的梯度

$$\mathbf{g}_{i+1} = \nabla f(\mathbf{x}_{i+1}) = -\mathbf{H} \mathbf{x}_{i+1} = -\mathbf{H}(\mathbf{x}_i + \lambda \mathbf{h}_{i+1})$$

Conjugate gradient

09-VASP 的 优化与并行

VASP 软件的特色
经典数值优化算法概
要
矩阵的迭代对角化
VASP 软件的主程
序结构
VASP 软件概要

与方向 \mathbf{h}_i 相比, 梯度的改变为

$$\Delta \mathbf{g} = -\lambda \mathbf{H} \mathbf{h}_{i+1}$$

根据最小化要求, 梯度的改变与 \mathbf{h}_i 方向正交

$$\mathbf{h}_i \cdot \mathbf{H} \cdot \mathbf{h}_{i+1} = 0$$

共轭梯度法算法: 对于给定函数

- 已知初值 \mathbf{x}_0 和梯度 \mathbf{g}_0 , 取初始方向 $\mathbf{h}_0 = \mathbf{g}_0$ (最陡下降)
- 根据递推关系确定

$$\begin{aligned} \mathbf{g}_{i+1} &= \mathbf{g}_{i+1} - \lambda_i \mathbf{H} \mathbf{h}_i & \lambda_i &= \frac{\mathbf{g}_i \cdot \mathbf{g}_i}{\mathbf{g}_i \cdot \mathbf{H} \mathbf{h}_i} \\ \mathbf{h}_{i+1} &= \mathbf{g}_{i+1} + \gamma_i \mathbf{h}_i & \gamma_i &= -\frac{\mathbf{g}_{i+1} \cdot \mathbf{H} \mathbf{h}_i}{\mathbf{h}_i \cdot \mathbf{H} \mathbf{h}_i} \\ \mathbf{x}_{i+1} &= \mathbf{x}_i + \lambda_i \mathbf{h}_i \end{aligned}$$

共轭梯度法的收敛: 步收敛性, 稳定性高, 不需要任何外来参数

最陡下降与共轭梯度

09-VASP 的 优化与并行

VASP 软件的特色
 经典数值优化算法概要
 矩阵的迭代对角化
 VASP 软件的主程序结构
 VASP 软件概要

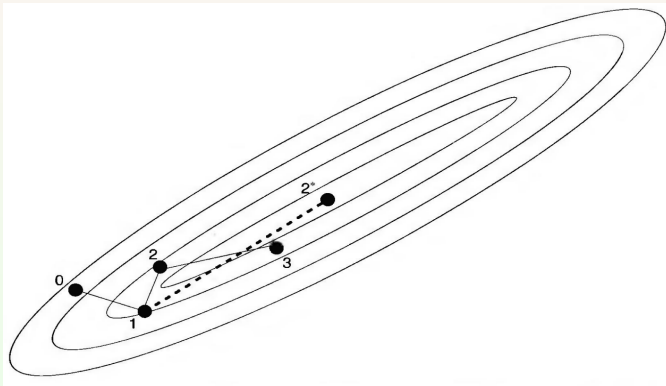


Fig.: Schematic illustration of minimization of a function in two dimensions. The steps 1,2,3, ... denote the steepest descent steps and the point ----- denote the conjugate gradient path that reaches the exact solution after two steps if the functional is quadratic.

Fixed Point

09-VASP 的 优化与并行

求解方程

$$f(x) = x$$

x 是函数 $f(x)$ 的**不动点**

对这类问题的求解, 可以利用迭代关系

$$x_{i+1} = f(x_i) \quad (i = 1, 2, 3, \dots)$$

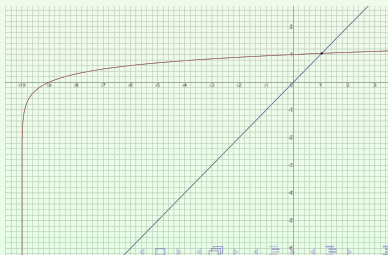
这称为**不动点迭代法**

例如求解方程

$$\lg(10 + x) = x \implies x \approx 1.04309063$$

```

x0 = 1
lg(10 + 1) = 1.041392685
x1 = 1.041392685
lg(10 + 1.041392685) = 1.0430238558
x2 = 1.043023856
lg(10 + 1.043023856) = 1.0430880104
x3 = 1.043088010
lg(10 + 1.043088010) = 1.0443090533
x4 = 1.043090533
lg(10 + 1.043090533) = 1.04430906326
x4 = 1.0430906326
lg(10 + 1.0430906326) = 1.04430906365
x5 = 1.0430906365
lg(10 + 1.0430906365) = 1.04430906366
x6 = 1.0430906366
    
```



Residue minimization Methods

09-VASP 的 优化与并行

VASP 软件的特色
经典数值优化算法概
要

矩阵的迭代对角化
VASP 软件的主程
序结构

VASP 软件概要

不动点迭代的主要问题是依赖初猜，很可能不收敛或线性收敛(收敛缓慢)，一种求解策略是定义残量

$$R[\mathbf{x}] = f(\mathbf{x}) - \mathbf{x}$$

最小化残量的模 $|R[\mathbf{x}]|$ ，特别是当残量 $R[\mathbf{x}]$ 近似是 \mathbf{x} 的线性函数，可有 **Jacobian 矩阵**

$$\mathbf{J} \equiv \frac{\delta R[\mathbf{x}]}{\delta \mathbf{x}}$$

然后可以用 Quasi-Newton 方法最小化残量
具体地可通过迭代关系求解原方程的解

$$\mathbf{x}_{i+1} = \mathbf{x}_i - \mathbf{J}^{-1} R[\mathbf{x}_i]$$

但是一般来说 Jacobian 矩阵很可能未知(或者很难求逆)，只有另图别策(在 Krylov 子空间中迭代求解)，一般常用的方法有

- 迭代子空间求逆 (Discret Inversion in the Iterative Subspace, DIIS)
- Anderson 加速或 Anderson 混合

Pulay DIIS full-subspace Method

09-VASP 的 优化与并行

DIIS 是 Pulay 引入的, 俗称的 **Pulay 混合**¹

其思想是: 对迭代逼近的矢量 \mathbf{x}_i , 可通过将之前得到的所有矢量 \mathbf{x} 线性组合得到, 组合系数则由对其残矢最小化确定。概要如下
在由 \mathbf{x}_i 构成的完全 Krylov 子空间内, 有矢量

$$\mathbf{x}_{i+1} = \sum_{j=0}^i a_j \mathbf{x}_j = c_0 \mathbf{x}_0 + \sum_{j=1}^i c_j \delta \mathbf{x}_j$$

若假设该矢量的残矢 $R[\mathbf{x}_{i+1}]$ 与矢量 \mathbf{x}_{i+1} 满足相同的线性化要求

$$R[\mathbf{x}_{i+1}] = R\left[\sum_{j=0}^i a_j \mathbf{x}_j\right] = \sum_{j=0}^i a_j R[\mathbf{x}_j]$$

通过最小化残矢的模

$$\langle R[\mathbf{x}_{i+1}] | R[\mathbf{x}_{i+1}] \rangle = \sum_{j,k} a_j a_k A_{j,k}; A_{j,k} = \langle R[\mathbf{x}_j] | R[\mathbf{x}_k] \rangle$$

即可确定矢量 \mathbf{x}_{i+1}

在电子结构计算中, 残矢最小化的约束条件最常用的有

- 电子能带的本征正交
- 电荷密度混合时, 电荷密度守恒, 即 $\sum_{j=0}^i a_j = 1$, 可有

$$a_j = \sum_k A_{j,i}^{-1} / \sum_{j,k} a_j a_k A_{j,k}^{-1} = \sum_j A_{j,i}^{-1} / \sum_{j,k} A_{j,k}^{-1}$$

¹一般地说, DIIS 是通用的不动点迭代的加速收敛方法; 对于线性问题, DIIS 等价于 GMRES 方法 ▶

Broyden Jacobian update method

09-VASP 的 优化与并行

VASP 软件的特色
经典数值优化算法概
要

矩阵的迭代对角化

VASP 软件的主程
序结构

VASP 软件概要

最初的 Broyden 方法是迭代过程中连续产生 Jacobian 逆阵的方案

- 作为迭代起点, 首先初猜合理的 \mathbf{J}_0^{-1} (比如 $\mathbf{J}_0^{-1} = \alpha \mathbf{I}$)
- 迭代开始的若干步内, 保持 \mathbf{J}^{-1} 为初猜形式; 后续 \mathbf{J}^{-1} 再逐步更新: 因为 \mathbf{J}^{-1} 始终是近似的, 有

$$\delta \mathbf{x}_i = \mathbf{x}_i - \mathbf{x}_{i-1} = -\mathbf{J}_{i-1}^{-1} R_{i-1}$$

$$\delta R_i = R_i - R_{i-1}$$

因此要求每一步迭代更新的 \mathbf{J}_i^{-1} 满足

$$0 = \delta \mathbf{x}_i - \mathbf{J}_i^{-1} \delta R_i$$

由此得到构成 $M \times M$ 的 \mathbf{J}_i^{-1} 的 M 个方程

- 最小化 Jacobian 逆阵的模变化

$$Q = \|\mathbf{J}_i^{-1} - \mathbf{J}_{i-1}^{-1}\|$$

$$\mathbf{J}_i^{-1} = \mathbf{J}_{i-1}^{-1} \frac{(\delta \mathbf{x}_i - \mathbf{J}_{i-1}^{-1} \delta R_i) \delta R_i}{\langle \delta R_i | \delta R_i \rangle}$$

改进的 Broyden 方法与 DIIS 方法的结果类似, 可以节约内存

$$Q^{\text{modified}} = \sum_{j=1}^i w_j |\delta \mathbf{x}_j - \mathbf{J}_i^{-1} \delta R_j|^2 + w_0 \|\mathbf{J}_i^{-1} - \mathbf{J}_0^{-1}\|$$

通过参数 w_j 筛选出迭代步中与之最密切关联的贡献

Anderson acceleration

09-VASP 的 优化与并行

Donald G. Anderson 给出了加速不动点迭代的求解思路

$$\mathbf{x}_1 = f(\mathbf{x}_0)$$

$$\forall k = 1, 2, \dots$$

$$m_k = \min\{m, k\}$$

$$R_k = [R_{k-m_k}, \dots, R_k]$$

$$\alpha_k = \operatorname{argmin}_{\alpha \in A_k} \|R_k \alpha\|_2, \text{ 这里 } A_k = \{\alpha = (\alpha_0, \alpha_1, \dots, \alpha_{m_k}) \text{ 并满足 } \sum_{i=0}^n \alpha_i = 1\}$$

$$\mathbf{x}_{k+1} = \sum_{i=0}^{m_k} (\alpha_k)_i f_{k-m_k+i}$$

- Anderson 本质上是 Quasi-Newton 方法求解非线性方程 (是割线方法的推广), 也可以归入 Broyden 方法一类
- Anderson 加速数学形式上可以看成 Generalized Minimal RESidual Method (GMRES) 迭代推广到非线性方程求解, 只是作了适当的截断

Anderson acceleration

09-VASP 的 优化与并行

具体地，在迭代过程中，引入中间变量

$$\mathbf{x}'_{i+1} = \alpha_k \mathbf{X}_k$$

这里 α_k 是组合系数 $\alpha_k \in A_k$, $\mathbf{X}_k = [\mathbf{x}_{k-m_k}, \dots, \mathbf{x}_k]$ 是含有最近 $m_k + 1$ 个矢量的矩阵

选择合适的 \mathbf{x}'_{k+1} 使得 $\|R(\mathbf{x}'_{k+1})\|$ 最小化

因为 α_k 的求和为 1，有一阶近似

$$R(\mathbf{X}_k \alpha_k) = R\left(\sum_{i=0}^{m_k} (\alpha_k)_i \mathbf{x}_{k-m_k+i}\right) \approx \sum_{i=0}^{m_k} (\alpha_k)_i R(\mathbf{x}_{k-m_k+i}) = R_k \alpha_k$$

因此可以通过最小化 $\|R_k \alpha\|_2$ 确定 α ，进而确定 \mathbf{x}'_{k+1}

考虑到 $f(\mathbf{x}) = \mathbf{x}$ 的精确解 \mathbf{x}^* ，因此 $f(\mathbf{x}'_{k+1})$ 可能比 \mathbf{x}'_{k+1} 更接近 \mathbf{x}^* ，因此最终方程的解选为

$\mathbf{x}_{k+1} = f(\mathbf{x}'_{k+1})$ 而非 $\mathbf{x}_{k+1} = \mathbf{x}'_{k+1}$

类似地，因为 α_k 的求和为 1，有一阶近似

$$f(\mathbf{x}'_{k+1}) = f\left(\sum_{i=0}^{m_k} (\alpha_k)_i \mathbf{x}_{k-m_k+i}\right) \approx \sum_{i=0}^{m_k} (\alpha_k)_i f(\mathbf{x}_{k-m_k+i}) = \sum_{i=0}^{m_k} (\alpha_k)_i f_{k-m_k+i}$$

最终确定方程的解为

$$\mathbf{x}_{k+1} = \sum_{i=0}^{m_k} (\alpha_k)_i f_{k-m_k+i}$$

矩阵的迭代对角化

09-VASP 的 优化与并行

VASP 软件的特色
经典数值优化算法概
要

矩阵的迭代对角化

VASP 软件的主程
序结构

VASP 软件概要

- 矩阵的直接对角化计算复杂复 $O(N^3)$
- 矩阵的迭代对角化计算复杂度 $O(N_0^2 \times N \ln N)$ $N_0 \ll N$

迭代求本征值的思想是 Jacobian 于 1846 年提出的^[1]
其基本思想是

$$(H - \varepsilon^n)|\psi^n\rangle = |R[\psi^n]\rangle$$

这里 n 是迭代步数, $|\psi^n\rangle$ 和 ε^n 分别是本征态和本征值, $|R[\psi^n]\rangle$ 是残差矢量

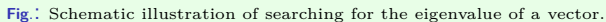
在电子态计算过程中, 选择适当的基函数, 可以使 Schrödinger 方程的矩阵接近对角阵
因此可有

$$\begin{aligned} |\psi^{n+1}\rangle &= \mathbf{D}^{-1}(\mathbf{H} - \varepsilon)|\psi^n\rangle + |\psi^n\rangle = \delta|\psi^{n+1}\rangle + |\psi^n\rangle \\ \mathbf{D}\delta\psi^{n+1} &= R[\psi^n] \quad \text{或} \quad \delta\psi^{n+1} = \mathbf{D}^{-1}R[\psi^n] \equiv \mathbf{K}R[\psi^n] \end{aligned}$$

这里 \mathbf{D} 是非奇异矩阵, 与 \mathbf{H} 矩阵有关
 $\mathbf{K} = \mathbf{D}^{-1}$, 也叫“预处理矩阵”, 可根据需要选取多种形式

- 要求 \mathbf{D} 比原始的 $\mathbf{H} - \varepsilon$ 更易求逆阵
- 要求 \mathbf{D} 使得修正项 $\delta\psi^{n+1}$ 能够使 ψ^n 尽可能更接近正确的本征矢

VASP 软件的特色
经典数值优化算法概要
矩阵的迭代对角化
VASP 软件的主程序结构
VASP 软件概要



Krylov 子空间与矩阵迭代

09-VASP 的 优化与并行

对于矩阵 \mathbf{A} , 取任意矢量 ψ_0 (要求归一), 构造矢量 ψ_1 (同样要求归一, 并与 ψ_0 正交), 满足

$$\mathbf{A}\psi_0 = a_0\psi_0 + b_0\psi_1$$

由此确定 a_0 、 b_0 、 ψ_1

$$a_0 = \langle \psi_0 | \mathbf{A} | \psi_0 \rangle$$

$$b_0\psi_1 = \mathbf{A}\psi_0 - a_0\psi_0$$

$$\|\psi_1\| = 1$$

进而可构造 ψ_2 :

$$\mathbf{A}\psi_1 = c_1\psi_0 + a_1\psi_1 + b_1\psi_2$$

要求 ψ_2 与 ψ_0 、 ψ_1 正交归一条件, 确定 ψ_2 , a_1 , b_1 , c_1

Krylov 子空间与矩阵迭代

09-VASP 的 优化与并行

VASP 软件的特色
经典数值优化算法概
要
矩阵的迭代对角化
VASP 软件的主程
序结构
VASP 软件概要

根据递推关系有

$$\mathbf{A}\psi_p = \sum_{q=0}^{p-1} c_p^{(q)} \psi_q + a_p \psi_p + b_p \psi_{p+1}$$

这里 ψ_{p+1} 将与所有之前的 ψ_q (processors) 正交
利用矩阵 \mathbf{A} 的 Hermitian, 因此对于 $q < p - 1$ 各项, 可有等式

$$c_p^{(q)} = \langle \psi_q | \mathbf{A} \psi_p \rangle = \langle \mathbf{A} \psi_q | \psi_p \rangle = 0$$

即矢量 ψ_p 垂直于矢量 $\mathbf{A}\psi_q$, 由此可得

$$c_p^{(p-1)} = \langle \psi_{p-1} | \mathbf{A} \psi_p \rangle = \langle \mathbf{A} \psi_{p-1} | \psi_p \rangle = b_{p-1}$$

经过 p 步迭代后

$$\mathbf{A}\psi_p = b_{p-1} \psi_{p-1} + a_p \psi_p + b_p \psi_{p+1}$$

这里 ψ_{p+1} 要求与 ψ_p 和 ψ_{p-1} 满足正交归一条件

矩阵对角化的 Lanczos 算法

09-VASP 的 优化与并行

VASP 软件的特色
经典数值优化算法概
要
矩阵的迭代对角化
VASP 软件的主程
序结构
VASP 软件概要

因此矩阵 A 可以用 ψ_p (称为 Lanczos 矢量) 为基组表示成三对角阵形式(稀疏矩阵)

$$A^p = \begin{pmatrix} a_1 & b_2 & & & & 0 \\ b_2 & a_2 & b_3 & & & \\ & b_3 & a_3 & \ddots & & \\ & & \ddots & \ddots & b_{p-1} & \\ 0 & & & b_{p-1} & a_{p-1} & b_p \\ & & & & b_p & a_p \end{pmatrix}$$

不难看出, 经过 p 次 Lanczos 迭代, 当 $b \rightarrow 0$ 即达到收敛, 意味着此时 $p \times p$ 三对角阵 A^p 的本征值也将收敛到矩阵 A 的本征值

- 稀疏矩阵 A^p 可通过快速 QR 分解得到本征值
- 三对角阵的最低和最高本征值随着迭代次数增加收敛最迅速
- Lanczos 方法适用于少量本征值与剩余本征有较大差值的体系

矩阵对角化的 Lanczos 算法

09-VASP 的 优化与并行

VASP 软件的特色
经典数值优化算法概
要

矩阵的迭代对角化

VASP 软件的主程
序结构

VASP 软件概要

大型稀疏矩阵对角化基本思路便是从一个试探向量 \mathbf{c}_0 出发，通过矩阵-向量乘操作²，同时保持矩阵的稀疏性，使得试探向量逐渐收敛到目标特征向量（往往是基态对应的特征向量）

本征值求解对应如下优化问题

$$\lambda_{\min} = \min_{\mathbf{c}} \rho(\mathbf{c}) = \min_{\mathbf{c}} \frac{\mathbf{c}^H \mathbf{c}}{\mathbf{c}^T \mathbf{c}}$$

利用最陡下降法求解上述优化问题，则需要计算函数的梯度

$$\nabla \rho(\mathbf{c})|_{\mathbf{c}=\mathbf{c}_0} = \frac{2}{\mathbf{c}^T \mathbf{c}} (\mathbf{H}\mathbf{c}_0 - \rho(\mathbf{c}_0)\mathbf{c}_0)$$

实际计算中无需求出梯度并作精确搜索，因为解一定在 Krylov 子空间 $\text{span}(\mathbf{c}_0, \mathbf{H}\mathbf{c}_0)$ 内；只需计算这两个向量之间的哈密顿矩阵元，并对角化所得到的矩阵便相当于做了一步最速下降法

经过 k 步迭代之后所得到的解在子空间 $\text{span}(\mathbf{c}_0, \mathbf{H}\mathbf{c}_0, \dots, \mathbf{H}^k \mathbf{c}_0)$

因此大型矩阵对角化的问题，转化为子空间内矩阵对角化的问题

选取合适的初猜，经过若干次迭代之后，子空间内最小特征值可能于真实的最小特征值非常接近³一般地， k 次迭代后第 j 个特征向量写成

$$\mathbf{c}_j^{(k)} = \sum_{i=0}^k c_{i,j}^{(k)} \mathbf{H}^i \mathbf{c}_0$$

² 由于矩阵是稀疏的，从而可以快速进行矩阵-向量乘这一基本操作（时间复杂度 $O(N^2)$ ）

³ 子空间内若干个最小特征值都可能于相应的真值非常接近

矩阵迭代对角化

09-VASP 的 优化与并行

VASP 软件的特色
经典数值优化算法概
要

矩阵的迭代对角化
VASP 软件的主程
序结构

VASP 软件概要

稀疏矩阵求解的 Lanczos 优化过程，只变动一个分量 \mathbf{c}_I 的前提下

$$\left. \frac{\partial \rho}{\partial \mathbf{c}_I} \right|_{\mathbf{c}_I + \delta_I} = 0$$

是可以精确求解的，其解为

$$\delta_I = (\rho(\mathbf{c}_0) - \mathbf{H}_{II})^{-1} \mathbf{q}_I \quad \text{这里 } \mathbf{q} = (\mathbf{H} - \rho \mathbf{I}) \mathbf{c}_0$$

不难看出，矢量 \mathbf{q} 就对应 Jacobi 迭代中用于判断收敛的残差矢量
更一般地，求解方程

$$\left. \frac{\partial \rho}{\partial \mathbf{c}} \right|_{\mathbf{c} + \delta} = 0$$

将方程展开到二阶近似，不难有

$$(\rho - \mathbf{H}_{II}) \delta_I \approx \mathbf{q}_I + \sum_{J \neq I} \delta_J + (\rho - \lambda) \mathbf{c}_I$$

实际计算中选则 $\delta_I = (\rho(\mathbf{c}_0) - \mathbf{H}_{II})^{-1} \mathbf{q}_I$ 并不是方程的解的好的近
似，**好处是计算比较简单**

Block-Davison algorithm

09-VASP 的 优化与并行

VASP 软件的特色
经典数值优化算法概
要

矩阵的迭代对角化
VASP 软件的主程
序结构

VASP 软件概要

Davidson 方法是求解大型稀疏矩阵的少量本征值问题提出来的, 结合了 Lanczos 优化和 Jacobi 迭代的优点, 简言之就是改进初猜, 不用 $\mathbf{H}\mathbf{c}_0$, 而改用计算简单的 $\delta_I = (\rho(\mathbf{c}_0) - \mathbf{H}_{II})^{-1}\mathbf{q}_I$ 形式

应用 Davison 方法可以快速地依次求解稀疏矩阵的少量本征值和本征矢, 将该方法推广为同时求解若干个本征态, 即块-Davidson 方法

- 选取合适数目正交归一的向量 $\mathbf{c}_1, \mathbf{c}_2, \dots, \mathbf{c}_n$ 作为初猜子空间的基组, 计算并储存向量 $\mathbf{H}\mathbf{c}_i$ 和矩阵元 $\mathbf{H} = \langle \mathbf{c}_i | \mathbf{H} | \mathbf{c}_j \rangle$
- 对角化矩阵, 得到本征值 λ^n 和本征矢 \mathbf{a}^n
- 构造残量矢量 $\mathbf{q}_M = (\mathbf{H} - \lambda^{(M)}\mathbf{I})\mathbf{a}^{(M)}$ 其中 $\mathbf{a}^{(M)} = \sum_{i=1}^M a_i^{(M)}\mathbf{a}_i$
- 根据模长 $\|\mathbf{q}_M\|$ 判断迭代收敛情况
- 构造 $\delta_{I,M+1} = (\lambda^{(M)} - \mathbf{H}_{II})^{-1}\mathbf{q}_{I,M}$, 与此前的基组正交归一化, 得到 \mathbf{c}_{M+1}
- 计算矩阵元 $\mathbf{H}_{i,M+1} \quad i = 1, 2, \dots, M+1$
- 对角化矩阵得到新的本征值和本征矢量, 继续迭代

RMM-DIIS

09-VASP 的 优化与并行

前述矩阵迭代对角化方法的优化策略都是

- 通过迭代优化得到最小本征值 (极值)
- 利用本征态正交, 依次获得其他各本征态和本征值

RMM-DIIS (Residual Minimization Method by Direct Inversion in the Iterative Subspace)⁴方法则可以不用引入正交条件而得到多个本征值, 因为该方法最小化的不是本征值而是残矢

其基本思想概要: 在 n 维 Krylov 子空间内, 生成矢量

$$\psi^{n+1} = c_0 \psi^0 + \sum_{j=1}^{n+1} c_j \delta \psi^j$$

通过改变选取一套合适的系数 c_j 来完成 ψ^{n+1} 的残矢 R^{n+1} 的最小化。等价于 c_j 由 $\{\psi^0, \psi^1, \dots, \psi^n\}$ 构成的 Krylov 子空间内求 Hermitian 本征值问题

$$\sum_{j=1}^n \langle R^i | R^j \rangle c_j = \epsilon \sum_{j=1}^n \psi^i | S | \psi^j \rangle c_j$$

每迭代一次, 子空间引入一个新波函数 ψ 和一个新残矢 $R(\psi)$

- RMM-DIIS 的计算量瓶颈将是后续的逐个矩阵-向量乘操作 $H\psi$
- 只要内存许可, RMM-DIIS 构造的完整的子空间内, 构成子空间的矢量本征值都可以求解出来
- 因为 RMM 方法对初猜的矢量敏感 (矢量收敛的位置到离初猜较近)

⁴ RMM-DIIS 的得名源自该方法的提出者 Pulay: 该方法的基本思想是在历次迭代产生的矢量构成的完整 Krylov 子空间内, 完成对残矢的最小化

VASP 的并行效率

09-VASP 的 优化与并行

与同类型软件相比，VASP 有着优异的并行能力

VASP 软件的特色
经典数值优化算法概要
矩阵的迭代对角化
VASP 软件的主程序结构
VASP 软件概要

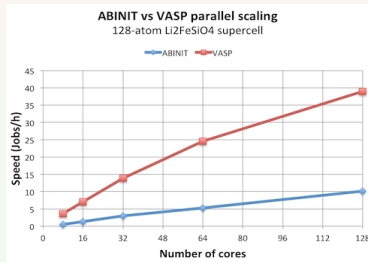
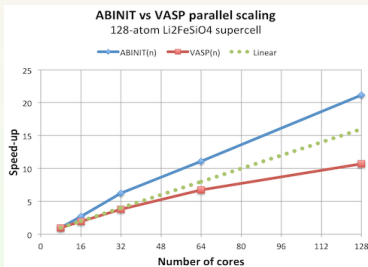


Fig.1 The comparison of parallel scaling for ABINIT vs VASP.

- VASP 迭代对角化约束了矩阵的维度，减少了对角化过程中的迭代次数，保证了 MPI 并行的规模和扩展性
- VASP 实施 FFT 变换时，保证各节点上处理的网格负载均衡

VASP 计算的 FFT 并行实现

09-VASP 的 优化与并行

VASP 软件的特色
经典数值优化算法概
要
矩阵的迭代对角化
VASP 软件的主程
序结构
VASP 软件概要

- 中间层设计: FFT 网格、实空间基组与计算节点的匹配
通过子程序 `mgrid.F` 生成中间层, 实现并行负载与计算节点分配的匹配, 减少 FFT 变换和实空间并行的节点间通信

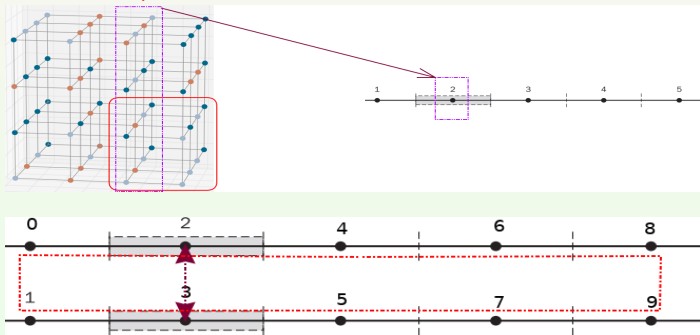


Fig.: VASP: Reciprocal-Real space layout for grids in MPI.

VASP 的通信开销

09-VASP 的 优化与并行

在高性能的计算队列中，VASP 的并行上限可以突破 256 核，但当并行核数超过百核数量级，并行效率下降非常明显

VASP 软件的特色
经典数值优化算法概
要

矩阵的迭代对角化

VASP 软件的主程
序结构

VASP 软件概要

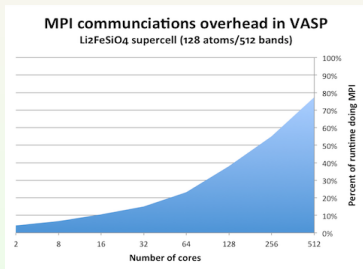


Fig.1 Time spent in MPI calls with increasing the number of ranks in a VASP calculation.

如能对并行系统与 VASP 结合作深度改造 (如国家超算天津中心方案), VASP 的并行扩展可以到 10^4 核级别, 但这一改造需要对底层代码和计算框架作较大规模改动

VASP 的 GPU 加速



09-VASP 的 优化与并行

VASP 软件的特色
经典数值优化算法概
要

矩阵的迭代对角化

VASP 软件的主程
序结构

VASP 软件概要

NVIDIA 多年来致力于 VASP 的 GPU 加速，取得了一定的成效

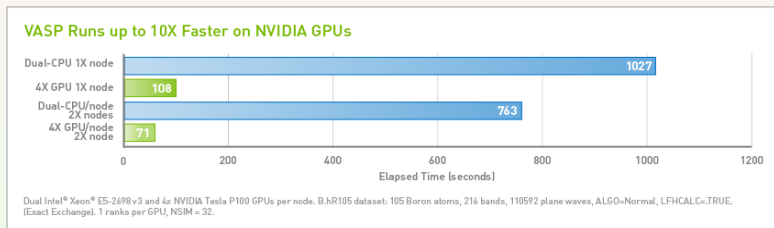


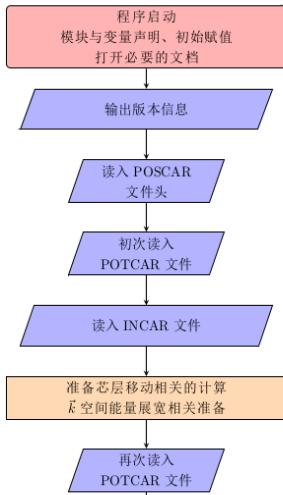
Fig.: Compare of VASP calculation with GPU and CPU.

- 通用配置下，GPU 对 VASP 计算有加速效果，一般可提升 4~6 倍
- 矩阵对角化的并行算法限制了 GPU 在第一原理计算中的应用
- GPU 加速的模式主要适合于分子动力学计算

VASP 的主程序结构

09-VASP 的 优化与并行

VASP 软件的特色
经典数值优化算法概要
矩阵的迭代对角化
VASP 软件的主程序结构
VASP 软件概要



VASP 的主程序结构 (cont.)

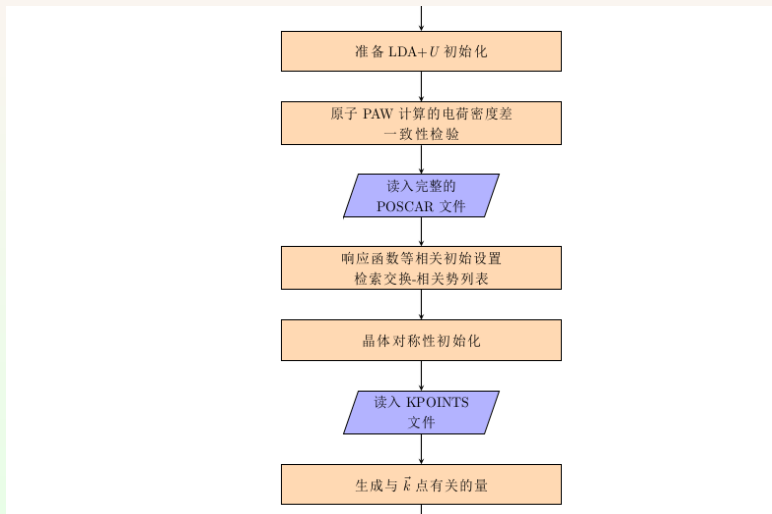
09-VASP 的 优化与并行

VASP 软件的特色
经典数值优化算法概
要

矩阵的迭代对角化

VASP 软件的主程
序结构

VASP 软件概要



VASP 的主程序结构 (cont.)

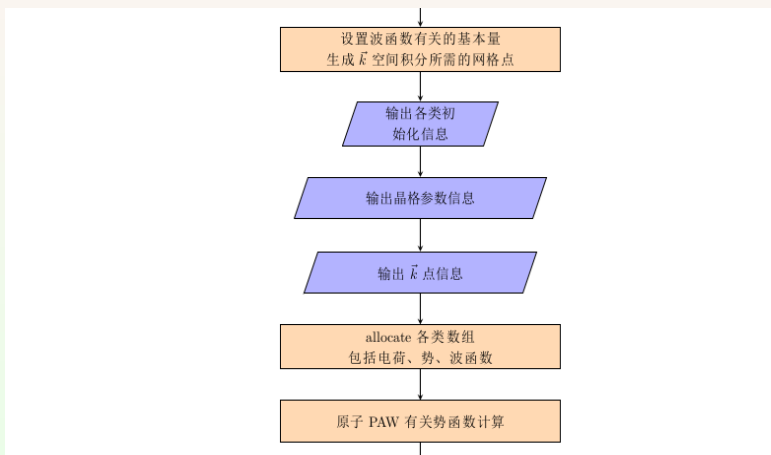
09-VASP 的 优化与并行

VASP 软件的特色
经典数值优化算法概
要

矩阵的迭代对角化

VASP 软件的主程
序结构

VASP 软件概要



VASP 的主程序结构 (cont.)

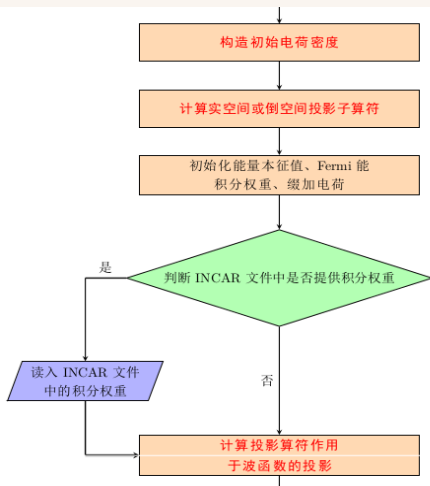
09-VASP 的 优化与并行

VASP 软件的特色
经典数值优化算法概
要

矩阵的迭代对角化

VASP 软件的主程
序结构

VASP 软件概要



VASP 的主程序结构 (cont.)

09-VASP 的 优化与并行

VASP 软件的特色
经典数值优化算法概
要
矩阵的迭代对角化
VASP 软件的主程
序结构
VASP 软件概要



VASP 的主程序结构 (cont.)

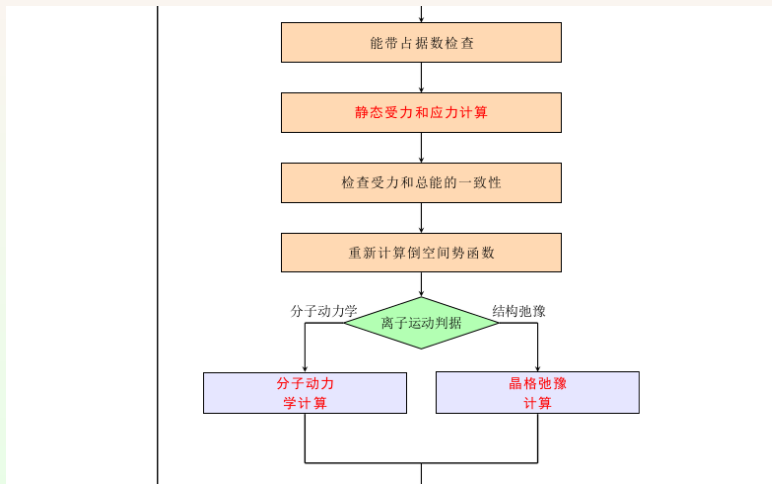
09-VASP 的 优化与并行

VASP 软件的特色
经典数值优化算法概
要

矩阵的迭代对角化

VASP 软件的主程
序结构

VASP 软件概要



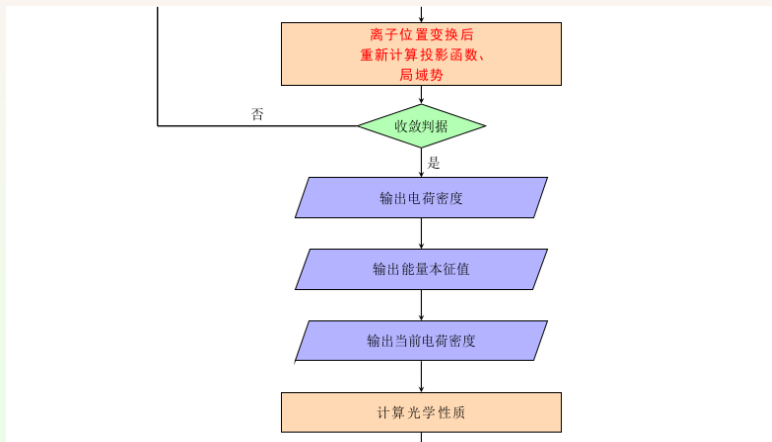
VASP 的主程序结构 (cont.)

09-VASP 的 优化与并行

VASP 软件的特色
经典数值优化算法概
要

矩阵的迭代对角化
VASP 软件的主程
序结构

VASP 软件概要



VASP 的主程序结构 (cont.)

09-VASP 的 优化与并行

VASP 软件的特色
经典数值优化算法概
要

矩阵的迭代对角化

VASP 软件的主程
序结构

VASP 软件概要

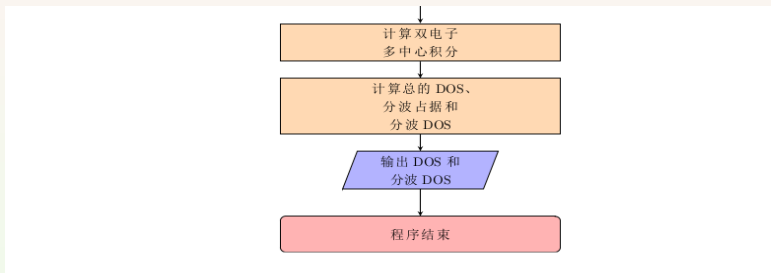


Fig.1 The Flow of main program for VASP.

VASP 的迭代收敛

09-VASP 的 优化与并行

VASP 软件的特色
经典数值优化算法概要
矩阵的迭代对角化
VASP 软件的主程序结构
VASP 软件概要

完整的 VASP 计算流程是离子-电子的耦合自洽迭代，称为从头算分子动力学 (Ab Initio Molecular Dynamics, AIMD)^[4, 5]

- AIMD 通过在动力学系统中引入经典力学的绝热能量，将 DFT 与 MD 关联起来，实现电子与离子运动在同一动力学框架内处理，同时又在时间尺度上保持分离
- AIMD 框架下，DFT 到 MD 跨尺度无需再借助势函数模拟，电子弛豫过程与分子动力学可以用类似的迭代方式计算，大大降低了程序的复杂度

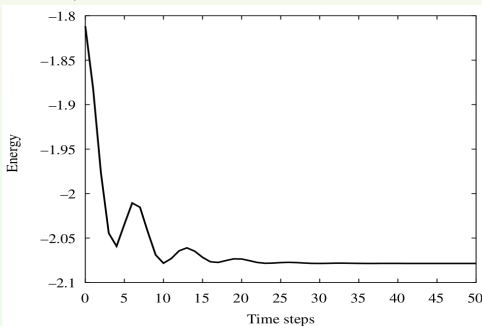


Fig.: The Flow of calculation for the AIMD.

作为第一性原理计算的商用软件，VASP 已成为计算材料学领域应用最广泛的软件之一。全球绝大多数超算中心都安装了 VASP，据统计，VASP 软件的作业机时占用全球总机时的 12~20%，但由于其属于重型浮点计算密集型应用，实际耗电量占比则高达 30~50%

- **物理上**，VASP 基于 DFT 近似，求解 Kohn-Sham 方程，并将粒子基态密度问题转化为矩阵的本征函数和本征值问题
- **数学上**，方程求解过程的核心是矩阵对角化与 PDE 的自洽迭代，即便对于简单体系，也需要完成数十次的迭代，而规模大的计算模拟体系则可能需要成千上万次迭代计算
- **计算过程上**，VASP 计算的时长开销主要是本征值求解的矩阵对角化；此外由于算法限制，Kohn-Sham 方程作为线性方程组作并行处理时，节点间存在密集的通信。在上千节点，上万计算核的大规模并行系统上，数据通信将严重影响程序的性能，这是当前 VASP 软件的主要瓶颈

有必要探索新的并行和优化策略来提升 VASP 的计算性能

主要参考文献

09-VASP 的 优化与并行

VASP 软件的特色
经典数值优化算法概
要

矩阵的迭代对角化
VASP 软件的主程
序结构

VASP 软件概要

- [1] C. G. Jacobi, *Über ein leichtes Verfahren die in der Theorie der Säculärstörungen vorkommenden Gleichungen numerisch aufzulösen*, *Crelle's J.* **30** (1846), 51-94
- [2] P. E. Blöchl. *Phys. Rev. B*, **50** (1994), 17953
- [3] G. Kresse and D. Joubert *Phys. Rev. B*, **59** (1999), 1758
- [4] R. Car and M. Parrinello *Phys. Rev. Lett.*, **55** (1985), 2471
- [5] K. Laasonen and A. Pasquarello and R. Car and C. Lee and D. Vanderbilt *Phys. Rev. B*, **47** (1993), 10142
- [6] Richard. M. Martin. *Electronic Structure: Basic Theory and Practical Methods* (Cambridge University Press, Cambridge, England, 2004)
- [7] D. J. Singh. *Plane Wave, PseudoPotential and the LAPW method* (Kluwer Academic, Boston, USA, 1994)

09-VASP 的 优化与并行

VASP 软件的特色
经典数值优化算法概
要
矩阵的迭代对角化
VASP 软件的主程
序结构
VASP 软件概要

谢谢大家！