# Iterative Diagonalisation

Dyutiman Das

**Abstract**: The diagonalisation of a Hamiltonian by constrained minimisation techniques is discussed. The minimisation is carried out by the **conjugate gradient** method. The conjugacy criterion is derived for the unconstrained case. The errors introduced by the constrains are also emphasised.

**Minimisation Methods**: The Hamiltonian can be diagonalised by iterative minimisation schemes such as "conjugate gradient". This is inspired by the fact that, of the set of all normalised vectors, the one which minimises the expectation value of the Hamiltonian (i.e. the energy) is the ground state eigen-vector.

The minimising (normalised) vector $|\varphi\rangle$, can always be written as some linear combination of the eigen-states $\{|\psi_i\rangle\}$, hence the expectation value of the Hamiltonian is some combination of the eigen-values and has the minimum value when $|\varphi\rangle$ is a combination of only the eigen-sates corresponding to the lowest eigenvalue. Mathematically,

$$|\varphi\rangle = \sum_i a_i|\psi_i\rangle, \qquad \sum_i |a_i|^2 = 1$$

The expectation value of the Hamiltonian between the state $|\varphi\rangle$ is

$$\langle\varphi|H|\varphi\rangle = \sum_i a_i^\star a_i\langle\psi_i|H|\psi_i\rangle = \sum_i |a_i|^2\lambda_i\langle\psi_i|\psi_i\rangle = \sum_i \lambda_i|a_i|^2$$

This is minimum when only the lowest state is present, since the presence of any other state increases the energy. Hence, in the absence of degeneracy, the ground state wave function minimises the energy.

So, the problem of finding the lowest eigen-state becomes one of minimising the expectation value of the Hamiltonian subject to the constraint of normalisation of the eigen-state. Once the lowest eigen-state has been determined, if the energy is minimised again under an additional constraint that the minimising vector be orthogonal to the already determined lowest eigen-state (since the eigen-states of a hermitian operator corresponding to different eigen-values are orthogonal), the eigen-vector corresponding to the next higher eigen-value will be determined (neglecting degeneracies). In this way, all the higher states could be obtained.

The "conjugate gradient" minimisation scheme is well suited for constrained multivariate minimisation, which is the task at hand.

**Steepest-Descent**: The process of determining the $i$-th eigen-vector, with all the lower eigen-vectors already determined, involves the minimisation of the expectation value of the Hamiltonian under the constraint that the minimising vector is normalised ( orthogonality constraints will be explicitly treated later). Thus,

$$E[|\psi_i\rangle] = \langle\psi_i|H|\psi_i\rangle, \qquad \text{under the constraint } \langle\psi_i|\psi_i\rangle = 1$$

where $|\psi_i\rangle$ is the $i$-th eigenvector or *band*. The constraint equation is included by a Lagrangian multiplier $\lambda$, which constraints the solution to be on the surface of an

1

N-dimensional sphere of unit radius.

$$\mathcal{L} = \langle \psi_i | (H - \lambda\mathbb{I}) | \psi_i \rangle$$

The iteration scheme starts with an arbitrary point on the sphere, i.e. with a starting vector $|\psi_i^0\rangle$ of unit norm[1]. The first step is to move in the direction in which the decreased of the function $E[|\psi_i^m\rangle]$ is fastest. Here, $m$ denotes the iteration index. This is the direction of *steepest-descent* and is determined by the negative gradient of the function at that point on the sphere (since the increase of the function is fastest *in* the direction of the gradient).

The steepest-descent direction for the $i$-th band at the $m$-th iteration step is thus given by the negative gradient of $\mathcal{L}$

$$|\zeta_i^m\rangle = -(H - \lambda\mathbb{I})|\psi_i^m\rangle$$

Now, this vector should be such that when we start moving from the tip of the starting vector, we are still on the surface of the N-dimensional sphere, so that the constraint is obeyed. Hence, $|\zeta_i^m\rangle$ must be orthogonal to $|\psi_i^m\rangle$. This determines the Lagrangian multiplier for that iteration step[2].

$$\lambda_i^m = \langle \psi_i^m | H | \psi_i^m \rangle$$

i.e. the best guess for the eigen-value at that point. This value, also called the *Ritz value* is a function of the approximate eigen-vector $|\psi_i^m\rangle$.

Orthogonality Constraint: The minimisation of the energy corresponding to the $i$-th band is under an additional constraint of orthogonality to all the lower bands. Hence, the search direction, which is constructed from the steepest-descent vector, should be orthogonalised to all the lower bands, as otherwise it would add some components along directions forbidden by the constraint. This is done by the Gram-Schmidt orthogonalisation

$$|\zeta_i'^m\rangle = |\zeta_i^m\rangle - \sum_{j<i} \langle \psi_j | \zeta_i^m \rangle |\psi_j\rangle$$

Since the lower bands $|\psi_j\rangle, \forall\ j\ <\ i$ have been determined, they do not have any iteration index. Its evident that $\langle \psi_i^m\ |\ \zeta_i'^m \rangle = 0$.

Preconditioning: If exact arithmetic is used, the number of iterations taken by the conjugate gradient method to reach the solution of the unconstrained minimisation problem is equal[3] to the number of *distinct* eigen-values of $H$, i.e. a highly degenerate problem converges faster. Thus, if the original system is replaced by an equivalent system, in which the matrix has many unit eigen-values, i.e. the condition number is as small as possible, a faster convergence can be achieved. The idea of *preconditioning* is to construct a transformation to have this effect on $H$, which is essentially constructing the inverse of $H$. However, the evaluation if the inverse is essentially solving the whole problem, so some inexpensive approximations have to be found.

To this effect, we observe that each iteration will reduce the magnitude of the steepest-descent vector which is a measure of the error in the eigen-state, vanishing at

---

[1]This vector is of course orthogonalised against all the lower and already determined bands.

[2]Set $\langle \psi_i^m | \zeta_i^m \rangle$ equal to zero and solve for $\lambda$.

[3]This will be motivated later

the solution. If the steepest-descent vector was just a scalar multiple of the error, then moving the correct distance along the steepest descent direction would completely eliminate the error in the eigen-state.

The relationship between the error in the eigen-state $|\delta\psi_i^m\rangle$ , and the steepest-descent direction $|\zeta_i^m\rangle$ is demonstrated by a simple expansion of the error in the basis of the eigen-states of the Hamiltonian

$$|\delta\psi_i^m\rangle = \sum_j c_{i,j}|\psi_j\rangle$$

Of course, by definition $|\psi_i\rangle = |\psi_i^m\rangle + |\delta\psi_i^m\rangle$. Substituting this $|\psi_i^m\rangle$ into the steepest-descent vector we get

$$|\zeta_i^m\rangle = -(H - \lambda_i^m\mathbb{I})|\psi_i\rangle + (H - \lambda_i^m\mathbb{I})|\delta\psi_i^m\rangle$$

Near the solution, the norm of the first term is much smaller than the second term[5], and hence can be neglected. So we have

$$
\begin{aligned}
|\,\zeta_i^m\rangle &= (H - \lambda_i^m\mathbb{I})\sum_j c_{i,j}\,|\,\psi_j\rangle \\
&= \sum_j (\lambda_j - \lambda_i^m)c_{i,j}\,|\,\psi_j\rangle
\end{aligned}
$$

It is observed that the steepest-descent vector is a scalar multiple of the error $|\delta\psi_i^m\rangle$ if the Hamiltonian is $n$-fold degenerate, i.e. $\lambda_j = \lambda, \ \forall j = 1\ldots n$. Thus, a *preconditioned* Hamiltonian with an increased degeneracy will converge faster.

We know that two matrices related by a similarity transformation have the same set of eigen-values[4], hence *preconditioning* can be achieved by a similarity transformation of the Hamiltonian which makes the system highly degenerate. Let us denote $G \equiv (H - \lambda\mathbb{I})$, i.e. neglecting the wave-function dependence of the *Ritz* value. Then, the minimisation problem reduces to solving

$$G|\psi_i^m\rangle \equiv -|\zeta_i^m\rangle = 0$$

If $K$ is a symmetric, positive-definite matrix, then we can replace the above system by

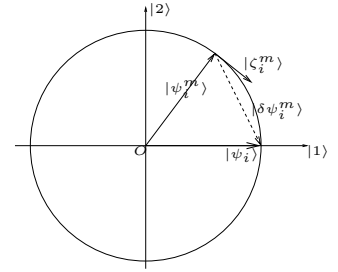$$KGK|\Psi_i^m\rangle = 0, \qquad \text{with } |\psi_i^m\rangle = K|\Psi_i^m\rangle$$

Let $R$ denote the matrix $KGK$; then $R$ is related to $K^2G$ by a similarity transformation, i.e. $KRK^{-1} = K^2G$. Since solving $G|\psi_i^m\rangle = 0$ is same as solving $K^2G|\psi_i^m\rangle = 0$, the whole problem transforms into minimising

$$\mathcal{L} = \langle\Psi_i|R|\Psi_i\rangle, \qquad \text{or solve } R|\Psi_i\rangle = 0$$

The matrix $K$, called the *preconditioner* should be chosen depending on the type of the Hamiltonian being dealt with.

The broad eigen-value spectrum of the Kohn-Sham Hamiltonian in pseudopotential caluclations that use plane-wave basis sets is associated with the wide range of energies of the basis states. The higher-energy states are dominated by the plane-wave basis states whose kinetic energies are close to the eigen-value of the state. To make these high-energy states degenerate, **the inverse of the kinetic energy operator**



PSfrag replacements

The error in the eigen-state for a $2 \times 2$ case

---

[4]A Similarity Transformation is $H \sim AHA^{-1}$ where $A$ is a non-singular matrix

**would be a good preconditioner**. However, this does not hold for the lower-nergy states, and hence the elements of the preconditioning matrix should become a constant for the lower-energy states, rather than varying as the inverse of the kinetic energy.

So to correct for the error in the steepset-descent vector, it is multiplied with a *preconditioning* matrix $K$ whose matrix elements are given by

$$K_{i,j} = \frac{27 + 18x + 12x^2 + 8x^3}{27 + 18x + 12x^2 + 8x^3 + 16x^4} \delta_{ij}$$

where $x$ is the ratio of the kinetic energy corresponding to that basis vector to the total kinetic energy of that state $|\psi_i^m\rangle$. The matrix elements $K_{ij}$ approach unity as $x \to 0$, with zero first, second, and third derivatives. This ensures that the small wave-vectors remain unchanged. Above $x = 1$, $K_{ij}$ asymptotically approach $1/[2(x-1)]$ with an asymptotic expansion correct to fourth order in $1/x$.

For the transformed problem, the steepest-descent vector is thus given by

$$|\eta_i^m\rangle = -R|\Psi_i^m\rangle = -(KGK)K^{-1}|\psi_i^m\rangle = K|\zeta_i'^m\rangle$$

This preconditioning destroys the orthogonality constraint, and so this vector has to be explicitly orthogonalised to all the lower bands and also the $i$-th band itself.

$$|\eta_i'^m\rangle = |\eta_i^m\rangle - \langle\psi_i^m|\eta_i^m\rangle|\psi_i^m\rangle - \sum_{j<i}\langle\psi_j|\eta_i^m\rangle|\psi_j\rangle$$

Conjugate Directions: After the initial step along the steepest-descent direction, the "Conjugate Gradient" method proceeds along the conjugate directions such that the minimisation at each iteration is independent of the of the previous ones. This is the condition for "conjugacy" [2]. These are generally derived for the unconstrained quadratic problem. For the case under consideration, the assumption of the Lagrangian multiplier being independent makes the problem quadratic, however the constraints make the problem highly non-linear. Progress is made by assuming that the constraints do not significantly alter the conjugate directions. This point will be discussed more in the next section.

In the problem thus far, $|\Psi_i^m\rangle$ is an approximation of the vector corresponding to the minimum of $\mathcal{L}$. Consider a subspace of linearly independent vectors $|\varphi_i^0\rangle, |\varphi_i^1\rangle, \ldots, |\varphi_i^m\rangle$ and let $\Phi_i^m$ be a matrix with columns $|\varphi_i^0\rangle, |\varphi_i^1\rangle, \ldots, |\varphi_i^m\rangle$. If the next approximation of the minimum is obtained by moving from the present approximation [5] over the manifold $\Psi_i^m + \Phi_i^m$, then the solution is obtained by

$$\min_{w \in \Re^{m+1}} \mathcal{L}(|\Psi_i^m\rangle + \Phi_i^m|w\rangle)$$

Minimising this Lagrangian function with respect to $|w\rangle$ we obtain

$$|w\rangle = (\Phi_i^{m\dagger} R \Phi_i^m)^{-1} \Phi_i^{m\dagger} |\eta_i'^m\rangle$$

where $|\eta_i'^m\rangle = R|\Psi_i^m\rangle$ is the preconditioned and orthogonalised steepest-descent vector. Hence the next approximation to the preconditioned eigen-state is given by

$$|\Psi_i^{m+1}\rangle = |\Psi_i^m\rangle + \Phi_i^m(\Phi_i^{m\dagger} R \Phi_i^m)^{-1} \Phi_i^{m\dagger} |\eta_i'^m\rangle$$

---

[5]This will however be modified by the constraints

The iterates in this procedure display several interesting properties. Firstly, $|\eta_i'^m\rangle$, the preconditioned gradient at $|\Psi_i^{m+1}\rangle$, is orthogonal to the vectors $\{|\phi_i^m\rangle\}$. Note that

$$
\begin{aligned}
\Phi_i^{m\dagger}|\eta_i^{m+1}\rangle &= -\Phi_i^{m\dagger}R|\Psi_i^{m+1}\rangle \\
&= \Phi_i^{m\dagger}|\eta_i^m\rangle - \Phi_i^{m\dagger}R\Phi_i^m(\Phi_i^{m\dagger}R\Phi_i^m)^{-1}\Phi_i^{m\dagger}|\eta_i'^m\rangle \\
&= 0
\end{aligned}
$$

Therefore
$$
\langle\eta_i'^{m+1}|\varphi_i^s\rangle = 0, \qquad s = 0,\ldots,m
$$

Hence, if each previous iterate $\Psi_i^s, \forall\ s \leq m$, then by induction it follows

$$
\langle\eta_i'^r|\varphi_i^s\rangle = 0, \qquad r > s
$$

Thus we get,
$$
|\Psi_i^{m+1}\rangle = |\Psi_i^m\rangle + \beta\Phi_i^m(\Phi_i^{m\dagger}R\Phi_i^m)^{-1}|e_m\rangle
$$

where $|e_m\rangle$ is the $m$-th column of the identity operator and $\beta = \langle\eta_i'^m|\varphi_i^m\rangle$.

To simplify the situation, the matrix $(\Phi_i^{m\dagger}R\Phi_i^m)^{-1}$ is constructed to be diagonal. Also, each of the columns of $\Phi_i^m$ are chosen to be *conjugate* with respect to the matrix $R$, i.e.
$$
\langle\varphi_i^r|R|\varphi_i^s\rangle = 0, \qquad r \neq s
$$

When this holds,
$$
|\Psi_i^{m+1}\rangle = |\Psi_i^m\rangle + \alpha^m|\varphi_i^m\rangle
$$

with $\alpha^m = \langle\eta_i'^m|\varphi_i^m\rangle/\langle\varphi_i^m|R|\varphi_i^m\rangle$. This is the step to the minimum for the unconstrained problem, however this is **not** used under the orthogonality constraints.

From the definitions we find that

$$
|\eta_i'^{m+1}\rangle - |\eta_i'^m\rangle = -R(|\Psi_i^{m+1}\rangle - |\Psi_i^m\rangle) = -\alpha^m R|\varphi_i^m\rangle
$$

Let $|y_i^m\rangle$ denote the vector $|\eta_i^{m+1}\rangle - |\eta_i^m\rangle$. Hence the conjugacy condition $\langle\varphi_i^r|R|\varphi_i^s\rangle = 0$ is equivalent to the orthogonality condition $\langle y_i^r|\varphi_i^s\rangle = 0$.

A set of mutually conjugate directions can be ontained by taking $|\varphi_i^0\rangle$ as the steepest-descent direction $|\eta_i'^m\rangle$ and computing each subsequent direction as a linear combination of $|\eta_i'^m\rangle$ and the previous $m$ search directions

$$
|\varphi_i^m\rangle = |\eta_i'^m\rangle + \sum_{j=0}^{m-1}\gamma_i^{mj}|\varphi_i^j\rangle
$$

All the components of the previous equation are linear combinations of $\{|\varphi_i^m\rangle\}$ and with $|\varphi_i^0\rangle = |\eta_i'^0\rangle$ it holds trivially

$$
\langle\eta_i'^m|\eta_i'^n\rangle = 0, \qquad n < m
$$

Moreover, $|\varphi_i^m\rangle$ can be constructed to be conjugate to $\{|\varphi_i^s\rangle\} \forall\ s < m$. Thus premultiplying $|\varphi_i^m\rangle$ with $\langle\varphi_i^n|R$ (note $R$ is symmetric) we obtain for $n = 0,\ldots,m-1$

$$
\begin{aligned}
\langle\varphi_i^n|R|\varphi_i^m\rangle &= \langle\varphi_i^n|R|\eta_i'^m\rangle + \sum_{j=0}^{m-1}\gamma_i^{mj}\langle\varphi_i^n|R|\varphi_i^j\rangle \\
&= -\frac{1}{\alpha^m}\langle y_i^m|\eta_i'^m\rangle + \gamma_i^{mn}\langle\varphi_i^n|R|\varphi_i^n\rangle
\end{aligned}
$$

The orthogonality of the $|\eta_i'^m\rangle$s implies that the first term vanishes for all $n < m - 1$, thus from the conjugacy condition we have $\gamma_i^{mn} = 0$, $\forall\ n < m - 1$. Thus only the coefficient $\gamma_i^{m,m-1} \equiv \gamma_i^{m-1}$ is non-zero. Hence, the value of $\gamma_i^{m-1}$ that ensures that $|\varphi_i^m\rangle$ is conjugate to $|\varphi_i^{m-1}\rangle$, we premultiply $|\varphi_i^m\rangle$ by $\langle y_i^{m-1}|$ and apply the orhtogonality condition that $\langle y_i^{m-1}|\varphi_i^m\rangle = 0$; this gives[2]

$$
\begin{aligned}
0 &= \langle y_i^{m-1}|\eta_i'^m\rangle + \gamma_i^{m-1}\langle y_i^{m-1}|\varphi_i^{m-1}\rangle \\
\gamma_i^{m-1} &= -\frac{\langle y_i^{m-1}|\eta_i'^m\rangle}{\langle y_i^{m-1}|\varphi_i^{m-1}\rangle}
\end{aligned}
\tag{1}
$$

Hence, the conjugate-gradient direction is constructed out of the steepest-descent direction as indicated above

$$
|\varphi_i^m\rangle = |\eta_i'^m\rangle + \gamma_i^{m-1}|\varphi_i^{m-1}\rangle
$$

where $\gamma_i^{m-1}$ is given as above, and $\gamma_i^1 = 0$, which means that the first step is along the regular steepest-descent direction, but then on its along the conjugate direction constructed as above.

This conjugate direction will be orthogonal to all the other bands, because it was constructed out of the steepest-descent vectors which had been orthogonalised against all other bands. However, the conjugate directions from the previous iterations will not be orthogonal to the present band, and hence this preconditioned conjugate direction will have to be explicitly orthonormaalised against the $i$-th band

$$
|\varphi_i''^m\rangle = |\varphi_i^m\rangle - \langle \psi_i^m|\varphi_i^m\rangle|\psi_i^m\rangle
$$

$$
|\varphi_i'^m\rangle = \frac{|\varphi_i''^m\rangle}{\langle \varphi_i''^m|\varphi_i''^m\rangle^{1/2}}
$$

A point of notation, in the last two equations we changed from the transformed eigen-state $|\Psi_i^m\rangle$ to the original $|\psi_i^m\rangle$.

**Line Minimisation:** So far the solution is still at the tip of the starting guess for the eigen-vector, and should move along the conjugate dirction derived above. However, the constraints restrict the solution to be on the surface of a sphere of unit radius. Moving from the surface along *any* direction orthogonal to the radius vector at that point will obviously violate this constraint. Hence, the next update of the eigen-vector is taken as a combination of $|\psi_i^m\rangle$ and $|\varphi_i'^m\rangle$, so that the resulting vector is still normalised [6]

$$
|\psi_i^{m+1}\rangle = \cos\theta|\psi_i^m\rangle + \sin\theta|\varphi_i'^m\rangle \qquad (\theta\ \text{real})
$$

where $\theta$ is some parameter which minimises the expectation of the hamiltonian. This vector is a normalised vector that is orthogonal to all the *lower* bands.

However, the imposition of this constraint violates the assumptions under which the conjugate direction was derived, namely that the next step would be taken from the tip of the previous solution.

To illustrate this point consider a $3 \times 3$ system, projected on a $2D$ plane for convinience. The current solution $|\psi_i^m\rangle$ is at the point $B$, and the conjugate direction *at* $B$ is $|\varphi_i'^m\rangle$ which is in the plane of the paper and orthogonal to $|\psi_i^m\rangle$. The updated

ements



The Constrints for a $3 \times 3$ system

---
[6]Note that we have already normalised the conjugate vector
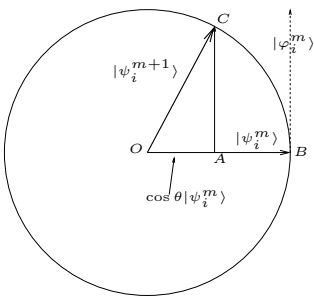
vector is $|\psi_i^{m+1}\rangle$ is obtained by moving from $A$ along $|\varphi_i'^m\rangle$ which is a conjugate direction at $B$ and not $A$. For a $3D$ problem, the conjugate direction at $A$ could have been in any direction perpendicular to $|\psi_i^m\rangle$, i.e. it need not lie in the plane of the paper.

An exact solution would require determination of the conjugate directions at all points along $OB$, the form of which is not necessarily the same as the one derived in the last section. However, if we assume that $\gamma_i^{m-1}$ has the same form, then $|\varphi_i^m\rangle$ has the same value.

To find the value of $\theta = \theta_{\min}$ for which the expectation of the Hamiltonian is minimised, we see that

$$E(\theta) = \langle \psi_i^m(\theta)|H|\psi_i^m(\theta)\rangle$$

and the $\theta_{\min}$ which minimises this is

$$\tan 2\theta = \frac{2\langle \varphi_i'^m|H|\psi_i^m\rangle}{\langle \psi_i^m|H|\psi_i^m\rangle - \langle \varphi_i'^m|H|\varphi_i'^m\rangle}$$

**The Effect of Constrains**:  In absence of the constrains the formula for $\gamma_i^{m-1}$ derived above is equivalent to the more popular form[1]

$$\gamma_i^{m-1} = \frac{\langle \eta_i'^m \mid \zeta_i'^m \rangle}{\langle \eta_i'^{m-1} \mid \zeta_i'^{m-1} \rangle} \tag{2}$$

But under the constrains, this is not true, since the vectors are only approximately conjugate, and most of the orthogonalisation conditions do not hold exactly. However, it should be noted that this formula requires additional storage of $\mid \zeta_i'^m\rangle$, whereas in the previous case, the preconditioned steepest-descent vector $\mid \eta_i'^m\rangle$ overwrites the steepest-descent vector.

The code however converges much faster for Eqn. 1 than for 2. The bulk of the computation is in the evaluation of the term $H|\phi_i'^m\rangle$, and this is done once per iteration, since $H|\psi_i^m\rangle$ can be evaluated in terms of previously evaluated terms $H|\psi_i^{m-1}\rangle$ and $H|\phi_i'^{m-1}\rangle$

Results:   To test this algorithm a simple matrix form was assumed

$$H_{i,j} = \begin{cases} i^{2/3} & \text{if } i = j \\ \sqrt{i+j} - \text{int}\sqrt{i+j} - 0.5 & \text{if } i \neq j \end{cases}$$

This is a diagonally dominant matrix with the diagonal elements representing the kinetic energy of the Hamiltonian whereas the off diagonal elements are like the potential energy terms.
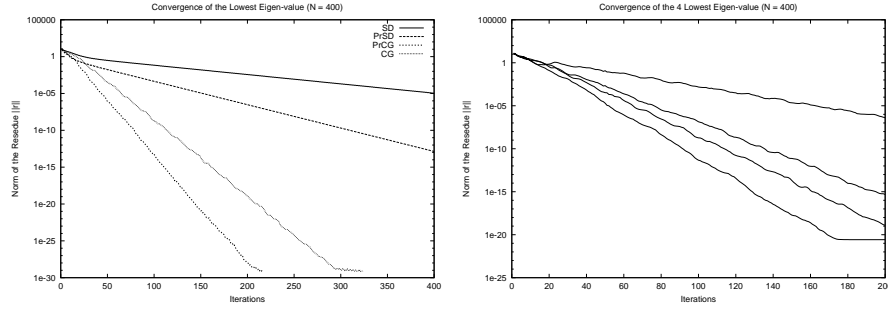


Figure 1: A Matrix of size 400 was tested. The first figure shows the improvements of convergence of the conjugate gradient method over the steepest-descent, and also how the preconditioner accelerates convergence. The second figure shows the convergence of some of the higher eige-nvalues .

# References

[1] Payne M. C., Teter M. P., Allan D. C., Arias T. A., and Joannopoulos J. D., 1992 *Rev. Mod. Phys.* **64** 1045

[2] Gill P. E., W. Murray, and M. H. Wright, 1981, *Practical Optimization* (Academic, London)

[3] Payne M. C., Teter M. P., Allan D. C. 1989 *Phys. Rev.* B **40** 12255

[4] Stich I., Car R., Parinello M., and Baroni S. 1989 *Phys. Rev.* B **39** 4997

[5] Denoting $\mid A \rangle = -(H - \lambda_i^m \mathbb{I}) \mid \psi_i \rangle$, the norm is

$$\langle A \mid A \rangle = \langle \psi_i \mid (H - \lambda_i^m \mathbb{I})^2 \mid \psi_i \rangle = (\Delta\lambda)^2 \quad \text{small near the solution}$$

while with $\mid B \rangle = (H - \lambda_i^m \mathbb{I}) \mid \psi_i \sum_j c_{i,j} \mid \psi_i \rangle$, the norm is

$$\langle B \mid B \rangle = \sum_j \mid c_{i,j} \mid^2 (\lambda_j - \lambda_i^m)^2 \qquad \text{not small}$$