On the use of local and global minimizers for optimization problems in quantum chemistry

P. Dutta and S. P. Bhattacharyya

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

Available local and global minimization algorithms are reviewed with special emphasis on their applicability in electronic structure calculation of atoms and molecules in the variational approximation which can often be cast as a constrained search for the global minimum on a multidimensional energy or energy-like surface in the space spanned by the parameters in the wave function. The traditional gradient-driven algorithms cannot guarantee the globality of the minimum reached. A global stochastic minimizer like the one offered by the method of simulated annealing is shown to be a powerful tool for handling these problems, in general. Specific applications are discussed to uncover the merits and demerits, if any, of the method proposed.

THE aim of this article is twofold; the first, has been to survey the available practical methods for minimizing a function or a functional of many variables and the second, to assess them in the context of quantum chemical optimization problems, specially problems arising in electronic structure theory. The very edifice of electronic structure theory has been largely built on the idea of expanding wave functions in a finite dimensional analytic basis set. Naturally, the task of finding an optimally conditioned basis set or the optimal form of such expansions has seen very major efforts in the arena of atomic and molecular electronic structure theory in general, and molecular electronic structure theory in particular[1]. A very natural choice of tool in this research has been the variational approximation. The connection between optimization theory and principles of variational electronic structure theory is a direct one, for the variational recipe casts the problem of finding an optimal expansion for ψ in the mould of an extremization problem of an appropriately constrained energy functional[2,3]. In the most general case, the extremization condition on the appropriate functional naturally leads to equations which are nonlinear and therefore offer multiple solutions physically representing maxima, minima or the saddle points of the function. How does one pick up the lowest energy solution from amongst all possible solutions? In what follows, we intend to carry out a general analysis of the problems involved and ultimately offer a workable strategy for accomplishing

the task. The strategy, as we will see later, involves the use of a stochastic global minimizer[4-10].

Variational principle and energy extremization

For the ground state of a system described by a Hamiltonian (H) bounded from below, the energy functional $E(\tilde{\psi}) = (\langle \tilde{\psi} | H | \tilde{\psi} \rangle / \langle \tilde{\psi} | \tilde{\psi} \rangle)$ attains a global minimum for the exact ground state wave function ($\tilde{\psi} = \psi_0$) while for all other admissible trial wave functions, $E(\tilde{\psi}) > E_0$. From the variational point of view, the problem of finding the ground state wave function thus boils down to a global minimization problem of energy. For the excited states, the problem is much less simple. The functional $E(\tilde{\psi})$ can be shown to have a stationary point $E(\tilde{\psi}')$ which is an upperbound to the corresponding exact excited state energy[1] (E_n) if and only if $\tilde{\psi}$ is constrained to remain orthogonal to all the exact lower eigenstates of H. This assertion of variational principle is operationally crippled from the point of view of optimization theory since the constraint conditions involve exact lower states which will be unknown for problems of real interest. For practical calculations on excited states one is therefore compelled to invoke a much more restricted variational recipe provided by the so-called linear variational principle succinctly put forward by the Hylleraas Undheim Mcdonald (HUM) theorem[11,12]. The only restriction in HUM is the linearity of the trial space. But it prevents variational collapse and offers systematic improvement of the wave function and energy (Figure 1).

The restriction of linear space is a severe one and is often manifested in slow convergence of the expansion

$$\psi^{(N)} = \sum_{i=1}^{N} c_i \phi_i .$$

It is possible to relax this restriction but only at the cost of bringing in additional (difficult to enforce!) constraints. Thus, for a general variational calculation on the kth excited state of a system, $E(\tilde{\psi}_k)_{optimum}$ is an upperbound to the exact energy of the kth state (E_k) if and only if $\tilde{\psi}_k$ is constrained to remain both orthogonal
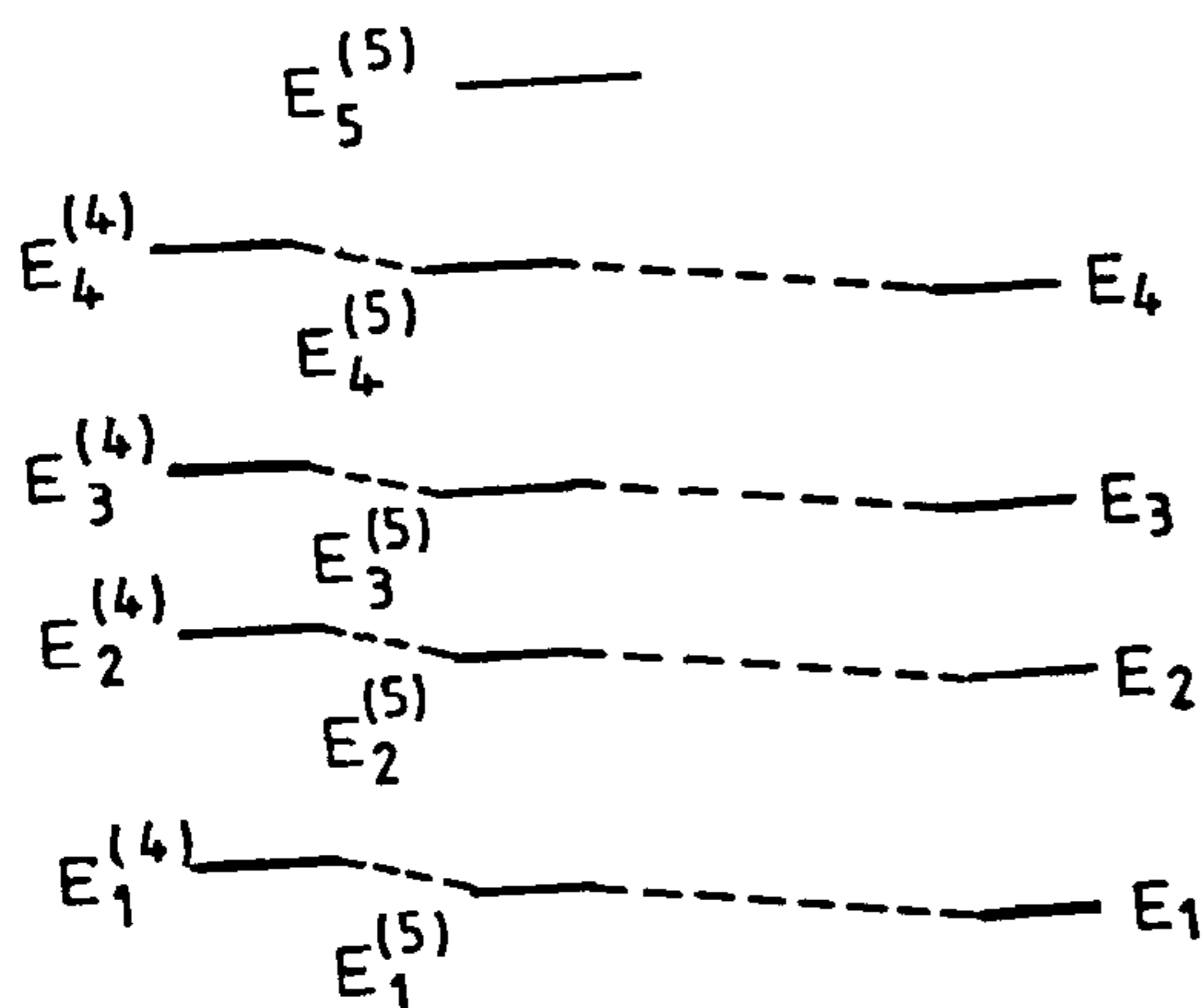
Figure 1. Monotonic convergence of approximate energy eigenvalues obtained from linear variation principle.

to and decoupled with the $(k-1)$ lower approximate eigenstates of $H(\tilde{\psi}_1 \tilde{\psi}_2, \ldots \tilde{\psi}_{k-1})$ which themselves must also be mutually orthogonal and decoupled[3]. The equations resulting from the extremization of the appropriate functional $[J(\tilde{\psi}_k)]$ are, however, highly non-linear and can be solved only by some practicable strategy of unconstrained global minimization. It is natural therefore to start exploring the possibility of casting the extremization problem of the energy functional $J(\tilde{\psi})$ into a global minimization problem of some other functional. The transcription may be effected by using the idea of penalized function(al) strategy of unconstrained optimization[13–17].

From extremization to minimization

Let us suppose that our problem is to perform a general variational calculation on the kth excited state of a system. Let $\tilde{\phi}_1, \tilde{\phi}_2, \ldots \tilde{\phi}_{k-1}$ be a set of orthonormal and mutually decoupled approximate eigenfunctions of H. The correct energy functional for the kth state is then

$$J(\tilde{\psi}_k) = \langle \tilde{\psi}_k \mid H \mid \tilde{\psi}_k \rangle / \langle \tilde{\psi}_k \mid \tilde{\psi}_k \rangle +$$

$$\sum_{i=1}^{k-1} \alpha_i \langle \tilde{\psi}_k \mid \tilde{\phi}_i \rangle + \sum_{j=1}^{k-1} \beta_j \langle \tilde{\psi}_k \mid H \mid \tilde{\phi}_j \rangle, \quad (1)$$

where $\alpha_i s$ are a set of Lagrangian multipliers (LM) for enforcing the relevant orthogonality constraints ($\langle \tilde{\phi}_k \mid \tilde{\phi}_i \rangle = 0$) while $\beta_i s$ stand for another set of LMs for satisfying the decoupling constraints ($\langle \tilde{\psi}_k \mid H \mid \tilde{\phi}_i \rangle = 0$). The extremum solution is obtained by setting the first derivative of $J(\tilde{\psi}_k)$ (with respect to $\tilde{\psi}_k$, $\alpha_i s$ and $\beta_i s$) zero.

The same problem cast as an absolute minimization problem with the aid of the penalized function(al) strategy[15,16] works with a different functional $\mathcal{F}(\tilde{\psi}_k)$:

$$\mathcal{F}(\tilde{\psi}_k) = (\langle \tilde{\psi}_k \mid H \mid \tilde{\psi}_k \rangle - E_L)^2 +$$

$$\sum_{i=1}^{k-1} \lambda_i \, g_i(\langle \tilde{\psi}_k \mid \tilde{\phi}_i \rangle) +$$

$$\sum_{j=1}^{k-1} \gamma_j \, h_j(\langle \tilde{\psi}_k \mid H \mid \tilde{\phi}_j \rangle) + \eta f(\langle \tilde{\psi}_k \mid \tilde{\psi}_k \rangle) \quad (2)$$

E_L in equation (2) is a lower bound to the constrained energy of the kth state,

$$g_i(\langle \tilde{\psi}_k \mid \tilde{\phi}_i \rangle) = |\langle \tilde{\psi}_k \mid \tilde{\phi}_i \rangle|^2, \quad h_j(\langle \tilde{\psi}_k \mid H \mid \tilde{\phi}_j \rangle)$$

$$= |\langle \tilde{\psi}_k \mid H \mid \tilde{\phi}_j \rangle|^2 \quad \text{and} \quad f(\langle \tilde{\psi}_k \mid \tilde{\psi}_k \rangle)$$

$$= (1 - \langle \tilde{\psi}_k \mid \tilde{\psi}_k \rangle)^2.$$

η, $\lambda_i s$ and $\gamma_i s$ are penalty weight factors of appropriate dimensions and magnitudes. It would perhaps not be out of place to explain the precise mathematical nature of the penalty function and weight factors introduced for constructing the functional $\mathcal{F}(\tilde{\psi}_k)$. Let us suppose that $F(x)$ is the function to be minimized subject to the constraints $f_i(x) \leq 0$, $i = 1, 2, \ldots, n$. To proceed further we introduce the following definitions:

$$g_0(t) = \begin{pmatrix} t^2, & t \geq 0 \\ 0, & t < 0 \end{pmatrix} \quad (3)$$

and define a new function $h(x, \beta)$, the penalty function, as follows:

$$h(x, \beta) = \beta \sum_{i=1}^{n} g_0[f_i(x)] \quad (4)$$

It is clear that $h(x, \beta) = 0$, if $x \subseteq D$, where $D = \{ x: f_i(x) \leq 0, i = 1, 2, \ldots, n \}$. If $x \notin D$, $h(x, \beta) > 0$. Also, $h(x, \beta) \rightarrow +\alpha$ as $\beta \rightarrow \alpha$ when $x \notin D$. β is called a penalty weight factor. It is now possible to define an auxiliary minimization problem with the help of $h(x, \beta)$ such that the constrained minimization problem for $F(x)$ reduces to the unconstrained minimization of an auxiliary function $\mathcal{F}(x, \beta)$ where

$$\mathcal{F}(x, \beta) = F(x) + h(x, \beta). \quad (5)$$

It is obvious that the construction of $h(x, \beta)$ as suggested here is not unique. Many alternative constructions are possible for $h(x, \beta)$ and the auxiliary function $\mathcal{F}(x, \beta)$. The specific one we have used is called Morrison function which is specially suitable for handling equality

constraints that we need to enforce. An inspection of the Morrison function $\mathcal{F}(\tilde{\psi}_k)$ shows that if any constraint condition is violated the value of $\mathcal{F}(\tilde{\psi}_k)$ rises steeply. The absolute minimum of $\mathcal{F}(\tilde{\psi}_k)$ is reached only when all the constraint conditions are unequivocally satisfied and $\tilde{E}_k$ is as close as possible to E_L. If the exact lower eigenstates of H were known, the third set of terms could be dropped from $\mathcal{F}(\tilde{\psi}_k)$. If the trial space were linear, again the same set of terms could be eliminated. For a ground state calculation, the only constraint on $\tilde{\psi}$ is the normalization constraint and both the second and third set of terms in $\mathcal{F}(\tilde{\psi}_k)$ naturally drop out. The lower bound E_L in $\mathcal{F}(\tilde{\psi}_k)$ is essentially arbitrary and a sufficiently large positive number would be acceptable. If $E_L < \langle \tilde{\psi}_k | H | \tilde{\psi}_k \rangle$ at any stage, one can readjust its value by using $E'_L = E_L - \text{const.} \{ \mathcal{F}(\tilde{\psi}_k) \}^{1/2}$. For the first excited state calculation, the optimized energy of the ground state (E_0) is the best choice for E_L. Similarly, the optimal energy of the first excited state can be used as E'_L in a calculation on the second excited state and so on sequentially to higher excited states. Whatever may be the case, one can either set $\delta F(\tilde{\psi}_k) = 0$ and proceed to solve the resulting equations or chalk out a strategy that directly looks for the minima of $\mathcal{F}(\tilde{\psi}_k)$. In either case, one needs a suitable optimization method.

The optimization methods

A number of methods for unconstrained minimization of nonlinear functions of many variables[17] are available. The choice of the best method for a particular situation depends on a number of factors, namely, the nature of the function to be minimized, the number of variables, the availability of analytical first (and higher) derivatives and the cost of evaluating the function and its derivatives, and so on. On the other hand, efficacy of an optimization procedure is usually calibrated in terms of a few important features of the algorithm, e.g. speed of convergence, storage requirement, stability and reliability of the optimization method and the overall cost of optimization. The algorithms are generally grouped into two categories depending on the type of information they use: (a) non-derivative methods and (b) derivative methods.

Non-derivative minimizers

Experience with non-derivative methods of optimization has been long. The features of a non-derivative method that account for its attraction and applicability are: conceptual simplicity, easy programmability and small storage demands on the computer. Of the non-derivative methods, the multivariate grid search method is the

oldest one. It has a long history in quantum chemistry, particularly in optimizing the energy with respect to nuclear positions and orbital exponents. But the sequential univariate search was perhaps the most frequently used method for molecular geometry optimization. Figure 2 shows how a sequential univariate search on a quadratic two parameter surface proceeds. A more sophisticated pattern search method due to Hooke and Jeeves[18] has also found applications in quantum mechanical minimization problems. Methods strongly related to the pattern search method are the Simplex method[19] and the method of Rosenbrock[20] which, however, have found little use in quantum chemistry. The conjugate directions method is potentially the most powerful among all the non-derivative methods and a particular implementation of the scheme that has become the most popular among the non-derivative minimizers has been due to Powell[21]. It generates conjugate directions without calculating first or second derivatives.

Derivative-based minimizers

The derivative methods[22–30] are based on a sequence of one-dimensional direction searches[22,23] with a few exceptions like the memory[24] and super memory[25] gradient methods which are not univariate in nature. In general, these algorithms approximate the function-surface at the kth step by a quadratic expression in terms of variables x_k, the computed function f_k, the computed gradient g_k, and the approximate or exact Hessian B_k:

$$f(x) = f_k + g_k^T (x - x_k) +$$

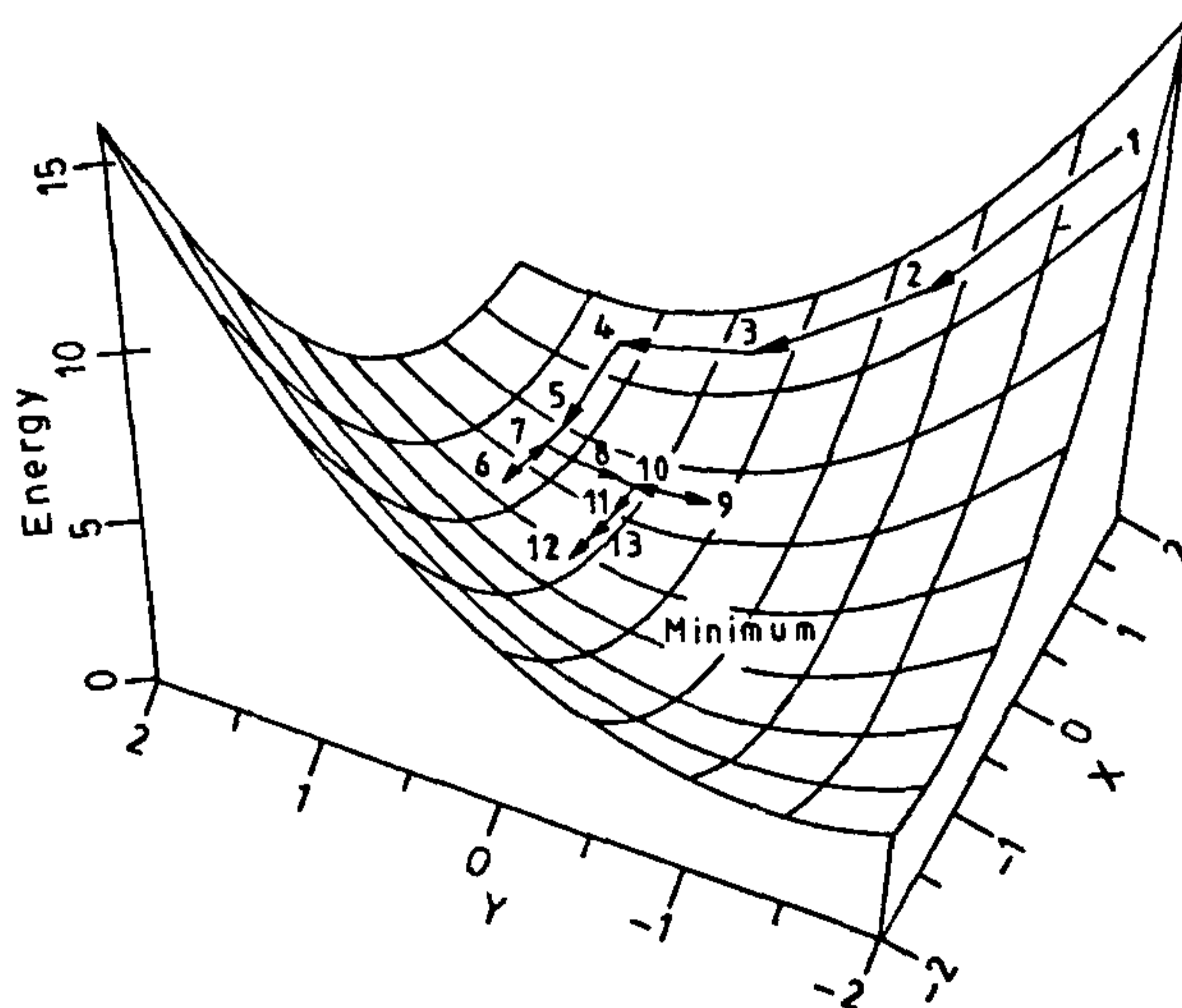$$(1/2) (x - x_k)^T B_k (x - x_k). \tag{6}$$

Figure 2. Steps in finding the minimum on a quadratic surface using the sequential univariate search or axial iteration method (ref. *Adv. Chem Phys.*, 1987, vol. LXVII, courtesy Academic Press).

Depending on the nature of derivative information (g_k only or both g_k and B_k), the line search scheme, formulae for updating B_k or the inverse of B_k ($= H_k$), there is a broad spectrum of the derivative methods which include, for example, the variable metric[26], Fletcher–Reeves[26], Fletcher–Powell[22], Davidon–Fletcher–Powell[27] (DFP), Murtagh–Sargent[28] (MS), Broyden–Fletcher–Goldfarb–Shanno[29] and the optimally conditioned algorithm[30]. The spectrum starts with the methods which employ only the first derivative of the function to be minimized and are called gradient methods. All these methods update variables by taking $x_{k+1} = x_k + \alpha_k B_k^{-1} p_k$, p_k being the direction of univariate search. Figure 3 shows how a Fletcher–Powell type algorithm operates its search for a minimum on a quadratic two-dimensional surface. The method at the other end of the spectrum is Newton's method which requires the full Hessian matrix. The most frequently used derivative algorithms fall between the extremes of the fixed metric methods and Newton's method and are called variable metric or quasi-Newton methods. These methods start with an approximate Hessian and using the gradient information improve it during the course of the optimization, instead of attempting direct calculation of the Hessian which is costly in terms of computing time.

The simplest of all the gradient methods is the steepest descent (SD) method. In the absence of any information about the function $f(x)$ to be minimized the optimum direction to move from a point $x = a$ to minimize $f(x)$ is to move along the direction of steepest descent $p = -\lambda g(a)$, where $g(a) = \partial f / \partial x$ at $x = a$. The method is, however, very slow. The reason for this rather disappointing performance of the SD method lies in

that each SD vector is orthogonal to the SD vector of the previous step. If the initial SD vector is not directed at right angles to the axis of the valley, successive SD vectors will be directed across the valley and not along it. The conjugate gradient technique (CG) overcomes this by combining information about the function and its gradient at a few previous steps such that each minimization step is independent of another. The initial direction is taken to be the negative of the gradient of the starting point. A subsequent conjugate direction is then constructed from a linear combination of the new gradient and the previous direction that minimized $f(x)$.

The CG, MS and DFP methods are special cases of the family of algorithms due to Huang[31] and Dixon[32], which suggest that the direction vector and the Hessian must obey the following relation

$$H_i B p_j = \rho p_j, \quad i - 1 \geq j \geq 0$$

to produce conjugate directions, where ρ is an arbitrary scalar, so that the limit $H_r = \rho B^{-1}$. As a general rule the overall computational cost and convergence of these methods depend on the accuracy of the line searches and the cost of gradient calculation. The line searches can be improved by means of fitting followed by interpolation at the minimum (by cubic polynomial interpolation[33], quartic polynomial[34], conic interpolation[35] or by following what has been suggested by McKelvey and Hamilton[36]). A few algorithms which do not fall into the categories outlined above have also been proposed. For example, the algorithm due to Csaszar and Pulay[37] uses a fixed Hessian with a k-dimensional search at the kth iteration and the direction vector (p_k) is determined in a least squares sense. It performs well. Finally, we come to those algorithms which depend on a knowledge of the second derivative of the function, i.e. the Hessian (B) or its inverse ($H = B^{-1}$). If B or H^{-1} are available, the minimum of a quadratic surface can be found in one step, so that we need not bother about the direction of descent. However, for non-quadratic functions, the choice of optimal directions again becomes important. For a well-behaved minimization, the Hessian must be positive definite and not ill-conditioned. In the latter case, however, the problem is remedied by means of a 'shifted' or 'dressed' Hessian obtained by adding a constant to the odd-eigenvalue(s). The most widely used scheme under this category is due to Marquardt–Levenberg[38]. An initial estimate of the Hessian is required for starting the quasi-Newton methods. A number of choices with different levels of cost and accuracy are available in the context of quantum chemistry. Thus one can use a unit matrix, make an empirical guess, use Hessian computed by a semi-empirical MO theory, construct an approximate Hessian
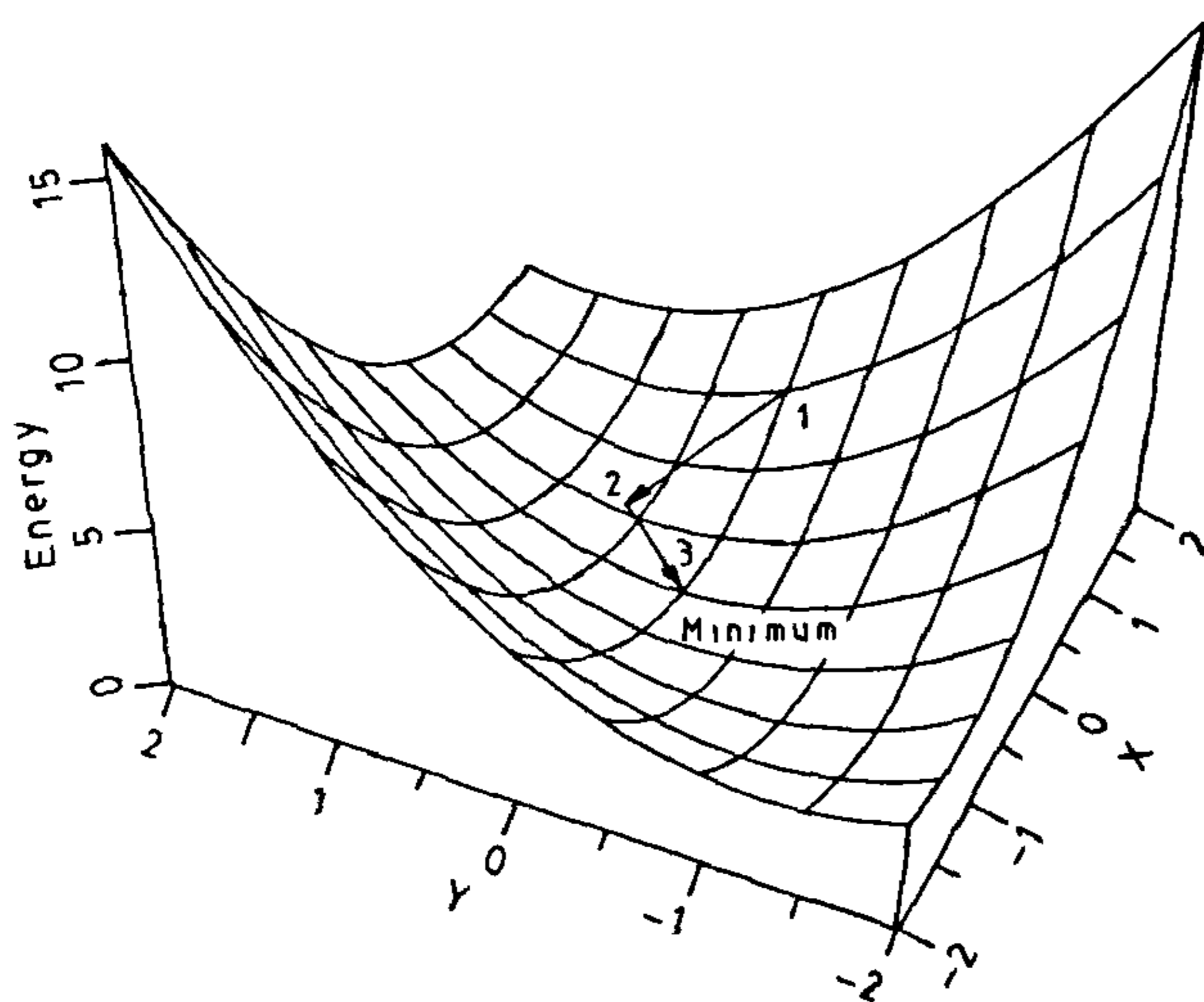

Figure 3. Steps in finding the minimum on a quadratic surface using a quasi-Newton algorithm with accurate line searches (ref. *Adv Chem Phys*, 1987, vol LXVII, courtesy Academic Press).

the diagonal elements of which are analytically or numerically calculated, use analytically computed second derivative matrix from a lower level *ab initio* calculations, etc. Since analytical second derivatives are not readily available and its calculation involves 5–10 to N-times the computational cost involved in a gradient calculation, the use of the first derivative and quasi-Newton methods seems to be more advantageous than the Newton-type methods. Figure 4 shows the line searches conducted by a quasi-Newton method on the quadratic surface of Figure 3. The Newton-type methods are invoked only when there are acute convergence difficulties in the optimization arising from strongly coupled variables, shallow wells or narrow curved valleys on the function surface. The performance of these methods is dependent upon the accuracy of the Hessian.

Experience with non-derivative and derivative-driven minimizers

In the context of quantum chemistry, the form of the function to be minimized depends on the nature of constraints and the way the constraints have been incorporated. For example, the scheme of incorporation of orthogonality constraints in closed shell SCF problem as used by Mcweeny and further investigated by several authors[39, 40] results in poor convergence of the steepest descent method. But the convergence improves following the scheme due to Fletcher. A much simpler scheme based on Lagrange multiplier has shown promise in closed shell MC-SCF theory[40]. The non-derivative methods are conceptually simple, easily programmable but often fail to converge in the cases of strongly

dependent variables[41] or in problems with large number of variables (Powell method[27]). The first derivative methods, on the other hand, are faster than non-derivative methods as they make use of extra information about the function. The second derivative methods are, in general, superior to either of these. But it is hard to assess the relative advantage of the methods in general terms. The problems arise basically due to the difficulty of construction and storage of the Hessian or an approximation to it. The use of Newton-like methods in optimizing nonlinear parameters is rather widespread. Understandably, MC-SCF theory has made large scale use of Newton or Newton-like methods for optimizing the parameters in wave function including nuclear coordinates so much so that quadratic MC-SCF equations are often referred to as MC-SCF Newton Raphson equations[42–48]. Experience with non-derivative methods of optimization is quite varied[49]. The chief attraction of these methods in the context of quantum chemical optimization problem is easily understandable. The most widely used non-derivative method in this context has been the sequential univariate search (e.g. in optimizing orbital exponents). Powell's method has been used in the optimization of exponents[50], nonlinear parameters in general calculations[51] and linear expansion coefficients in SCF calculations[52]. The method performs well for small number of variables. However, for a large number of variables the method becomes inefficient requiring one to switch over to derivative methods.

Stochastic minimizers

From the discussion in the preceding sections, one can conclude that for (a) simple potential energy or energy-like surfaces with one or a very few minima which are sparsely distributed, gradient search could be quite efficient in locating these minima. However, for systems with many degrees of freedom, the gradient methods are slow and often fail to reach the global minima. This is because the search can get trapped in any one of the local minima and the algorithms on which the searches are based do not have any built-in mechanism which can drive it out of such minima. Techniques based on random search do not have this problem. The surface is scanned at random and the new point being sampled is a function of predetermined step sizes for various degrees of freedom. The method of simulated annealing[4–10] is an example of such a technique. Its objective is the location of the global minimum of a multivariate function possessing many local minima. Though stochastic in nature, the method of simulated annealing is not information free; it derives much of its power from the fact that the underlying probability distribution is the Gibbs distribution[10]. Numerical experiment[4–7] indicates that a sort of adaptive 'divide and
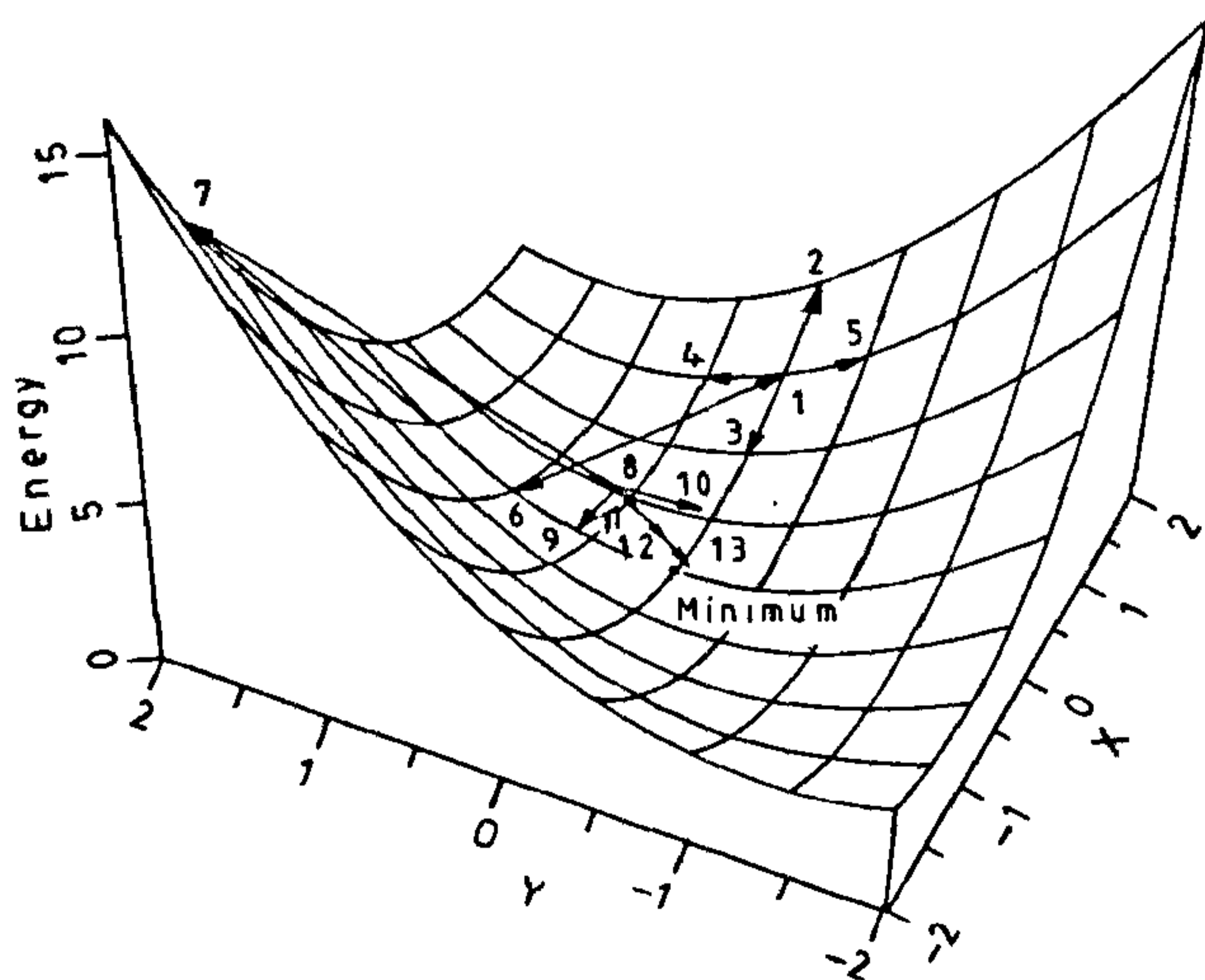
conquer' strategy emerges as a result of the annealing process. The gross features of the surface are sampled at high temperatures while the finer details are scanned at low temperatures. However, the optimal annealing schedule that ensures global minimization is tediously slow. Generating a practicable annealing schedule is therefore very important. We would now have a look at the method of simulated annealing rather qualitatively and recast the variational problem as a stochastic global minimization of an appropriate objective function(al).

Metropolis simulated annealing: the basic idea

The method of simulated annealing[4-8], first proposed by Kirkpatrick et al.[7,8], exploits similarities that exist between a multivariate optimization problem and the behaviour of a many particle system under cooling. What does really happen when liquids freeze or metals cool and anneal? At high temperatures, the molecules in the liquid move rather freely. When cooled slowly, the pure liquid freezes into a perfect crystal. One notes that the perfect crystalline state is the state of minimum energy of the system. If the cooling has been quick, the liquid could be 'quenched' and instead of reaching the perfectly crystalline state of minimum energy, it could end up in a polycrystalline or an amorphous metastable state having higher energy. The process of freezing of a liquid can therefore be viewed as a natural urge of the system to minimize its energy absolutely. Although ascribing such a definite *objective* to a natural process is fraught with danger and philosophically confusing, it helps us to visualize a mathematical minimization problem and map it onto a physical model. How can we do it? In what follows, we endeavour to achieve precisely this mapping.

Let $\mathcal{F}(\xi)$ be the constrained function (functional) to be minimized with respect to the parameters (linear or nonlinear or both). The constrained minimum as we have already seen, can certainly be found by adopting a gradient-based minimization technique (e.g. Marquardt–Levenberg algorithm)[38] or any other similar procedure[49]. Although the analogy is not perfect, one would be tempted to argue that the gradient-based algorithms mimic the thermodynamic behaviour of a liquid under *rapid* cooling. They follow the path of *'quick descent'* and more often than not end up in a *local minimum* (metastable state with respect to the global minimum). In contrast, the 'natural minimization' process discards the rapid descent path in favour of a more democratic philosophy. It adopts the attitude of a 'random walker' walking on an uneven terrain. The walker travels both uphill and downhill with a certain probability distribution that is controlled by the available 'energy fluctuation'. This is where the thermodynamic analogy comes handy. When a system is thermally equilibrated at a temperature

T, the probabilistic energy distribution $(P(E))$ among all the available energy states, called the Boltzmann distribution, is given by

$$P(E) \approx \exp(-E/kT).$$

The distribution is such that even at a low temperature, there is always a probability, however small, that the system is in a higher energy state. Naturally, there is also a probability that the system jumps out of the *local* energy minimum and moves further downhill to a more *global* one. But lower the temperature, less frequent is the probability of an uphill movement. In a typical multivariate minimization problem, one would therefore like to bring in the possibility of *both uphill and downhill movements, if one wants to shake off the possibility of being always trapped in a local minimum.*

Metropolis and co-workers[53] first suggested an algorithm which incorporates these features. Given a series of probable configurations, a simulated thermodynamic system is assumed to change its configuration from energy E_1 to E_2 with the probability $P = \exp\{-(E_2 - E_1)/kT\}$. Clearly, $P > 1$, if $E_2 < E_1$. In such a situation one assumes that $P = 1$, meaning that the system always favours and actually undergoes such a change. Even if $E_2 > E_1$, the system may sometimes undergo the change depending on the values of P (see later). If ΔE is within the scale of thermal fluctuation that is fixed by the thermodynamic temperature of the system, one can assume that the change is a plausible one and may actually occur albeit with a probability that is small and depends upon the temperature. Now the question is: how to bring in these features in a constrained minimization problem encountered by a quantum chemist.

Metropolis-simulated annealing and constrained minimization

In a problem of constrained minimization, the first step is to construct the constrained function(al) with all the relevant constraints incorporated suitably the minimization of which could be the goal. Obviously, the nature of the function(al) will depend on the nature of the variational parameters (linear or nonlinear), the number and type of constraints (normalization, orthogonality, decoupling, etc.) and the way the constraints are incorporated in the functional. For convenience, let us proceed with a functional $\mathcal{F}(\xi)$, $\{\xi\}$ denoting a set of variational parameters in the trial wave function. In a non-thermodynamical system, the use of Metropolis algorithm in a minimization procedure requires the following:

(i) a clear description of the possible configurations the system can assume,

(ii) a generator that can randomly reconfigure the system,

(iii) an objective function $\mathcal{F}$ (analogue of energy) the global minimum of which would be the target,

(iv) a control parameter T (analogue of temperature) that would fix the scale of fluctuation,

(v) practicable annealing schedule which controls the rate of cooling or heating.

In our problem, $\mathcal{F}(\xi_1, \xi_2, \ldots \xi_n)$ can be taken as the objective function. Any set of allowed values of the parameters $(\xi_1, \xi_2, \ldots \xi_n)$ can be used to define a *configuration* of the system in the parameter space. One can randomly choose one (say ξ_i) from among the n-parameters and update it as follows:

$$\bar{\xi}_i = \xi_i + (-1)^q \, \delta\xi_i ,$$

where q is a randomly generated integer and

$$\delta\xi_i = s \cdot r_i ,$$

where s is a scale factor $(s < 1)$ and r_i is a random number between 0 and 1.

(i) If $\mathcal{F}(\xi_1, \xi_2, \ldots \tilde{\xi}_i, \ldots \xi_n) < \mathcal{F}(\xi_1, \xi_2, \ldots \xi_i, \ldots \xi_n)$, the move is accepted as it lowers the value of the objective function.

(ii) If $\mathcal{F}(\xi_1, \ldots \tilde{\xi}_i, \ldots \xi_n) > \mathcal{F}(\xi_1, \ldots \xi_i, \ldots \xi_n)$ then we evaluate the quantity

$$P(F) = \exp(-\Delta F / T') ,$$

where $\Delta F = | \mathcal{F}(\xi_1, \ldots \tilde{\xi}_i, \ldots \xi_n) - \mathcal{F}(\xi_1, \ldots \xi_i, \ldots \xi_n)|^{1/2}$, T' has the dimension of energy and is proportional to what may be called the 'parameter temperature (T)' of the system. If $P(F) > r_i$ where r_i is a random number between 0 and 1 generated at the variable updating step, then also the move is accepted although it increases the value of the objective function. The rationale behind this choice is that the reconfigured system corresponds to an energy state that is within the limits of 'thermal fluctuation' at the parameter temperature T of the system. If $P(F) < r_i$, the move is rejected. T is held fixed and the system is reconfigured M-times (M sufficiently large, typically, $M = 100\,n$, where n is the number of parameters to be optimized) after which T is lowered by say $t\%$ (t is provided by the annealing schedule fixed beforehand). When the temperature (T) is sufficiently low, or attempts to reduce $\mathcal{F}(\xi_1, \xi_2, \ldots \xi_n)$ further become frustrating, it is time to stop and analyse the results. It may be noted that as $T \to 0$, the minimizing sequence of $\mathcal{F}(\xi)$ leads to the absolute minimum of it $(\mathcal{F}(\xi_0))$. Although the search is stochastic, it is not correct to assume that the SAM uses no information; indeed much of its power is derived from the fact that the underlying probability distribution is Gibbs distribution[51]. We may regard SAM as a

non-equilibrium procedure designed to cope with large barriers. This task is accomplished by starting the stochastic minimization at a very high temperature. Since the simulated dynamical system has large 'kinetic energy' at such temperatures, all barriers appear transparent to the 'walker' and the entire configuration space is therefore sampled. Given a long enough time for simulation and a slow enough rate of cooling, the algorithm, as we have stated already will guide the system to the global minimum with unit probability. The only difficulty with SAM is that the optimal cooling schedule, $T(t) = c/\ln(t)$, where c is some large number (constant) and t is the simulation 'time'[9], makes the algorithm costly in terms of computational labour. A compromise solution is to cool at faster rate and heat up the system whenever it tends to get struck during an annealing run. The other non-local stochastic minimizers, e.g. Monte Carlo and Molecular Dynamics methods work at finite non-zero temperatures. These minimizers can therefore surmount barriers $O(kT)$ and are not global minimizers. The gradient methods described already can be viewed as zero temperature exploratory minimizers which are strictly local in character.

Quantum Monte Carlo and the SAM based strategy

A few novel numerical techniques collectively called the Quantum Monte Carlo (QMC) methods attempt to represent the Schrödinger equation for many interacting fermions by random walk in the many-dimensional space in such a way that physical averages can be exactly calculated. The more well-known among the QMC methods of solving the Schrödinger equation are the diffusion Monte Carlo and Green's function Monte Carlo[54, 55]. The basis of diffusion Monte Carlo is that Schrödinger equation written in imaginary time will converge exponentially fast to the ground state. A wave function that is everywhere ≥ 0 can be directly interpreted as probability density so that Schrödinger equation for N interacting particles in imaginary time can be interpreted as a diffusion and branching process in 3N-dimensional space. To elaborate the idea a little more, we note that the Schrödinger equation in imaginary time reads as follows:

$$\frac{\partial \psi(x, t)}{\partial t'} = [D \nabla^2 + (E_T - V(x))] \psi(x, t), \quad (7)$$

where $D = \hbar^2/2m_e$ is the diffusion constant, $x = x_1, \ldots x_{3N}$ is the 3N dimensional coordinate vector of N-particles corresponding to electrons in real time, t' is the imaginary time, $V(x)$ is the usual coulomb potential for molecular system and E_T is an energy offset. It may be noted that for $t \to \alpha$, $\partial \psi / \partial t = 0$ and equation (7)

reduces to time-independent Schrödinger equation.

The formal solution of equation (7) can be written in terms of the eigenfunctions $\{\phi\}$ of H:

$$\psi(x,\,t) = C_0\,\phi_0\,\exp[-(E_0-E_T)t] +$$

$$\sum_i C_i\,\phi_i\,\exp[-(E_i-E_T)t].\qquad (8)$$

A look at $\psi(x,\,t)$ will immediately convince us about the truth of the assertion already made that the higher energy terms in the expansion decay exponentially with respect to the ground state. If one adjusts E_T to be E_0, one arrives at the steady state solution.

From a practical point of view, let us note that if the second term is dropped from equation (7), it becomes just the usual diffusion equation which can be simulated by Brownian motion of particles of 3N-dimensions. If $D\,\nabla^2$ is dropped, the result is a first-order rate equation with (E_T-V) as the rate constant. To simulate this equation, V is to be evaluated following each step of the random walk and the walker is either destroyed or copied depending upon the size of this term. The exact wave function is represented by a distribution of the random walkers resulting from the balance between the kinetic (diffusion) and potential (branching) processes.

Green's function Monte Carlo is a reformulation of the diffusion process such that no systematic error due to finite time step creeps in, the reformulation converting the differential equation into an integral equation the kernel of which is sampled exactly. The procedure may be viewed as a generalization of Von Neuman and Ulam's method for solving systems of linear equations. In addition, there is the so-called variational Monte Carlo method[56] in which one uses the MC methods to obtain expectation values employing the distribution $|\tilde{\psi}\rangle$ where $\tilde{\psi}$ is an analytical trial function as close as possible to ϕ_0.

The method elaborated in detail in the previous section can perhaps be regarded as a variation on the basic QMC theme. Thus, unlike in diffusion or Green's function MC we do not try to solve the time-dependent Schrödinger equation in imaginary time directed by random walks. Instead, we deal with the time-independent Schrödinger equation but recast the energy eigenvalue problem in the mould of stochastic search for the global minimum of an appropriately constrained energy functional at a sequence of temperatures $T_1,\,T_2,\,\ldots$ $T_n\to 0$ such that the global minimum of the functional in the given parameter space is reached in the limit of zero parameter temperature of the system. One may therefore feel that this amounts to replacing the diffusion part of QMC by simulated annealing. In that sense the SAM-based strategy advocated by us bears a relation, however distant, with the variational MC methods.

Quantum chemical applications based on MSA strategy

The unique feature of the MSA method lies in its ability to reach the global minimum without the knowledge of derivatives. We demonstrate this much sought for ability by locating the global minimum of a well-known model potential called Muller-Brown potential[57] $V^{MB}(x,y)$ (Figure 5). The three minima $(M_1,\,M_2,\,M_3)$ are clearly marked in the figure. The simplest objective function for the minimization is

$$F(x,y,V_L) = [V^{MB}(x,y)-V_L]^2,$$

where V_L is the lower bound to $V(x,y)$ which is held fixed during the annealing. Alternatively one can use

$$F = F'(x,y,V_L) = [V^{MB}(x,y)-V_L]^2 +$$

$$\lambda\{(\partial V/\partial x)^2 + (\partial V/\partial y)^2\}$$

A third choice is possible where

$$F = F''(x,y,V_L) = (\partial V/\partial x)^2 + (\partial V/\partial y)^2.$$

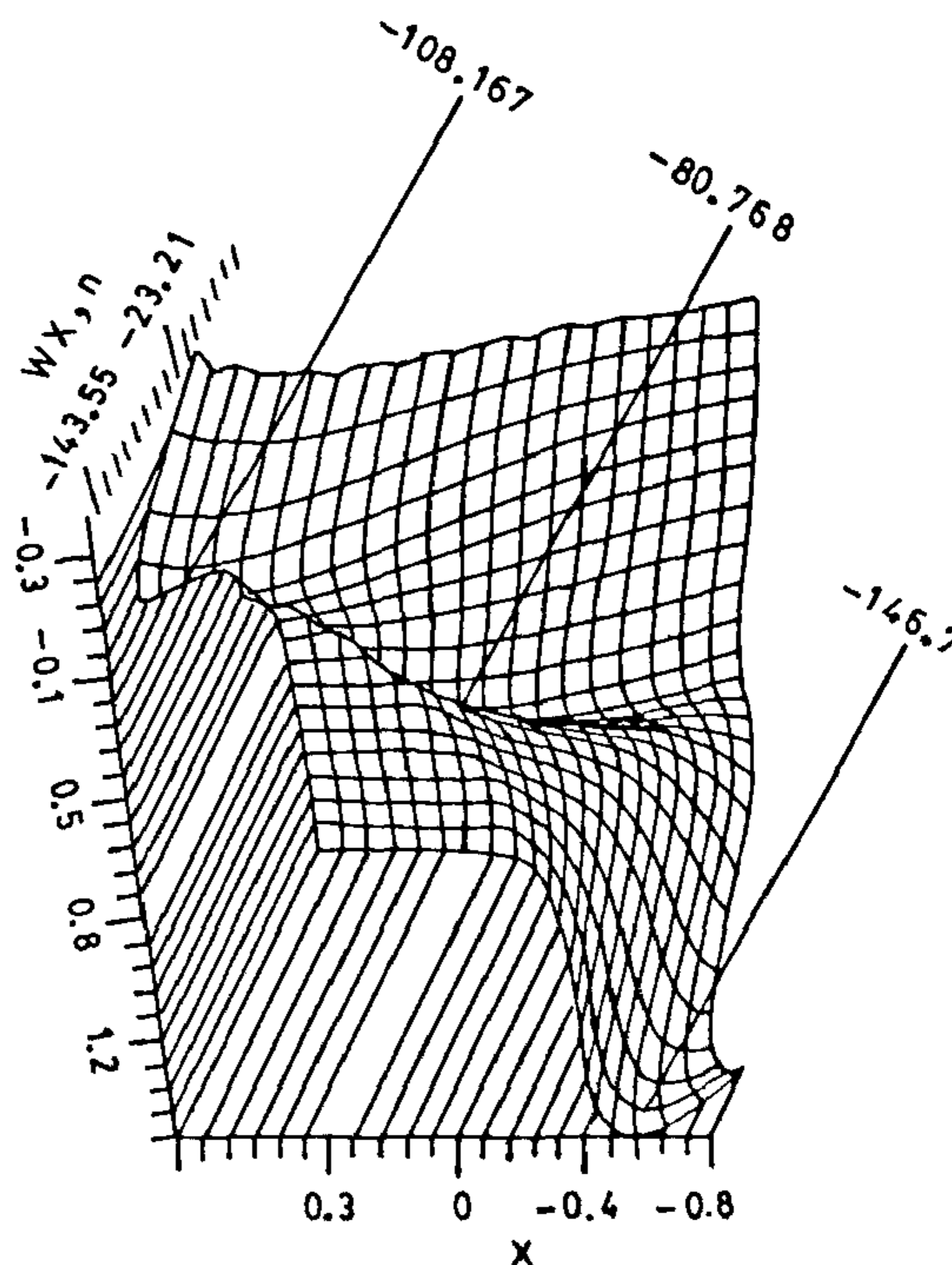A fourth strategy could be to set $\delta F = 2[V^{MB}(x,y)-V_L]\delta V = 0$ and solve the resulting equa-

Figure 5. The Muller-Brown potential (V_{MB}) consists of a sum of four terms of the form $V_{MB}(x,y) = A\,\exp[a\,(x-x_0)^2 + b\,(x-x_0)\,(y-y_0)+c\,(y-y_0)^2]$ (ref 57)

tions. Of these, the first, second and the fourth choices of the objective function will ensure that a minimum closest to the chosen V_L is reached as $T \rightarrow 0$. If V_L is well chosen, the global minimum is safely reached. But minimization based on F'' will lead to any stationary point, be it a true minimum, a saddle point or a maximum. To discriminate between these possibilities F'' needs to be augmented with curvature-controlled constraints, a strategy that will not be discussed in this review any further. Figure 6 shows how the average value of the objective function

$$F_{Av} = \sum_i p_i F(x_i, y_i, V_L) / \sum_i p_i,$$

p_i being the probability of generating the particular configuration (x_i, y_i) at a particular parameter temperature, changes as the system is cooled very slowly. The point to note is that F_{Av} decreases with fluctuation, the scale of fluctuation being determined by the parameter temperature T at that stage of cooling. Ultimately the search attains the global minimum. That this happens irrespective of the starting point is demonstrated in Figure 7. One can immediately see the tremendous potential of the method in the field of geometry optimization of large molecules, specially where a number of configurations of not-too-widely different energies are present. Figure 8 shows the energy profile during a MSA-based SCF geometry optimization run on the cyclohexane molecule in ground state, starting from a purely arbitrary reference geometry. The global minimum is reached easily if the cooling is slow. The efficacy

of the method is presently under further evaluation in the context of geometry optimization in clusters, biopolymers, etc.

When it comes to optimizing the parameters in the trial wave function, a variety of situations can arise as discussed already. The most interesting case is encountered when one must optimize $\tilde{\psi}$ for an excited state having the same symmetry as the ground state.
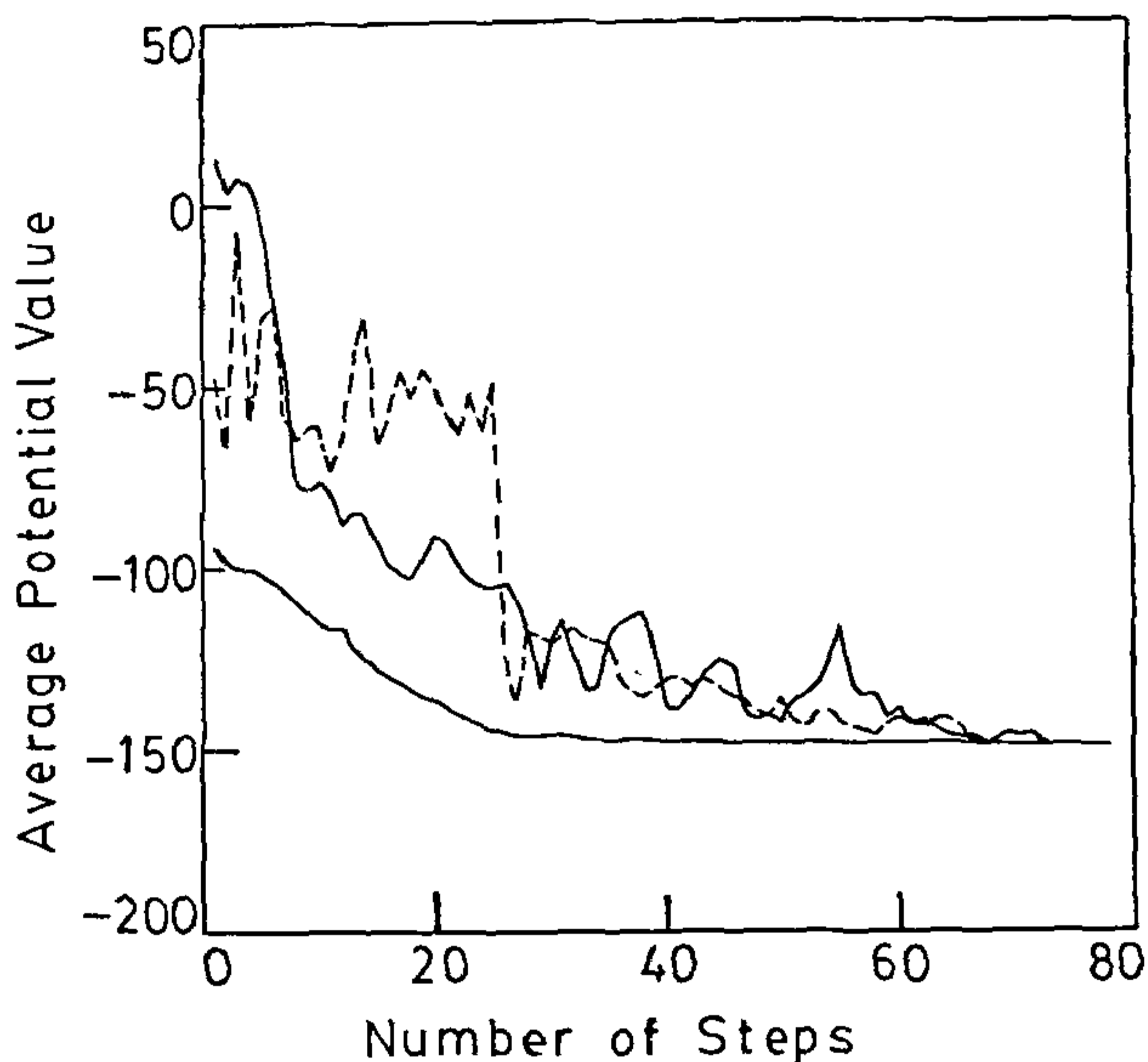

Figure 7. Profiles of V_{MB} during the SAM-based minimization starting from different points on the surface. The global minimum is reached irrespective of the starting point.
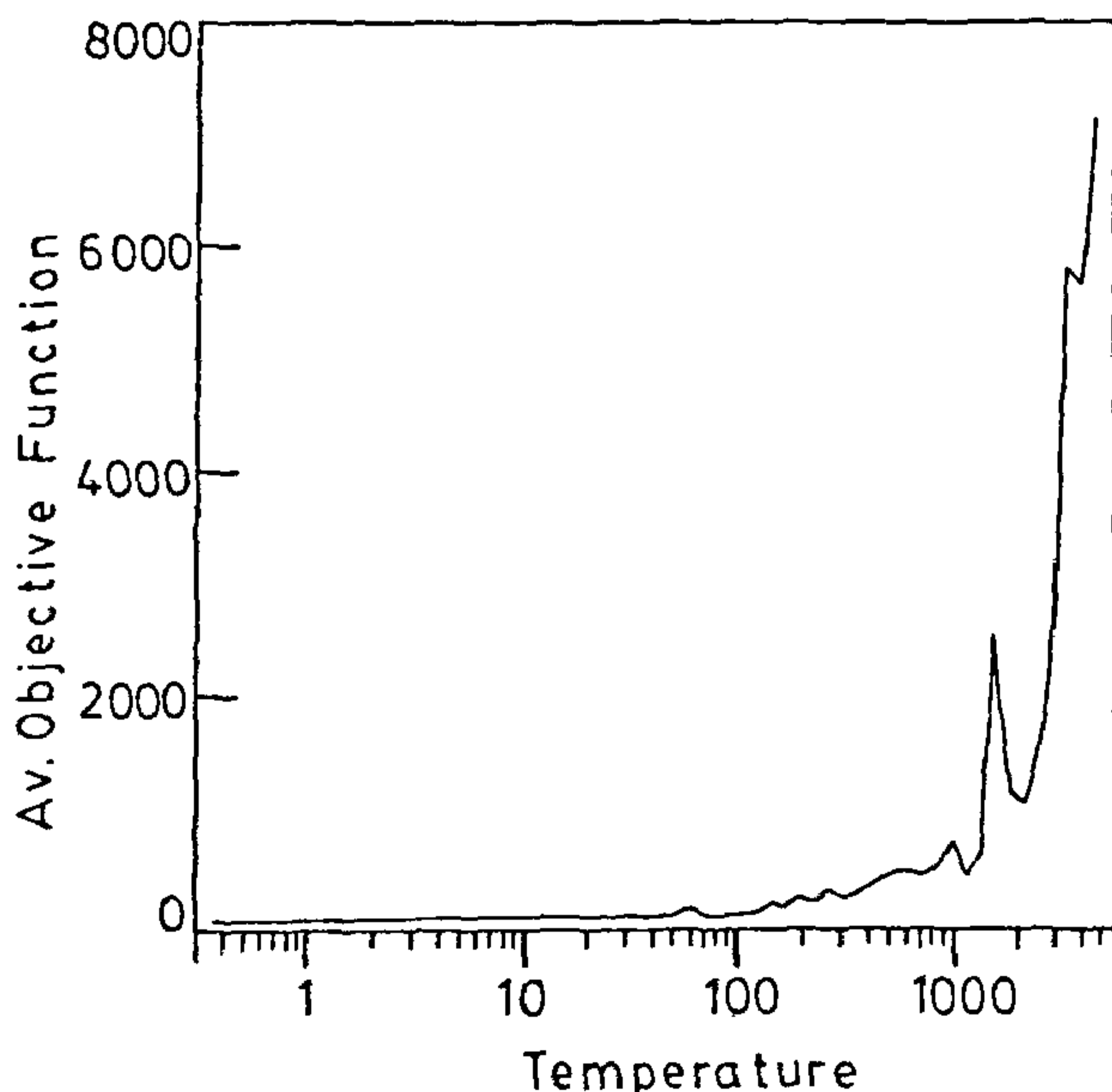

Figure 6. Profile of the average of the objective function ($\mathcal{F}$) used in the minimization of V_{MB} against temperature under the SAM.
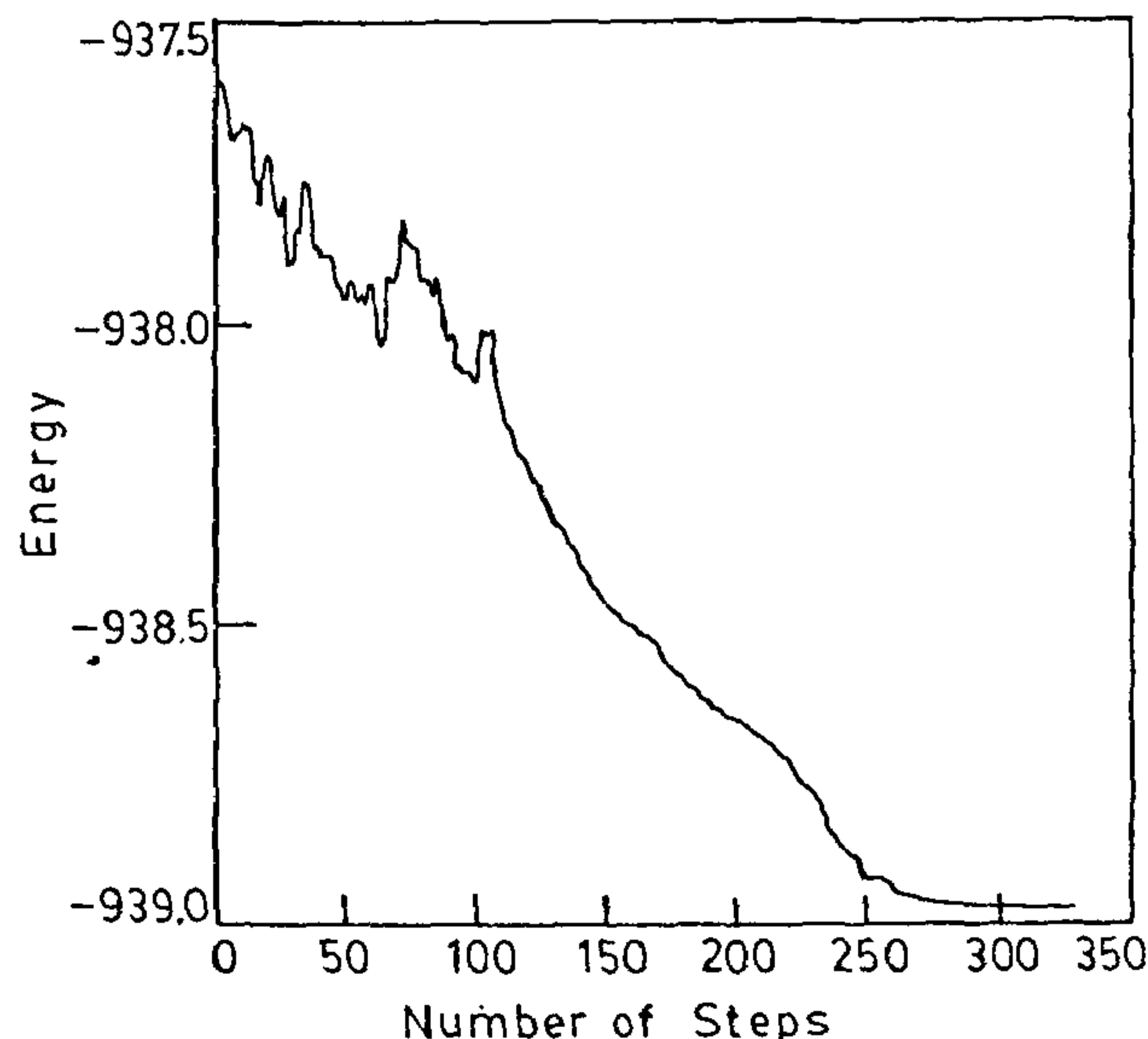

Figure 8. Energy profile obtained by MNDO-SAM-based geometry optimization of cyclohexane. Note the globality of the minimum.

Figure 9 shows the behaviour of the appropriate objective function during a typical annealing run with a 10-term trial CI-wave function for Li^+ ion. Figures 9 and 10 represent the behaviour of the ground-average objective function and energy, respectively. Breaks in Figure 9
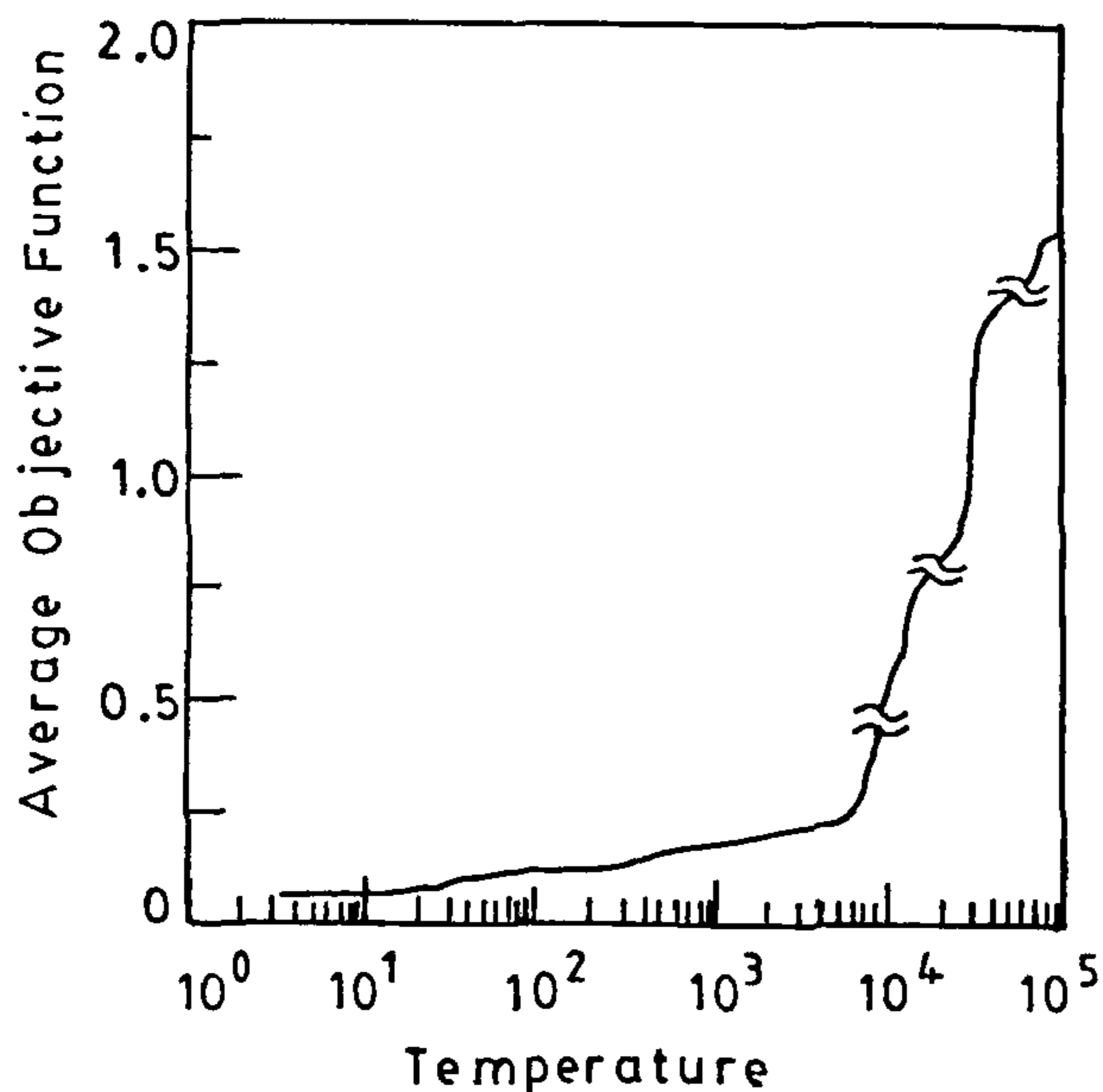

Figure 9. Profile of average objective function as a function of temperature in the optimization of the ground state wave function and energy of Li^+.
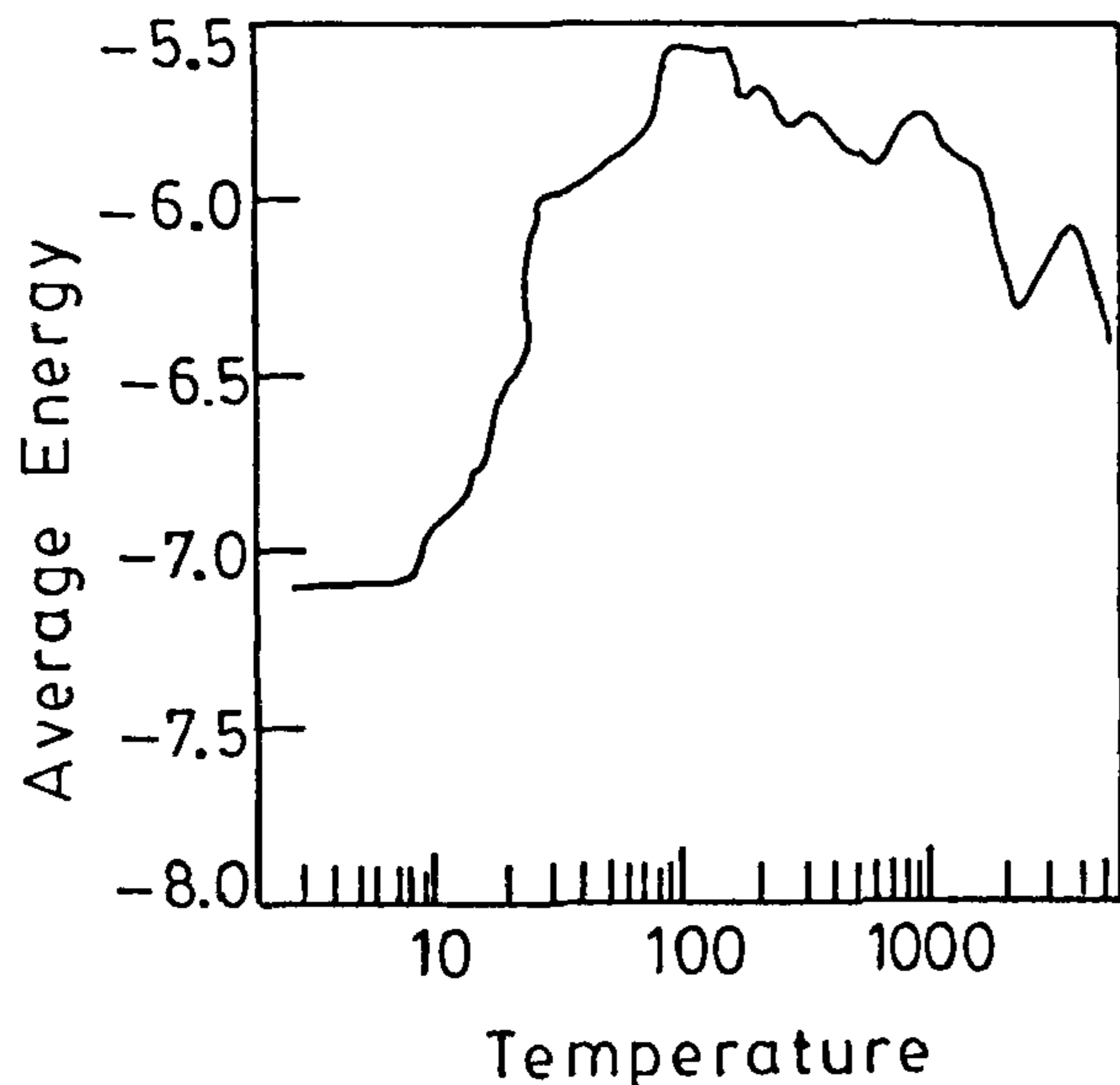

Figure 10. Profile of average energy as a function of temperature in the optimization of the ground state wave function and energy of Li^+ within the framework of SAM.

indicate changes in annealing schedule when the system is heated up after cooling for a predetermined length of time. In Figure 10 the average energy is seen to increase initially as the temperature is lowered. This happens as a consequence of a kind of adaptive 'divide and conquer' strategy built into the algorithm. In the initial stage, the normalization constraint dominates the minimization of the objective function leading to an increase in the average energy. In the end, however, energy minimization begins to control the proceedings and as a result the objective function or the target function as well as average energy decreases as $T \rightarrow 0$.

The scope of applicability of MSAM is wide and it is not possible to mention all the applications made so far in this brief review. However, we would like to conclude this section with an example of an MSAM-based first order MC-SCF calculation in a pathologically divergent problem. Generally, first order MC-SCF methods are prone to convergence difficulties due to the neglect of coupling between orbital and configuration spaces. A standard remedy is to invoke quadratic MC-SCF theory in such cases. However, it is much more desirable to have a remedy that works at the level of first order MC-SCF theory in view of its simplicity. The MSAM-based MC-SCF theory achieves this objective quite elegantly as can be seen from Figure 11. No external level or root shifting is necessary for achieving convergence.

Molecular dynamics: MSA-based approach

The molecular dynamics (MD) method is essentially a dynamical method for applying variational principle, in which eigenfunctions for lower energy electronic states
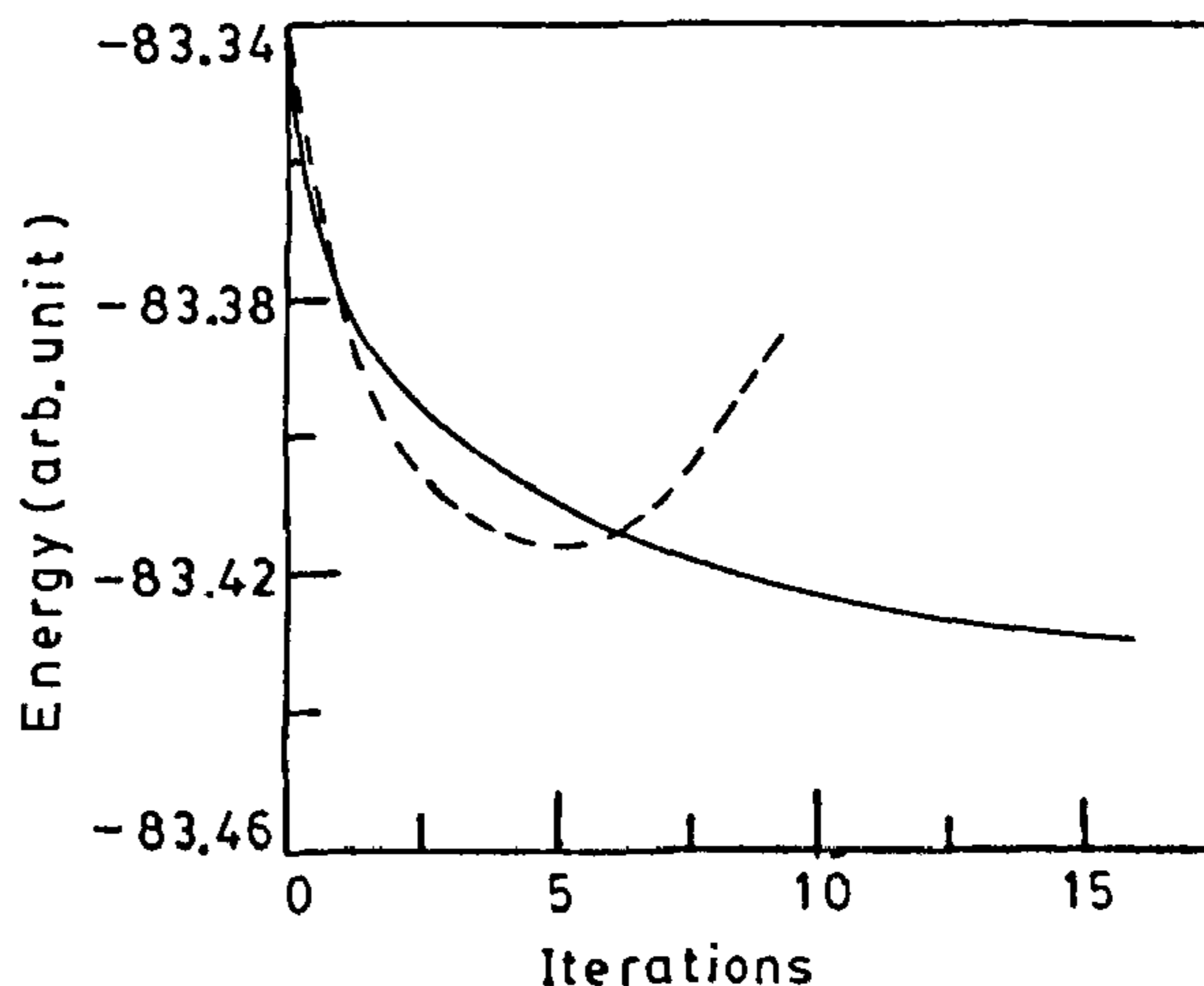

Figure 11. The energy profile in level shifted (—) and unshifted (– –) OGM-based calculations on the $^1\pi\pi^*$ state of propynal. Note the divergence when shifting is withdrawn.

are determined simultaneously. The essential step is to treat the electronic wave function as a dynamical variable and define a fictitious Lagrangian, called the Car–Parrinello Lagrangian for the electronic system[58]. The method has been found to be specially useful for searching out locally stable structures of atomic clusters and condensed phases. It is easily understood that the search requires electronic energies at thousands of different nuclear configurations for locating the minimum energy structures. Car and Parrinello demonstrated for the first time how the electronic structure problem can be solved with great facility in conjunction with classical dynamics for nuclear motion using simulated annealing. The classical Lagrangian for nuclear motion in an adiabatic electronic state α is

$$L = (1/2) \sum_{i=1}^{N} M_i R_i^2 - \text{Min} \langle \psi_\alpha(p) \mid H \mid \psi_\alpha(p) \rangle .$$

The index i runs over all the N nucleii. p is the vector of parameters in ψ. When the motion takes place on the ground state surface ($\alpha = 0$), the only constraint on ψ is that it must be normalized. For motion on excited surface, orthogonality to lower electronic states must be enforced. The expectation value $\langle \psi_\alpha(p) \mid H \mid \psi_\alpha(p) \rangle$ will in general be a complicated function of the parameters in ψ. The minimization is therefore a complex problem and the possibility of the minimization getting trapped in undesired local minima is always there.

Car and Parrinello's implementation brings in simulated annealing in the operation of the molecular dynamics scheme and rescues the minimization from the problem of arriving at metastable configurations corresponding to local minima. To achieve this, they introduce a fictitious kinetic energy associated with electronic parameters:

$$L' = (1/2) \sum_{i=1}^{N} M_i R_i^2 +$$

$$(1/2) \sum_{s=1}^{N_e} m_s \dot{p}_s^2 - \langle \psi_\alpha \mid H \mid \psi_\alpha \rangle .$$

One then carries out a MD simulation using standard techniques of constrained dynamics both for the nuclear coordinate and the N_e number of fictitious electronic degrees of freedom. To minimize the electronic energy functional kinetic energy is slowly removed from the electronic degrees of freedom which amounts to cooling the system to lower temperatures. When the fictitious electronic velocities $(\dot{p}_s)$ are cooled to a near zero temperature, classical statistical mechanics ensures that the system has reached the minimum of potential energy with respect to the electronic parameters $\{p_s\}$, thus

solving the electronic variational problem. At this stage, where the electronic parameter velocities are frozen, the nuclei undergo physically meaningful classical motion on the αth BO surface.

$$L = \sum \mu \langle \psi_i \mid \psi_i \rangle + (1/2) \sum_I M_I R_I^2 +$$

$$(1/2) \sum_v \mu_v \alpha_v'^2 - E [\{ \psi_i \} \{ R_I \} \{ \alpha_v \}] ,$$

where μ is a fictitious mass associated with the electronic wave functions, E is the energy functional, R_Is are the nuclear positional coordinates, α_vs are relevant system parameters present in ψ. M_I and μ_v may be regarded as arbitrary parameters in appropriate units. The stationarity condition on L generates dynamics of evolution of ψ_i, R_I and α_v. By varying velocities, i.e., $\dot{\psi}_i$, $\dot{R}_I$, and $\dot{\alpha}_v$, the system temperature can be reduced and in the limit $T \to 0$, the minimum energy state (equilibrium) can be reached. This alternative scheme which may be called dynamical simulated annealing method has also found many successful applications[59,60].

Some problems with SAM

The SAM has many positive features and its performance as a stochastic global minimizer has been noteworthy. To date, it is the only global minimizer that has proven capability of reaching the global minimum. Having noted this, we must also mention that it has some weakness as well. One of it is endemic to all Monte Carlo methods, viz., they are very slow towards the end of the search. Secondly, unless a suitable annealing schedule is chosen, the problem of getting stuck at a metastable state cannot be eliminated. The theoretically optimum schedule being very slow, practicable schedules are usually constructed by trial and error. More work in this direction is necessary to make the method operationally attractive as an all-purpose minimizer.

Postscript

Although we have focused our discussion on quantum chemical optimization problems, the MSAM is a perfectly general all purpose global optimizer, provided a suitable objective function can be constructed for the specific problem. We have employed this recipe for saddle point location[61], solving large systems of linear and nonlinear equations, matrix inversion, variance minimization[62], etc. with remarkable success. It would not be an exaggeration to comment that stochastic minimizers will play a very important role in optimization problems in the coming years. Of course, MSAM or the Car–Parrinello dynamical SAM will occupy the centre stage whenever global

minimization is the target of the search, be it in quantum chemical optimization problems or in any other field. However, in the recent years the genetic algorithm[63] has been emerging as a strong contender for acceptance as an all-purpose stochastic minimizer. Probably, a hybrid of the two could be a welcome alternative. At the moment, the search is on.

1. McWeeny, R. and Sutcliffe, B. T., in *Methods of Molecular Quantum Mechanics*, Academic Press, London and New York, 1969, 67–91.
2. Miller, H. G. and Ddreizler, R. M , *Nucl. Phys. A.*, 1979, **316**, 32–39.
3. Hendikovic, J , *Chem. Phys. Lett*, 1982, **90**, 198–201.
4. Dutta, P. and Bhattacharyya, S. P., *Chem. Phys. Lett.*, 1989, **162**, 67–71; 1990, **167**, 309–314.
5 Dutta, P. and Bhattacharyya, S. P., *Phys Lett A.*, 1990, **148**, 331–337.
6. Dutta, P., Mazumdar, D. and Bhattacharyya, S. P., *Chem Phys. Lett.*, 1991, **181**, 293–297; Dutta, P. and Bhattacharyya, S. P., *Chem Phys Lett.*, 1991, **184**, 330–334.
7. Kirkpatrick, S., Gelatt, C. D. and Vecchi, M. P., *Science*, 1983, **220**, 671–680.
8. Kirkpatrick, S., *J Stat. Phys.*, 1984, **34**, 975.
9. Gidas, B., *J. Stat. Phys*, 1985, **39**, 73.
10. Donnelly, R. A., *Chem. Phys Lett.*, 1987, **136**, 274–278
11. Hylleraas, E A. and Undheim, B., *Z Phys.*, 1930, **65**, 759–763; Macdonald, J. K. L., *Phys. Rev.*, 1933, **43**, 830–833.
12. Fiacco, A. V. and McCormik, G. P., in *Non-Linear Programming: Sequential Unconstrained Optimization Technique*, Wiley, New York, 87–107; Morrison, D. D., *SIAM J Numer. Anal.*, 1968, **5**, 83
13. Weber, T. A and Handy, N. C., *J. Chem Phys*, 1969, **50**, 2214–2215.
14 Wang, P. S, C., Benston, M. L. and Chong, D. P., *J. Chem. Phys.*, 1973, **59**, 1721–1725.
15. Mukherjee, D , *Int. J. Quantum Chem.*, 1975, **9**, 943–949.
16. Das, K., Khan, P. and Bhattacharyya, S. P , *Pramana*, 1985, **25**, 281–290.
17. Powell, M J D. (ed), in *Non-Linear Optimization*, Academic Press, New York, 1982, 33–89.
18. Hooke, R. and Jeeves, T. A., *J. Assoc. Comput Machin.*, 1961, **8**, 212.
19. Kowalik. J. and Osborne, M R , in *Methods for Unconstrained Optimization Problems*, Elsevier, New York, 1968, 67–102.
20 Rosenbrock, H. H , *Comput. J.*, 1960, **3**, 175.
21. Powell, M J. D., *Comput. J.*, 1964, **7**, 155.
22. Fletcher, R. and Powell, M J D., *Comput. J.*, 1963, **6**, 163.
23. Collins, J. B., Schleyer, P. V. R., Binkley, J. S. and Pople, J. A., *J. Chem. Phys*, 1976, **64**, 5142–5151.
24. Miele, A. and Cantrell, J. W., *J. Opt Appl.*, 1969, **3**, 459.
25 Cragg, E. E. and Levy, A. V., *J. Optim Theory Appl.*, 1969, **4**, 191
26 Fletcher, R. and Reeves, C. M., *Comput J*, 1964, **7**, 149.
27. Davidon, W., in *Argone National Lab. Report*, ANL-5990.
28. Murtagh, B. A. and Sargent, R. W. H., *Comput J*, 1972, **13**, 185.
29. Broyden, C. G., *J. Inst. Math. Appl*, 1979, **6**, 76, Fletcher, R , *Comput. J.*, 1970, **13**, 317, Goldfarb, D , *Math. Comput.*, 1970, **24**, 23, Shanno, D. F., *Math. Comput.*, 1970, **24**, 647.
30. Davidon, W C., *Math Program*, 1975, **9**, 1
31. Huang, H. Y., *J. Optim. Theor. Appl.*, 1970, **5**, 405.
32. Dixon, L. C. W , *Math. Program.*, 1972, **2**, 383.
33. Biggs, M. C., *J. Inst. Math. Appl*, 1973, **12**, 337.
34. Schlegel, H. B , *J Comput. Chem.*, 1982, **3**, 214–218.
35 Davidon, W. C., *Math. Program.*, 1975, **9**, 28.
36. McKelvey, J. M. and Hamilton (Jr), J. F., *J. Chem. Phys.*, 1984, **80**, 579–581.
37. Csaszar, P. and Pulay, P., *J. Mol. Struct.*, 1984, **114**, 31–34.
38. Murray, W., in *Numerical Methods for Unconstrained Optimization*, Academic Press, London and New York, 1972, 29–54.
39. Sleeman, D H., *Theor. Chim. Acta*, 1968, **11**, 135–144, Hillier, I. H. and Saunders, V. R., *Proc. R. Soc.*, 1970–71, **A320**, 161–173.
40. Mukherjee, D., *Proc Indian Acad. Sci.*, 1978, **A87**, 37–43; Adnan, S. S. Z., Bhattacharyya, S. P. and Mukherjee, D., *Int. J. Quantum Chem*, 1978, **14**, 289–297; Bhattacharyya, S. P. and Mukherjee, D., *Chem Phys. Lett.*, 1979, **66**, 511–516, *Int. J. Quantum Chem.*, 1981, **20**, 1165–1177.
41. Clementi, E., *J Chem. Phys.*, 1964, **40**, 1944–1945
42. Ransil, B. J., *Rev. Mod. Phys*, 1960, **32**, 239–244.
43. Boys, S. F. and Handy, W. C., *Proc. R. Soc.*, 1969, **A310**, 63–78, 1969, **A311**, 309–329
44. Bishop, D M. and Leclerc, J. C., *Mol. Phys.*, 1972, **24**, 979–992.
45 Hillier, I. H and Saunders, V. R., *Int. J Quant. Chem.*, 1970, **4**, 503.
46. Yeager, D. L. and Jorgensen, P , *J. Chem. Phys*, 1979, **71**, 755–760.
47. Roothaan, C. C. J., Detrich, J. and Hopper, D. G., *Int. J. Quant Chem.*, 1979, **S13**, 93.
48. Hans-Joachim Werner., *Adv. Chem. Phys.*, 1987, **67**, 1–62; Shepard, R., *Adv. Chem. Phys*, 1987, **67**, 63–200.
49. Dutta, P. and Bhattacharyya, S. P., *Pramana*, 1990, **34**, 13–21.
50 Solomon, C. E., *Int. J. Quantum Chem.*, 1971, **5**, 319.
51. Pan, K. C. and King, H. F., *J. Chem Phys.*, 1972, **56**, 4667–4688.
52. Bonacic, V. and Koutecky, J., *J. Chem. Phys.*, 1972, **56**, 4563–4573.
53. Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H and Teller, E., *J Chem. Phys.*, 1953, **21**, 1087–1092.
54 Vanderbilt, D and Louie, S. G , *J. Comput. Phys.*, 1984, **56**, 259–271; Laarhoven, P. J. M. and Aarts, E H. L., in *Simulated Annealing: Theory and Applications*, Reidel, Dordrecht, Holland, 1987, 187–203; Corana, A., Marchesi, M., Martini, C. and Ridella, S., *ACM Trans. Math. Software*, 1987, **13**, 262–280.
55. Lester, (Jr.) W. A. and Hammond, B. L., *Annu. Rev. Phys. Chem.*, 1990, **41**, 283–311 and references cited therein.
56. Umrigar, C. J., Wilson, K. G. and Wilkins, J W., *Phys. Rev. Lett.*, 1988, **60**, 1719–1722; Reynolds, P. J., Ceperley, D. M , Alder, B. J. and Lester, W. A (Jr), *J. Chem. Phys*, 1982, **77**, 5593–5603, Harrison, R J. and Handy, N. C., *Chem. Phys. Lett.*, 1985, **113**, 257–263.
57. Müller, K. and Brown, L. D., *Theor. Chim. Acta.*, 1979, **53**, 75–93.
58. Car, R. and Parrinello, M., *Phys. Rev. Lett.*, 1985, **55**, 2471–2474, *Chem Phys. Lett*, 1987, **139**, 540–545; *Phys. Rev. Lett*, 1988, **60**, 204–207.
59. Estrin, D. A., Tsoo, C. and Singer, S. J., *J Chem. Phys.*, 1990, **93**, 7201–7212.
60 Chaehaum, H. and Mohallem, J. R , *Mol. Phys.*, 1990, **70**, 391–400.
61. Dutta, P., Chaudhury, P., Mukhopadhyay, P and Bhattacharyya, S. P , *Chem Phys. Lett.*, (submitted, 1994).
62. Dutta, P. and Bhattacharyya, S. P., *J Comput Phys.* (submitted, 1994).
63 Xiao, Y. and Williams, D E., *Chem Phys Lett*, 1993, **215**, 17–24; Hartke, B , *J. Phys. Chem.*, 1993, **97**, 9973–9976.