

Diagonalization of a real-symmetric Hamiltonian by genetic algorithm: A recipe based on minimization of Rayleigh quotient

SUBHAJIT NANDY¹, PINAKI CHAUDHURY² and S P BHATTACHARYYA*

Department of Physical Chemistry, Indian Association for the Cultivation of Science, Jadavpur, Kolkata 700 032, India

Present address: ¹Andrew's High (H.S.) School, Kolkata 700 031, India

²Ramakrishna Mission Vidyamandira, Belur Math, Howrah 711 202, India

e-mail: pcspb@mahendra.iacs.res.in

MS received 6 October 2003; revised 17 June 2004

Abstract. A genetic algorithm-based recipe involving minimization of the Rayleigh quotient is proposed for the sequential extraction of eigenvalues and eigenvectors of a real symmetric matrix with and without basis optimization. Important features of the method are analysed, and possible directions of development suggested.

Keywords. Rayleigh quotient minimization; real-symmetric Hamiltonian eigenvalue problem; genetic algorithm based diagonalisation; stochastic diagonalisation.

1. Introduction

Solution of the time-independent Schrödinger equation $\hat{H}\mathbf{y}_n = E_n\mathbf{y}_n$, in principle, provides us with complete knowledge of the system. Exact solution of the equation is feasible only for a handful of problems, while for most of the systems one is compelled to introduce a basis, preferably an orthonormal basis $\{\mathbf{f}\}$ in terms of which the stationary state \mathbf{y} is expanded, viz. $\mathbf{y} = \sum_i c_i \mathbf{f}_i$. A variational ansatz then leads to a matrix eigenvalue problem,

$$HC = EC, \quad \mathbf{y} = \mathbf{f}C,$$

H being a real symmetric matrix constructed in the basis $\{\mathbf{f}\}$. Diagonalization of H leads to the eigenvalues E_n and the corresponding eigenvectors C_n . Many numerical methods are available^{1–5} for diagonalization of H , all of them being deterministic methods. It is not always relevant to find all the eigenvalues and eigenvectors – only a few eigenvalues and vectors including the lowest one may often be the target. Can we develop a direct search method for extracting the few eigenvalues and eigenvectors of interest from the total search space? It is indeed possible to do so by reducing the eigenvalue problem to an extremization problem. Even then, no de-

terministic search seems feasible as the best-known ones require computation of gradient and/or Hessian of the function being minimized or maximized during the search.⁶ Non-deterministic or stochastic search methods may be good candidates if one can construct a suitable function whose maximization or minimization would be equivalent to finding the desired eigenvalue and would thus be the target of the search. We must mention here that such a direct stochastic search-based scheme of diagonalization would not generally be competitive with the traditional deterministic diagonalization techniques when a fixed basis set is used. Stochastic diagonalization could be the method of choice when combined with simultaneous basis optimization, for diagonalization along with basis optimization is a complex nonlinear problem and it is for such nonlinear problems, that stochastic diagonalization schemes could outperform traditional techniques.

In quantum chemistry, one often makes use of the linear variation method which leads to a real symmetric matrix eigenvalue problem. In a traditional CI method the basis set is fixed.

The search space in the problem of diagonalization in a fixed basis $\{\mathbf{f}\}$ consists of linear expansion parameters $\{c_i\}$ only. To cut down the length of the expansion, it may also be necessary to search for the optimal nonlinear parameters in basis functions

*For correspondence

for expansion while diagonalizing H . This immediately brings us to the problem of a complex simultaneous search for optimal linear and nonlinear parameters in the search space to determine the optimal expansion parameters for $\{\mathbf{y}_n\}$ that make H diagonal. In what follows, we explore the possibility of developing a genetic algorithm^{7,8} driven method for extracting a few eigenvalues of a real symmetric Hamiltonian matrix with a simultaneous search for an optimal basis set of expansion. Earlier, we reported⁹ a modified Jacobi–GA route to eigenvalues and vectors. In the present communication, we explore a method based on the minimization of Rayleigh quotient which is easier to implement.

2. The method

Genetic algorithms^{7,8} simulate evolution of a population of probable solutions on a fitness landscape under the action of a set of genetic operators like selection (mimicking natural selection), crossover (mimicking chromosomal crossover in genetics) and mutation (mimicking sudden changes in one or more genes). The genetic operators act on individual members of the evolving population, each member being represented in the form of a linear string, an integer string or a string of floating point numbers. In our problem, each member of the population would represent a distribution of probability amplitudes of the evolving wavefunction solution strings projected on members of the particular basis set of expansion. Let \mathbf{y}_i be the i th member of the evolving population when

$$\mathbf{y}_i = \sum_{p=1}^m c_{pi} \mathbf{f}_p, \quad i = 1, 2, \dots, n,$$

$$\langle \mathbf{f}_p | \mathbf{y}_i \rangle = c_{pi}, \quad p = 1, 2, \dots, m. \quad (1)$$

The string S_i corresponding to \mathbf{y}_i is given by

$$S_i \equiv (c_{0i}, c_{1i}, c_{2i} \cdots c_{pi} \cdots c_{mi}), \quad (2)$$

while for another member \mathbf{y}_k of the population, we have the string S_k where

$$S_k \equiv (c_{0k}, c_{1k}, c_{2k} \cdots c_{pk} \cdots c_{mk}). \quad (3)$$

Each one of the strings (S_i) is characterized by a unique value of fitness (f_i) which is assigned in the following way. The Rayleigh quotient \mathbf{r}_i for the i th wavefunction string ($S_i \equiv \mathbf{y}_i$), we note, is defined as follows²

$$\mathbf{r}_i = \frac{C_i^\dagger H C_i}{C_i^\dagger C_i}. \quad (4)$$

If C_i happens to be the ground eigenvector of H (i.e. \mathbf{y}_i is the ground eigenvector of \hat{H}),^{4,5}

$$\nabla \mathbf{r}_i = \frac{2[H - \mathbf{r}_i]C_i}{C_i^\dagger C_i} = 0. \quad (5)$$

Condition (5) would not generally be satisfied by any wavefunction string in the initial population, except by accident. We may exploit this fact to generate a fitness landscape f by taking,

$$f_i = \exp(-\mathbf{I}(\nabla \mathbf{r}_i)^\top \mathbf{r}_i). \quad (6)$$

Clearly, $f_i \rightarrow 1$, as $\nabla \mathbf{r}_i \rightarrow 0$, signalling that the evolution has hit the true ground state eigenvector of H in the vector C_i . The factor \mathbf{I} in (6) is chosen to ensure that there is no exponential over flow or under flow. There are many other ways of constructing the fitness landscape. One needs experimentation to ascertain the best form of f . In our case, the present form has been found to be adequate. The corresponding eigenvalue E_0 is simply the value of the Rayleigh quotient for C_i :

$$\mathbf{r}_0 = \frac{C_i H C_i}{C_i^\dagger C_i} = E_0. \quad (7)$$

The evolution on the fitness landscape, would not however take place automatically. It is to be brought about by the action of a well-defined set of genetic operators. The first among these operators to act upon the initial population is the selection operator. A roulette wheel with slot widths proportional to fitness values is spun n times, n being the number of strings in the population. The string S_i is selected if the pointer points to the i th slot. It is expected that the selection operator allows us to copy a larger number of strings with higher fitness onto the mating pool, since higher fitness means wider slots on the wheel. Once the mating pool with n strings is created by the selection process, the crossover operation is allowed to take place with a preset probability p_c . Suppose, a pair of strings (S_k, S_l) have been randomly chosen to undergo crossover at the p th site with a probability p_c . The crossover creates a pair of new strings (S'_k, S'_l) where,

$$S'_k \equiv (c_{1k}, c_{2k}, \cdots c'_{pk} c_{p+1,k} \cdots c_{nk}),$$

$$S'_l \equiv (c_{1l}, c_{2l}, \dots, c'_{pl}c_{p+1,l}, \dots, c_{nl}), \quad (8)$$

and

$$\begin{aligned} c'_{pk} &= fc_{pk} + (1-f)c_{pl}, \\ c'_{pl} &= (1-f)c_{pk} + fc_{pl}, \end{aligned} \quad (9)$$

with $f(0 < f < 1)$ being drawn from a set of randomly distributed values. f thus plays the role of a mixing parameter, that creates new information. The crossover operator is allowed to act on a certain fraction of the strings (70–75%), the remaining ones being allowed to remain unchanged. Once the crossover operation is complete, the strings are made to undergo a low intensity and low probability event called mutation. In our case, we define the mutation operation in the following way. Let the q th site on the k th string be chosen to undergo mutation with probability p_m . The amplitude to suffer mutation is then c'_{qk} which is randomly mutated to c''_{qk} , where

$$c''_{qk} = c'_{qk} + (-1)^L r \Delta, \quad (10)$$

L being a random integer, r a random number ($0 \leq r \leq 1$) and Δ the intensity of mutation (10^{-2} – 10^{-3}). Every string in the post-crossover operation is allowed to mutate with a probability p_m . The conclusion of one sequence of selection-crossover-mutation operations signals that one generation has elapsed. We enforce now the most strongly elitist criterion of selecting n -strings out of a population of n -parents and n -offspring. Accordingly, the best n among the set of n parent strings and n offspring strings are chosen for the selection operation and the entire sequence of operation is repeated, till the average fitness of the population has converged. The wavefunction string with the highest fitness at this stage is likely to be the ground eigenvector of the problem. Let us note that one of the consequences of the strongly elitist strategy is that the highest fitness value available in the population at any stage continuously rises up to a maximum of 1 – it never oscillates. It may, however, force premature convergence in some cases. The scheme outlined concerns only the ground eigenvalue. For extracting excited eigenvalues the scheme has to be modified. Let us note that the Rayleigh quotient has a minimum at the ground vector (c_0) corresponding to the lowest eigenvalue and a saddle point for every other eigenvector. Extraction of excited eigenvalues and associated eigenvectors therefore requires a saddle-point search which

is much more difficult. Instead, we may still use the search for a minimum of $\{(\nabla \mathbf{r}_i)^\dagger \nabla \mathbf{r}_i\}$ but use a projected Hamiltonian H_p for extracting the higher eigenvalues.

Let $|\mathbf{y}_0\rangle$ be the ground eigenvector of the problem and $P_0 = |\mathbf{y}_0\rangle\langle\mathbf{y}_0|$ be the corresponding projector. Let the eigenvalues of H satisfy the following inequalities ($E_i \leq 0, i = 0, 1, \dots$) and

$$E_0 < E_1 < E_2 \cdots E_{k-1} < E_k < \cdots < E_n.$$

The projected Hamiltonian H_p , where

$$H_p = (1 - P_0)^\dagger H (1 - P_0),$$

has clearly the following eigenvalue structure: $E_1 < E_2 \cdots E_n < E_0$, as $\hat{H}_p |\mathbf{y}_0\rangle = 0$. The lowest eigenvalue of H_p is clearly E_1 and the procedure outlined above for extracting the lowest eigenvalue of H can be applied to extract E_1 and $|\mathbf{y}_1\rangle$, by working with H_p in place of H . The procedure can be applied sequentially, to extract $|\mathbf{y}_2\rangle, |\mathbf{y}_3\rangle, \dots$ etc. For the $(k+1)$ th eigenstate, H_p would be then given by

$$H_p = (1 - P)^\dagger H_0 (1 - P), \quad (11)$$

where

$$P = \sum_{i=0}^{k-1} P_i = \sum_{i=0}^{k-1} |\mathbf{y}_i\rangle\langle\mathbf{y}_i|.$$

If all the eigenvalues of H are positive ($E_i > 0$), a simple recipe that works would be to replace H by H_s^k , where

$$H_s^k = H - \mathbf{I} \sum_{i=0}^{k-1} (1 - P_i). \quad (12)$$

H_s obviously has all but the first k eigenvalues shifted by $-\mathbf{I}$ so that the eigenvalue structure of H_s is (with properly chosen $-\mathbf{I}$),

$$E_{k+1} < E_{k+2} \cdots E_n < E_0 < E_1 \cdots.$$

We may also optimize the basis set used simultaneously by merely extending the definition of the strings so as to allow each string to carry information concerning basis parameters as well as the linear expansions coefficients. The generalized solution string now would have the following structure,

$$S_k = (c_{0k}, c_{1k}, \dots, c_{mk} | \mathbf{a}_k, \mathbf{b}_k, \mathbf{g}_k \cdots),$$

where the vertical bar separates the linear parameters c_{ik} s from the nonlinear ones. The rest of the procedure remains unaltered. In what follows, we give results that indicate the workability of the GA-based method both with and without basis optimization.

3. Results and discussion

We present here two sets (A and B) of results and applications of the GA-driven diagonalization based on minimization of Rayleigh quotient. The first set makes use of a *fixed basis* while the second set refers to diagonalization with simultaneous optimization of the basis set. The applications are not spectacular – they only demonstrate workability of the method proposed and its advantages when coupled with the basis optimization problem.

3.1 (Set A) Ground and excited eigenvalues and vectors of a Harmonic oscillator – diagonalization in a fixed basis

We have used the Fourier Grid Hamiltonian method (FGH) for constructing the H -matrix in a coordinate representation, using 101 grid points over a grid length of 21 a.u. The population contains 20 individual strings ($n = 20$), and a crossover probability $p_c = 0.75$ and mutation probability $p_m = 0.5$ are used uniformly for all generations. The initial population was created randomly from a set of nodeless ampli-

tude distributions and the best string in the initial population had a fitness $f \approx 0.1$.

Figure 1a shows the profile of evolution of the fitness of the best string during a typical GA run. One can see the steep rise in the fitness value from 0.1 to 0.98 in the first 5000 generations. Figure 1b shows how the energy eigenvalue corresponding to the same string evolved during the generations. The rapid rise in fitness and fall in energy during the first 5000 generations and the subsequent slowdown of the rate of improvement is typical of a GA-run. The initial phase is dominated by crossover when gross changes take place in the strings. The evolution is mutation-dominated in the final stages.

After obtaining the ground eigenvector, $|\mathbf{y}_0\rangle$ and hence p_0 , we construct the shifted Hamiltonian of (12) with a shift $I = 3.0$ a.u. of energy.

$$\begin{aligned} H_s &= H - 3.0 (1 - |\mathbf{y}_0\rangle\langle\mathbf{y}_0|) \\ &= H - 3.0 (1 - \hat{P}_0). \end{aligned}$$

The initial population is made up of wavefunction strings with one node, each being made orthogonal to that already computed $|\mathbf{y}_0\rangle$ by Schmidt orthogonalization. The evolution profiles for the fitness and energy are displayed in figures 2a and b respectively. A comparison with corresponding profiles for the ground state reveals that the search is relatively slow – presumably because of the need to satisfy two mutually competing constraints – the need to

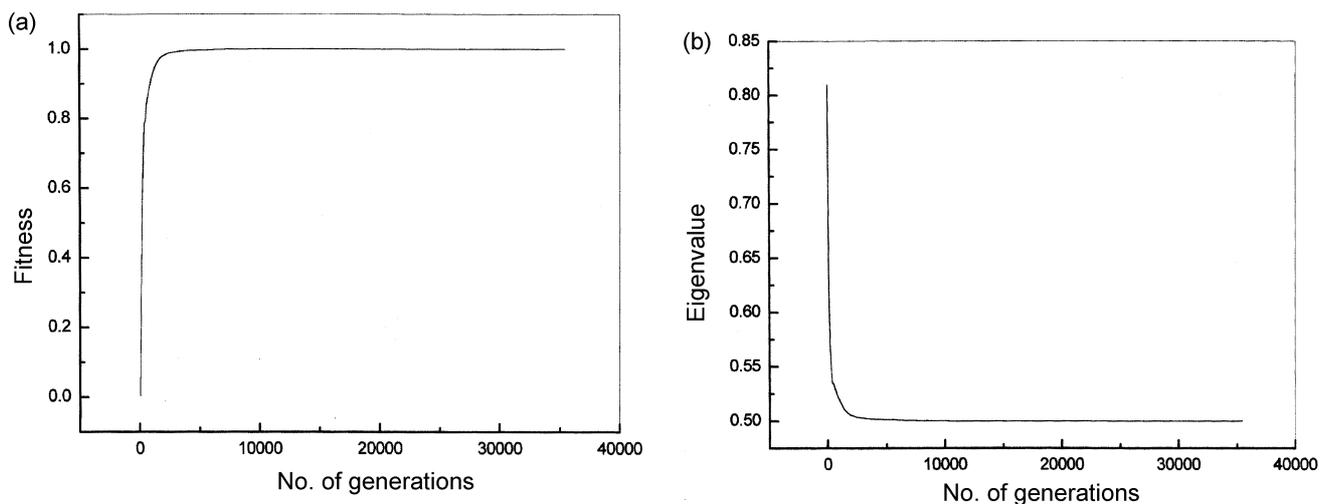


Figure 1. (a) Fitness evolution of the best string during the calculation of the lowest eigenvalue of a Harmonic Oscillator in an 101 point Fourier Grid Hamiltonian based calculation. (b) Evolution of the lowest eigenvalue of the same Harmonic Oscillator as in 1(a).

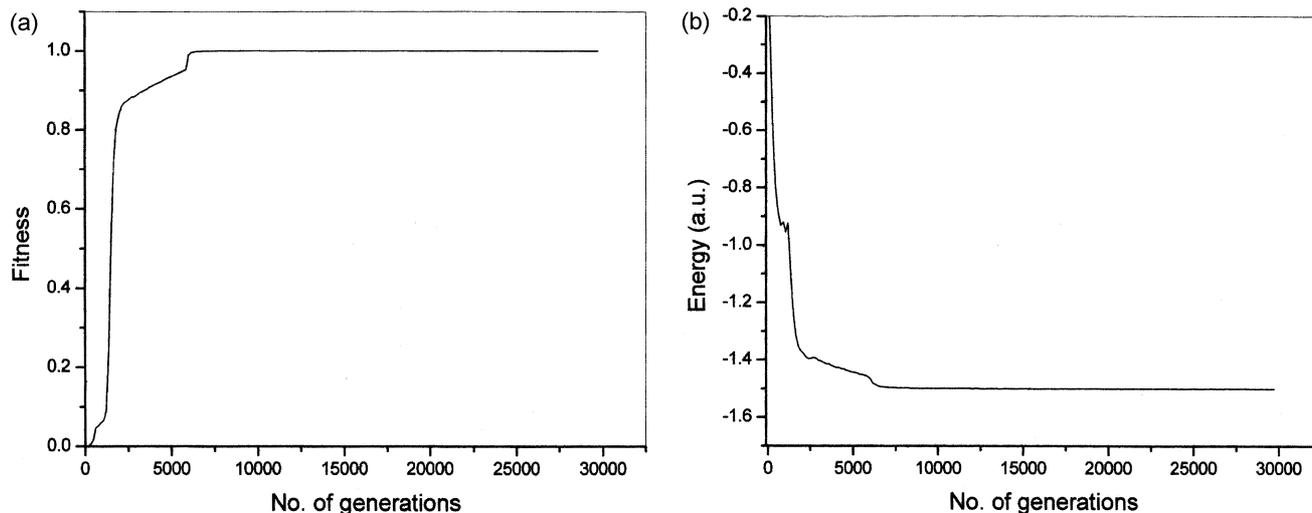


Figure 2. (a) Fitness evolution of the best string while calculating the first excited state of a Harmonic oscillator during a GA-RQ run. (b) Evolution of energy for the excited state (shifted by -3.0 a.u.) of the same Harmonic oscillator.

minimize energy as well as the cost of computation. We note here that the lowest energy reached with H_s is -1.5 a.u. (figure 2b). In fact, this is the shifted value $E'_1 = (E_1 - I)$ of E_1 where $I = 3$. If the shift value is subtracted from E'_1 , we get the true first excited eigenvalue. The fitness evolution profiles suggest that the present method would ultimately turn out to be slower compared to the methods of Davidson⁵ or Lanczos,⁴ although the initial rise of fitness is remarkable. The growth of fitness slows down towards the end when fine-tuning of the eigenvector takes place. We note, however, that even if it is 10 times slower, straightforward parallelization of the string evaluation step would easily make up for this slowness. One can further cut down the cost of computation by stopping the GA-based search as soon as the fitness value has increased to 0.8 – 0.9 and use an inverse iteration procedure² that can shorten the plateau region of the fitness evolution profile drastically. Both the option are now being evaluated.

3.2 (Set B) Ground state of H^- by diagonalization along with basis optimization

The trial wavefunction strings $\mathbf{y}_k(r_1, r_2)$ for the ground state $H^{(-)}$ has been chosen in the following form:

$$\mathbf{y}_k(r_1, r_2) = \sum_{ij} c_{ij}^{(k)} \mathbf{f}_{ij}^{(k)}(r_1, r_2, n_i, n_j, l_i = l_j, \mathbf{a}_i^k, \mathbf{a}_j) \Phi_{\text{spin}}(1, 2),$$

where

$$\mathbf{f}_{ij}^{(k)} = N[\mathbf{c}(r_1, \mathbf{a}_{n_i}^{(k)}) \mathbf{c}(r_2, \mathbf{a}_{n_j}^{(k)}) + \mathbf{c}(r_1, \mathbf{a}_{n_k}^{(k)}) \mathbf{c}(r_2, \mathbf{a}_{n_i}^{(k)})],$$

is $\mathbf{f}_{\text{spin}}^{(1,2)}$ singlet ($s = 0$) spin eigenfunction for the two electrons. We have used only six distinct n_i values ($n_i = 0, 2, 4, 6, 8, 10$). Each string now consists of M amplitudes $c_{ij}^{(k)}$ and m nonlinear parameters (\mathbf{a}_{ni}), all of them being allowed to evolve simultaneously. $\mathbf{c}(\mathbf{a}_n)s$ are actual eigenfunctions of the appropriate 3-D oscillator. We have used $l_i = l_j = 0$ for this calculation. The limiting energy is thus expected to be the s -limit energy.

Figures 3a and b exhibit the profiles of fitness and the corresponding energy eigenvalue of the best evolving string in the population representing the ground state of the $H^{(-)}$ ion. Diagonalization with simultaneous basis optimization proceeds smoothly – the initial improvement being rather steep. The fine adjustment takes place in the at regions of the profiles. Since the nonlinear parameters of the basis functions undergo evolution due to crossover and mutation, the Hamiltonian matrix needs to be recalculated whenever the \mathbf{a}_{ni} 's undergo changes. The advantage here is that we need not completely diagonalize the corresponding H whenever it is reconstructed. The search treats all the parameters on an equal footing and the genetic algorithm uses the information present in the search space to drive the evolution to the desired goal.

It appears that extracting a few eigenvalue of a real symmetric Hamiltonian matrix of moderate

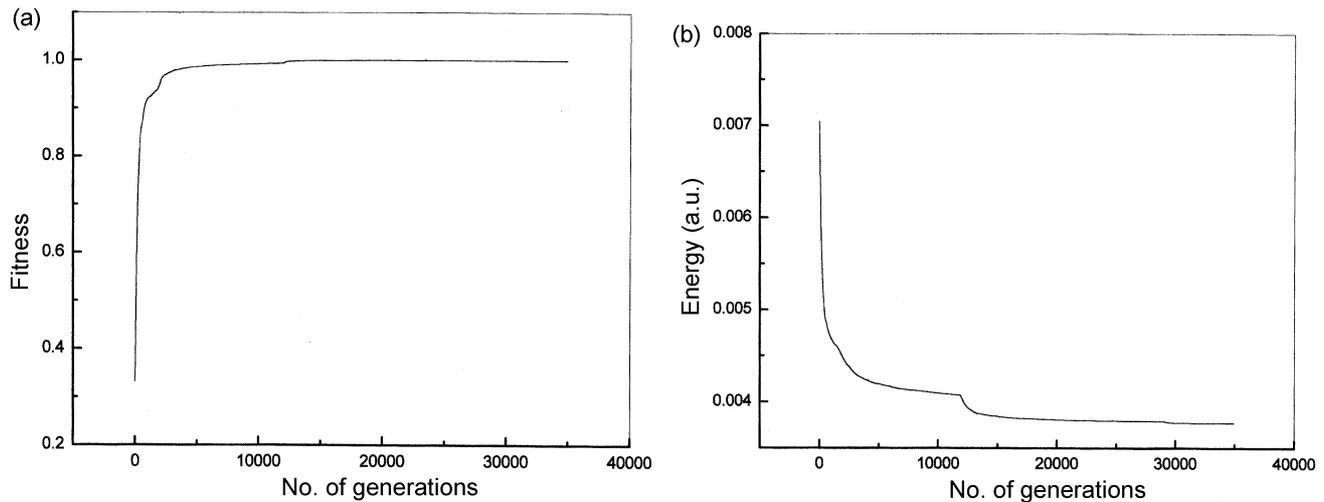


Figure 3. (a) Fitness evolution of the best string during the calculation of the lowest eigenvalue of a Morse oscillator (101 point FGH calculation). (b) Evolution of the lowest eigenvalue of the same Morse oscillator.

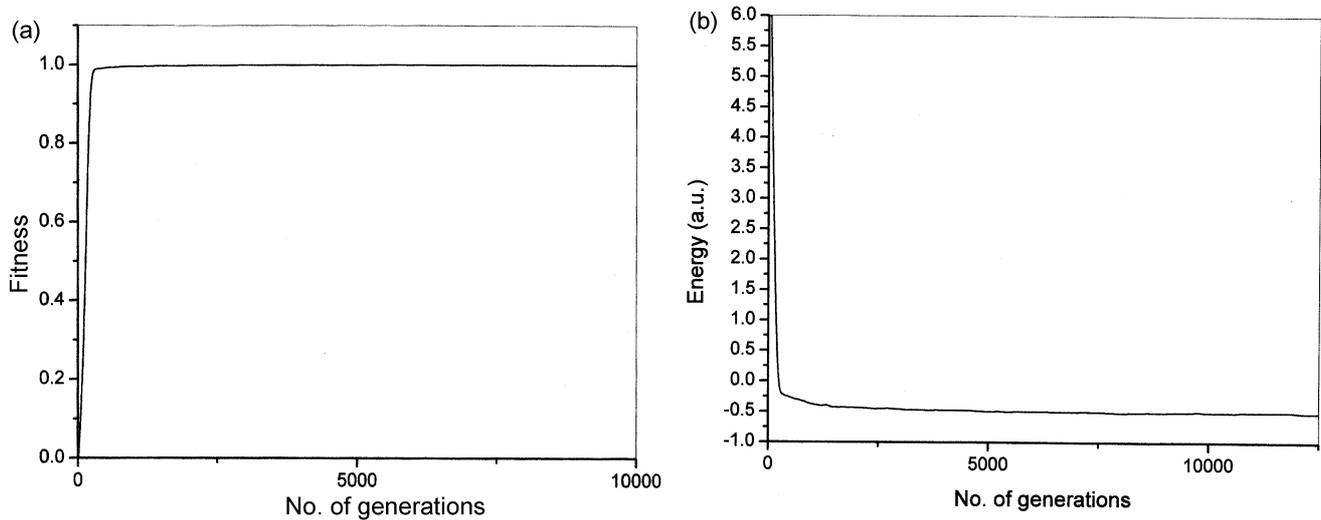


Figure 4. (a) Fitness profile for the $H^{(-)}$ - ground state where diagonalization is carried out with simultaneous optimization of basis parameters (3D-harmonic oscillator basis is used). (b) Evolution of the energy for the same $H^{(-)}$ problem during the GA-RQ run.

sizes by a GA-based minimization of Rayleigh quotient is feasible along with simultaneous basis optimization. To become competitive with the standard recipes now available, the proposed algorithm needs to be parallelized as completely as possible and coupled to an inverse iteration procedure. For the ‘few eigenvalues’ problem, one can effectively make use of the partitioning¹⁰⁻¹² method according to which the desired ‘ a ’ eigenvalues can be found from the partitioned equation,

$$\{H_{aa} + H_{ab}(E.I_b - H_{bb})^{-1}H_{ba}\}|\mathbf{y}_a\rangle = E|\mathbf{y}_a\rangle,$$

$$H_{\text{eff}}(E)|\mathbf{y}_a\rangle = E|\mathbf{y}_a\rangle,$$

where

$$|\mathbf{y}\rangle = (|\mathbf{y}_a\rangle|\mathbf{y}_b\rangle),$$

$$\langle\mathbf{y}_a|H|\mathbf{y}_a\rangle = H_{aa}, \quad \langle\mathbf{y}_a|H|\mathbf{y}_b\rangle = H_{ab},$$

$$\langle\mathbf{y}_b|H|\mathbf{y}_a\rangle = H_{ba}, \quad \langle\mathbf{y}_b|H|\mathbf{y}_b\rangle = H_{bb}.$$

Since $H_{\text{eff}}(E)$ acts on the a -dimensional subspace ($a \ll a + b$), it can reduce the computational labour. However, $H_{\text{eff}}(E)$ depends on the unknown E , making it a nonlinear problem. The genetic algorithm can handle such nonlinear optimization problem rather

well and our experience so far confirms the expectation.

4. Conclusion

GA-based minimization of the Rayleigh quotient leads to a workable algorithm for extracting a few eigenvalues and vectors of a real symmetric matrix. The method is specially useful when a simultaneous search for optimal basis is carried out. For realizing the full power of the algorithm, the string evaluation step needs to be parallelized. It appears that the partitioning method together with parallelization could make the GA-based search for a few eigenvalues a viable alternative to the presently available methods.

Acknowledgement

We thank the Council of Scientific and Industrial Research, New Delhi for financial support through a research grant.

References

1. Parlett B N 1981 *The symmetric eigenvalue problem* (Englewood Cliffs, NJ: Prentice Hall)
2. Wilkinson J H 1965 *The algebraic eigenvalue problem* (Oxford: Clarendon)
3. Kerner W J 1989 *J. Comput. Phys.* **85** 1
4. Cullum J K and Willoughby R A 1985 *Lanczos algorithm for large symmetric eigenvalue computations* (Boston: Birkhauss)
5. Davidson E R 1975 *J. Comput. Phys.* **17** 87
6. Fletcher R 1989 *Practical methods of optimization* 2nd edn (Chichester: John-Wiley)
7. Holland J H 1975 *Adaptation in natural and artificial systems* (Ann Arbor, MI: University of Michigan Press)
8. Goldberg D E 1989 *Genetic algorithms in search, optimisation and machine learning* (Reading, MA: Addison Wesley)
9. Nandy S, Chaudhury P and Bhattacharyya S P 2002 *Int. J. Quantum Chem.* **90** 188
10. Lowdin P O 1951 *J. Chem. Phys.* **19** 1396
11. Lowdin P O 1968 *Int. J. Quantum Chem.* **2** 167
12. Killingbeck J P and Jolicard G 2003 *J. Phys.* **A36** R105