Lecture XI : Iterative solutions for solving the Kohn-Sham equation

I. THE DIAGONALIZATION PROBLEM IN PLANEWAVE CALCULATIONS

The Kohn-Sham Hamiltonian when written in the planewave basis yields a matrix with a large dimension, which creates a very tight bottleneck in calculation speed since exact diagonalization scales like N^3. Moreover, in most electronic structure calculations only the first few eigenvalues are needed among the large number of eigenvalues that necessarily result from traditional diagonalization techniques. In order to overcome this difficulty, several alternative diagonalization schemes were either borrowed from linear algebra or deviced from scratch by physicists. All of these methods yield approximate solutions to the eignevalue equation however by careful tuning, results come out to be very close to the exact eigenvalues and eigenvectors. This problem is avoided in localized bases where the size of the basis set is much smaller and direct diagonalization costs little time.

Most of these methods rely on a repeated application of the Hamiltonian operator on an initial guess for an eigenvector and thus creating a sequence of vectors that span a vector space.

II. ITERATIVE METHODS

The common features of iterative methods are discussed extensively in literature (see, for example, Wood and Zunger *J. Phys. A : Math. Gen.*, **18**,1985).

The problem we have at hand is finding the n lowest eigenvalues of a generalized eigenvalue problem

$$\hat{H}|\psi_i\rangle = \lambda \hat{S}|\psi_i\rangle \tag{1}$$

where the dimension of the problem is N and $n \ll N$. In the so-called iterative methods, there are three common basis sets to be determined

- The basis set, $\{|\phi_i\rangle\}$ that is used in expressing the Hamiltonian, $\hat{H}$. In our case, this is the planewave basis.
- A complete N-dimensional set $\{\psi_i\}$ that spans the entire Hilbert space of the Hamiltonian and by means of which all vectors in the Hilbert space may be expressed.
- A much smaller N_b-dimensional spanning set $\{|b_i\rangle\}$. The assumption common to all iterative methods is that all eigenvectors of interest may be expanded in terms of this N_b dimensional vector space. N_b here satisfies $n \sim N_B \ll N$.

In all of the iterative methods, one starts with an $N_0 \times N_0$ submatrix of the Hamiltonian, $\hat{H}_0$ and finds its eigenvalues $\{\lambda_j^0\}$ and eigenvectors $\{|a_j^0\rangle\}$exactly. In the subsequent iterations, these starting eigenvectors are improved iteratively to reach convergence. The approximations to the eigenvalues and eigenvectors are calculated via the solution of the eigenvalue equation

$$\hat{H}^r|c\rangle = \epsilon \hat{S}^r|c\rangle \tag{2}$$

where the reduced operators $\hat{H}^r$ amd $\hat{S}^r$ are written in the basis $\{|B_i\rangle\}$ as

$$H_{ij}^r = \langle b_i|H|b_j\rangle \tag{3}$$

$$S_{ij}^r = \langle b_i|S|b_j\rangle \tag{4}$$

$$\tag{5}$$

At the kth step then the approximation to the real eigenvector is given by ϵ_k and the approximation to the corresponding eigenvector is given by

$$|a_k\rangle = \sum_i \langle b_i|c_k\rangle |b_i\rangle. \tag{6}$$

After each step new vectors are calculated either to add to the expansion set or directly to improve the approximation to the eigenvector being sought.

Iterative methods may either be *sequential* meaning that a single eigenvector is calculated at a time or in the *block* form where several eigenvalues emerge simultaneously.

Different methods are characterized by their choice of the spanning spaces $\{|b_i\rangle\}$ and $\{|\psi_i\rangle\}$. Some of these methods are briefly discussed below :

A. The Newton method

The quality of the approximation is measured by a residual vector defined by

$$|R(|A^a\rangle, E^a)\rangle = (\hat{H} - E^a\hat{S})|A^a\rangle. \tag{7}$$

where E^a and $|A^a\rangle$ are the approximate eigenvalues and eigenvectors respectivey. The residual as defined in Eq. 7 is in turn used to defined the quantity

$$R = \left(\frac{\langle R|R\rangle}{\langle A^a|S|A\rangle}\right)^{1/2}. \tag{8}$$

R is a measure of how far the approximation to the eigenvalue and the eigenvector is from the true ones and for the exact solution would be zero. The goal for each iteration then would be to make an addition $|\delta A\rangle$ to the approximation vector $|A^a\rangle$ such that the residual is a minimum. This is satisfied by demanding

$$|R(|A^a + \delta A\rangle, E^a)\rangle = |R(|A^a\rangle, E^a)\rangle + (\hat{H} - E^a\hat{S})|\delta A\rangle = 0. \tag{9}$$

The $|\delta A\rangle$ that satisfies Eq. 9 is given by

$$|\delta A\rangle = -(\hat{H} - E^a\hat{S})^{-1}|R(|A^a\rangle, E^a)\rangle \tag{10}$$

which however takes us back to the initial problem and thus cannot be solved exactly. Instead several methods (such as Newton,...) make a an approximation to the inverse operator in Eq. 10

$$|\delta A\rangle = -D^{-1}|R\rangle \tag{11}$$

where D is the diagonal of the $\hat{H} - E^a\hat{S}$. Thus in the next iteration, the approximate eigenvector is replaced by

$$|A^{new}\rangle = |A^a\rangle + |\delta A\rangle \tag{12}$$

where $|\delta A\rangle$ is defined in Eq. 11. This update step may be written in basis expansion form as

$$|\delta A\rangle = -\sum_i{}' \frac{\langle x_i|R\rangle|x_i\rangle}{\langle x_i|\hat{H} - E^a\hat{S}|x_i\rangle} \tag{13}$$

where the prime on the sum enforces the omission of such term that give a denominator that is below a certain threshold.

B. The Lanczos method

The Lanczos method is based on a repeated improvement of the basis set in the following manner. The vectors in the basis set in the ith step is formed by the repeated application of the Hamiltonian on initial guess vector $|a_j^0\rangle$[1].

$$\{|b_j\rangle^k\} = \left[|a_j^0\rangle, \hat{H}|a_j^0\rangle, \hat{H}^2|a_j^0\rangle, \cdots \hat{H}^k|a_j^0\rangle\right] \tag{14}$$

The remaining operations for each iteration are as described above in the Newton method. The vectors in Eq. 14 formed at each iteration must be orthonormalized to all the previous vectors.

C. The Davidson method

The Davidson method is described by

$$\begin{aligned} \{|x_i\rangle\} &= \{|e_i\rangle\} \\ \{|b_i\rangle^k\} &= [|a_j^0\rangle, |b_j^1\rangle = |\delta A_1\rangle - \sum_{i=1}^{j-1}\langle b_i|\delta A_1\rangle|b_i\rangle] \end{aligned} \tag{15}$$

III. OPTIMIZATION METHODS

A. Steepest descent

B. Conjugate gradients

---

[1] The iterative vector space thus formed is called a Krylov subspace