*Journal of Global Optimization* **29:** 297–314, 2004.
© 2004 *Kluwer Academic Publishers. Printed in the Netherlands.*

# A Geometric Terrain Methodology for Global Optimization

A. LUCIA, P.A. DIMAGGIO and P. DEPA
*Department of Chemical Engineering, University of Rhode Island, Kingston, RI 02881-0805, USA. (e-mail: lucia@egr.uri.edu; pdim2813@postoffice.uri.edu; pravind@egr.uri.edu)*

(Received 31 July 2003; accepted 7 August 2003)

**Abstract.** Global optimization remains an important area of active research. Many macroscopic and microscopic applications in science and engineering still present formidable challenges to current global optimization techniques. In this work, a completely different, novel and general geometric framework for continuous global optimization is described. The proposed methodology is based on intelligent movement along the valleys and ridges of an appropriate objective function using downhill, local minimization calculations defined in terms of a trust region method and uphill integration of the Newton-like vector field combined with intermittent SQP corrector steps. The novel features of the proposed methodology include new rigorous mathematical definitions of valleys and ridges, the combined use of objective function and gradient surfaces to guide movement, and techniques to assist both exploration and termination. Collisions with boundaries of the feasible region, integral curve bifurcations, and the presence of non-differentiabilities are also discussed. A variety of examples are used to make key concepts clear and to demonstrate the reliability, efficiency and robustness of terrain methods for global optimization.

**Key words:** geometric optimization, terrain methods, valleys and ridges.

## 1. Introduction

Global optimization is still a very important and challenging interdisciplinary area of research and with the recent thrust in multi-scale simulation and optimization, global methods will continue to be challenged. However, many good global optimization techniques are presently available. Some of these methods are general-purpose while others are specific to a given application area. There are general-purpose deterministic methods like the tunneling algorithms of Levy and Montalvo [8] and Bahren and Protopopescu [2], $\alpha$BB by Maranas and Floudas [11], and the interval analysis methods of Hansen [6] and Schnepper and Stadtherr [15]. There are also general-purpose stochastic methods for global optimization like the stochastic differential equations approach of Zirilli and co-workers [1] and the work of Bilbro [4]. Most general-purpose methods tend not to exploit any problem specific information and are therefore widely applicable. There are also deterministic and stochastic algorithms designed for particular classes of problems. Methods here include the tangent plane method of Michelsen [12] for phase equilibrium by Gibbs energy minimization, the chain-of-states method of Sevick et al. [16] and the nudged elastic band (NEB) methods of Henkelman et al. [7]

for transition state and reaction pathway calculations, and Monte Carlo and molecular dynamics approaches for conformational analysis by potential energy minimization like the work of Bolhius et al. [5]. There are many other methods, far too many to list here.

Recently Lucia and Yang [9, 10] have developed a novel geometric methodology for the global optimization of least-squares objective functions. The ideas that form the theoretical and computational framework for this geometric methodology are (1) that stationary points are connected along valleys and ridges and (2) that it is rather straightforward to move from one stationary point to another by moving up and down the landscape or terrain of the least squares function using the Newton and tangent vector fields. While there are considerable details to this terrain methodology, the overall strategy is quite simple and made up of a unique blend of equation solving, nonlinear programming and eigenvalue–eigenvector tasks. Moreover, Lucia and Yang [9, 10] have applied their terrain methodology to a variety of small and large-scale problems including equations of state, reactor and distillation examples and the associated numerical results clearly illustrate that terrain methods represent a reliable and efficient global equation solving methodology.

This paper describes the unique theoretical framework and related numerical results for the global optimization of general objective functions by terrain methods and is organized in the following way. The details of our framework for terrain methods for general objective functions are presented in Section 2. More subtle considerations concerning integral curve bifurcations and non-differentiabilities are then discussed in Section 3. In Section 4, numerical results for phase split/Gibbs energy calculations and transition state and reaction pathway calculations in molecular modeling are presented. Conclusions are given in Section 5.

## 2. Mathematical Background

The terrain methodology is an overall philosophy for moving from one stationary point to another that requires reliable local equation solving tools as well as reliable and efficient uphill exploration. Reliable local equation solving is required to find both a first stationary point from an arbitrary initial guess and any subsequent minima, saddle points and singular points to desired accuracy. Reliable and efficient uphill exploration, on the other hand, is needed to move from one stationary point to the next along a valley. Other peripheral tasks such as reliable and efficient eigenvalue–eigenvector calculations and solving nonlinearly constrained optimization problems are also needed. With each choice of method for a given task comes a slightly different terrain method.

### 2.1. PRELIMINARIES

Let $\phi$ be a C$^3$ objective function in the unknown variables $Z$ that takes $R^n$ into $R$ and let g and H be the gradient and Hessian matrix associated with $\phi$. Furthermore, let $L$ denote any level curve of $\phi$, $\mathcal{L}$ be any set or collection of level curves and let $\| \ \|$ denote the two norm of any vector. Finally, let $M = \mathrm{H}^T\mathrm{H} + \Sigma \mathrm{g}_i \mathrm{G}_i$, where $\mathrm{G}_i$ is the second derivative matrix of the $i$th component of the gradient.

For general global optimization calculations we use two surfaces to guide movement—a primary objective function, $\mathrm{g}^T\mathrm{g} = \|\mathrm{g}\|^2$, and a secondary objective function surface $\phi$. Note that saddle points and minima of $\phi$ are actually minima of $\mathrm{g}^T\mathrm{g}$ and that saddle points on $\mathrm{g}^T\mathrm{g}$ correspond to points where H is singular. Thus the surface $\mathrm{g}^T\mathrm{g}$ contains all of the stationary points of $\phi$ plus addition stationary points which correspond to singular points of H. The stationary points on each of these surfaces will be connected and will define valleys and ridges on each surface. However, it is important for the reader to realize that valleys on $\phi$ may or may not coincide with valleys on $\mathrm{g}^T\mathrm{g}$.

### 2.2. INITIALIZATION

Terrain methods are made up of up and downhill movement. Initial movement is always downhill and for this the starting point is arbitrary. Subsequent movement from any stationary point, whether up or downhill, is always initialized using eigen-information from either the primary or secondary objective function. Thus reliable and efficient calculation of some of the eigenvalues and eigenvectors of the matrices H and M is needed. The general rules for movement from a stationary point of $\mathrm{g}^T\mathrm{g}$ are straightforward and are based on the eigenvalues of M. That is, uphill movement is generally initialized in the eigen-direction associated with the smallest positive eigenvalue of M while downhill movement is initiated in the eigen-direction associated with the largest negative eigenvalue of M.

There are occasions, however, when eigen-information from $\mathrm{g}^T\mathrm{g}$ can be misleading. Thus some of the eigenvalues and eigenvectors for both H and M are needed. Let $\lambda_{\mathrm{M}}$ and $\mathrm{c}_{\mathrm{M}}$ be the eigenvalue and normalized eigenvector of M selected by the rules described in the previous paragraph. Also let $\lambda_{\mathrm{H}}$ and $\mathrm{c}_{\mathrm{H}}$ be the eigenvalue and normalized eigenvector of H selected by the rules described in the previous paragraph. If $57.295 \arccos[\mathrm{c}_{\mathrm{H}}^T \mathrm{c}_{\mathrm{M}}] > 1$ and $\lambda_{\mathrm{H}} < \lambda_{\mathrm{M}}$, then $\lambda_{\mathrm{H}}$ and $\mathrm{c}_{\mathrm{H}}$ are used to initiate movement; otherwise $\lambda_{\mathrm{M}}$ and $\mathrm{c}_{\mathrm{M}}$ are used. This test first checks the angle between the normalized eigenvectors of H and M. If that angle is less than 1° and the principle eigenvalue of H is less than the principle eigenvalue of M, then eigen-information from the secondary objective function is used to initiate movement to the next stationary point. The usefulness of this test is illustrated using a molecular modeling example in the section containing numerical results.

### 2.3. DOWNHILL MOVEMENT

Downhill movement is always calculated using a trust region method. Thus downhill steps are defined by

$$\Delta = -\beta \Delta_{\mathrm{N}} + (\beta - 1)\mathrm{g} \tag{1}$$

where $\Delta_N = H^{-1}g$ is the Newton direction and where $\beta \in [0, 1]$ is determined by the following simple rules. If $\|\Delta_N\| \leqslant R$, then $\beta = 1$, where $R$ is the trust region radius. If $\|\Delta_N\| > R$ and $\|g\| \geqslant R$, then $\beta = 0$. Otherwise, $\beta$ is the unique value in Equation (1) on [0,1] that satisfies $\|\Delta\| = R$. The new iterate is accepted if it reduces $\|g\|$. Otherwise, the new iterate is rejected, the trust region radius is reduced and the calculations are repeated until a reduction in $\|g\|$ occurs. Furthermore, if during downhill (or uphill) movement, $\|g\|/\|\Delta_N\| \leqslant \zeta$, then the quadratic acceleration step given by

$$\Delta = -M^{-1}Hg \tag{2}$$

is used. Downhill movement is terminated when either $\|g\| \leqslant \varepsilon$ or $\|Hg\| \leqslant \varepsilon$ where $\varepsilon$ is a convergence tolerance and can result in convergence to either a minimum, saddle point or singular point of $\phi$.

### 2.4. UPHILL MOVEMENT

Uphill movement is based on predictor-corrector calculations. Uphill predictor steps are simply controlled uphill Newton steps defined by

$$\Delta = \alpha \Delta_N \tag{3}$$

where the step size $\alpha \in (0, 1]$. These uphill Newton steps tend to follow valleys reasonably well but do drift some. Therefore, corrector steps are used intermittently to return iterates to the current valley and are invoked when

$$\theta = 57.295 \arccos\left(\frac{\Delta_N^T c}{\|\Delta_N\|\,\|c\|}\right) \geqslant \Theta \tag{4}$$

where c is the current estimate of the eigenvector associated with the appropriate eigenvalue of M and $\Theta$ is $5^\circ$. Corrector steps are defined by iteratively solving the nonlinear programming (NLP) problem

$$\text{opt } g^T H^T Hg \text{ such that } g^T g = L \tag{5}$$

The optimality conditions associated with Equation (5) are

$$g^T g - L = 0 \tag{6}$$

$$MHg - \lambda Hg = 0 \tag{7}$$

where $L$ is the value of the level set at the last predictor iterate. Note that by iteratively solving Equation (5), the terrain methodology generates a sequence of corrector steps each time it is necessary to return to a valley.

What distinguishes terrain following methods from all other global optimization methods is their unique rigorous nonlinear programming (NLP) characterization of valleys and ridges. Let $V$ denote a valley or ridge defined by the set

$$V = \{\text{opt}\, \mathrm{g}^T \mathrm{H}^T \mathrm{Hg} \text{ such that } \mathrm{g}^T \mathrm{g} = L \text{ for all } L \in \mathcal{L}\} \tag{8}$$

That is, $V$ is a set of local minima of $\mathrm{g}^T\mathrm{H}^T\mathrm{Hg}$ over a given set of neighboring level curves. Note that Equations (6) and (7) clearly show that valleys and ridges are defined by a sequence of constrained eigenvalue-eigenvector problems and that $\lambda$ in Equation (7) is actually the Lagrange multiplier associated with the level constraint defined in Equation (5). Moreover, the task of intermittently solving Equation (5) provides a straightforward means of updating both the principle eigenvector that defines the current valley, Hg, and its associated eigenvalue, $\lambda$.

### 2.5. EIGENVALUE–EIGENVECTOR CALCULATIONS

Except for the initial downhill or uphill movement, eigenvalues and eigenvectors of H and M are often required to begin either uphill or downhill movement. However, only a few eigenvalues and eigenvectors are needed. For these calculations we use the inverse power method together with incomplete factorization to compute a few eigenvalues and eigenvectors. Also we never actually form matrix products like $\mathrm{H}^T\mathrm{H}$ to avoid rounding errors.

### 2.6. TERMINATION

Lucia and Yang [9, 10] base their termination criterion on something they call limited connnectedness. All that this means is that stationary points are really only connected to neighboring stationary points along specific eigendirections. We assume that the number of important connections between stationary points along valleys and ridges is limited to four or less and is related to dominant geometric distortions (i.e., $+/-$ the smallest positive eigendirection and $+/-$ the most negative eigendirection) caused by the strongest 'attractions' between neighboring stationary points. This makes it possible to dynamically catalogue connections in a set, C, and to conclude that all of the important connections between stationary points have been explored when C is empty. Thus termination occurs when C is the empty set.

### 2.7. AN ALGORITHM

The basic steps of our terrain algorithms are outlined briefly below.

1) Set $j = 1$, choose a starting point, $Z^0$, define the feasible region, set $\mathrm{C} = \{\ \}$ and initialize the set of stationary points, $\mathrm{S} = \{\ \}$.
2) Find an initial stationary point, $s_j^*$ of $\mathrm{g}^T\mathrm{g}$ and put $s_j^* \in S$. Set $k = 1$.

```
3) For $s_j^* \in S$
    3a) perform a partial eigendecomposition of M and H at $s_j^*$.
    3b) add to C only those dominant eigendirections, $c_i$, not previously explored.
4) Set $n_k = s_k^*$ and
    4a) explore each dominant eigenconnection $c_i \in C$ associated with $s_k^*$.
    4b) add each new stationary point found to $S$, delete $C_i$, set $j = j+1$ and go to 3.
    4c) for each boundary encountered, delete $c_i$, and go to 5.
5) Set $k = k+1$ and go to 4.
6) Repeat steps 4 and 5 for all $s_j^* \in S$ until C is empty.
```

Step 1 is an initialization step that also includes the initialization of any problem-solving parameters like a trust region radius, the maximum allowable function and gradient evaluations, convergence tolerance, etc. Step 2 finds an initial stationary point from an arbitrary starting point. Here downhill calculations using a trust region method are tried first. If no stationary point is found, then uphill calculations are used to look for a saddle point or singular point from the initial starting point. If both downhill and uphill calculations result in no stationary point, we conclude that there is no stationary point for the given problem. Step 3a calculates the necessary eigenvalues and eigenvectors of the important Hessian matrices. Step 3b adds only the important eigen-directions to the set of connections, C, that haven't been explored before. This prevents unnecessary back tracking. Step 4 identifies the $k$-th stationary point and in step 4a each of its associated eigen-connections is explored by moving either up or downhill, depending on the sign of the associated eigenvalue of either M and H (see section 2.2), using the appropriate tools (i.e., local equation solving, acceleration and/or predictor-corrector calculations) to accomplish the task. Step 4b increments the set of stationary points, deletes each eigen-direction associated with the $k$-th stationary point after it has been explored, and loops to step 3 to add any new eigen-directions associated with any new stationary point that has been found. In contrast, step 4c deletes any eigen-connection associated with the $k$-th stationary point if a boundary has been encountered. Step 4 and 5 constitute the main iterative loop for terrain following. Note that the sets S and C keep running totals of the number of nodes (or stationary points) and the unexplored branches (or connections), systematically exploring, adding and deleting branches for each stationary point as required. Step 6 provides the termination criterion for the algorithm.

## 3. Some Subtle Issues in Terrain-Following

In this section, some additional theoretical details for integral curve bifurcations and points of non-differentiability are discussed in the context of terrain following.

### 3.1. INTEGRAL CURVE BIFURCATIONS

We have encountered several problems that exhibit integral curve bifurcations. Tangent, pitchfork and other bifurcations in integral curves occur at points where $g^T H^T Hg$ is simultaneously a local minimum and a local maximum on some level curve of $g^T g$. These bifurcation points are easily characterized by the singularity of the projected Hessian matrix of the Lagrangian function associated with Equation (5).

Tangent bifurcations in integral curves are similar to the bifurcations that occur in parametric $S$-shaped solution curves and characterized by the merging or splitting of a single minimum and single maximum on some level curve of $g^T g$. This type of bifurcation corresponds to points where an integral curve becomes tangent to a level curve—something that is very easy to measure numerically because of the eigenvalue–eigenvector and corrector calculations that are part of the terrain methodology. That is, a tangent bifurcation is characterized by the fact that Hg no longer corresponds to the smallest positive or largest negative eigendirection. Tangent bifurcations also typically exhibit hysteresis. Lucia and Yang [10] illustrate tangent bifurcations in integral curves using the simple problem of finding all azeotropes for isopropyl alcohol and water.

Pitchfork bifurcations can also occur in integral curves. This type of bifurcation is usually the result of competition between neighboring stationary points and corresponds to a point where two minima and one maximum in $g^T H^T Hg$ occur simultaneously on some level curve of $g^T g$. Pitchfork bifurcations are also easily determined in a numerical setting since they correspond to a point where the projection of the Hessian matrix of the Lagrangian function onto the tangent subspace of the level constraint has a zero eigenvalue. Lucia and Yang [10] also give an illustration of pitchfork bifurcations in integral curves using the steady-state Lorenz equations.

### 3.2. NON-DIFFERENTIABILITIES

There are also applications in which there are points of non-differentiability. In process engineering, models of high-pressure vapor-liquid phase equilibrium using an equation of state often exhibit non-differentiabilities at compositions where there is a switch between vapor and liquid. Consider as an example high-pressure vapor-liquid equilibrium (VLE) at fixed temperature and pressure. As compositions pass through the boundaries of the two-phase region, the relative values of the hypothetical single vapor Gibbs free energy and the hypothetical single liquid Gibbs free energy will go through an exchange. That is, either the Gibbs free energy of the liquid will become lower than that of the vapor or the Gibbs free energy of the vapor will become lower than that of the liquid. Thus the overall Gibbs free energy function on which phase stability is based is a composite function formed by selecting the lower of the two hypothetical single phase Gibbs energies at all compositions. The exchange from vapor to liquid or vice versa that occurs at certain compositions results in a cusp or fold because of the distinct differences between vapor and liquid compressibilities at these compositions. However, these non-differentiable points present no difficulties for the terrain methodology because both the liquid and vapor Gibbs energy surfaces can be easily monitored and appropriate measures can be taken when a switch occurs. Similar remarks apply to VLLE and VLLLE as well as any situation that involves a switch or exchange of models for different regions of the feasible region.

## 4. Numerical Examples

In this section, phase split calculations and molecular modeling examples are used to illustrate the numerical performance of terrain methods. For these numerical experiments, the initial trust region radius, $R$, was 1. The convergence tolerance for all stationary points of $g^T g$, $\varepsilon$, was set to $1 \times 10^{-8}$, the trigger for quadratic acceleration, $\zeta$, was $1 \times 10^{-6}$ and the convergence tolerance for Kuhn-Tucker conditions for the individual NLP problems used in the corrector calculations (i.e., Equation (5)) was $\varepsilon^{1/2}$. Furthermore the value of $\Theta$ in Equation (4), which is used to invoke corrector steps, was 5. All calculations were performed in double precision arithmetic on a PC equipped with a Pentium III processor and a Lahey F77/90L-EM32 compiler.

### 4.1. PHASE SPLIT CALCULATIONS

Phase split calculations are often used to provide good initial values for macroscopic multiphase equilibrium calculations by Gibbs free energy minimization. Here we are generally interested in finding all minima in the tangent plane distance function because combinations of these minima define candidate phases and corresponding phase compositions that are useful in determining the global minimum Gibbs free energy for a given feed mixture at specified temperature and pressure. See, for example, Michelsen [12]. In fact, providing assurance that all minima of the tangent plane distance function can be found significantly strengthens the robustness of multiphase equilibrium codes based on tangent plane analysis. Moreover, for kinetic reasons we might also be interested in finding all saddle points since these saddle points represent energy barriers to macroscopic phase transitions.

Consider a ternary mixture of aniline, heptane and water at 313.15 K and $1.013 \times 10^6$ Pa. It is well known that this mixture exhibits three liquid phases at this temperature and pressure. For this illustration, the UNIQUAC equation described in Prausnitz et al. [14] was used to model liquid phase activity coefficients. The binary interaction parameters for this mixture are shown in Table 1.

Figure 1 shows the dimensionless Gibbs free energy of mixing projected onto the mass balance constraint (see Equation 10) for a single hypothetical liquid

*Table 1.* UNIQUAC binary interaction parameters for aniline-heptane-water

| | Aniline | Heptane | Water |
|---|---|---|---|
| Aniline | | 2.00 | 133.07 |
| Heptane | 288.92 | | 1149.7 |
| Water | 134.34 | 618.90 | |

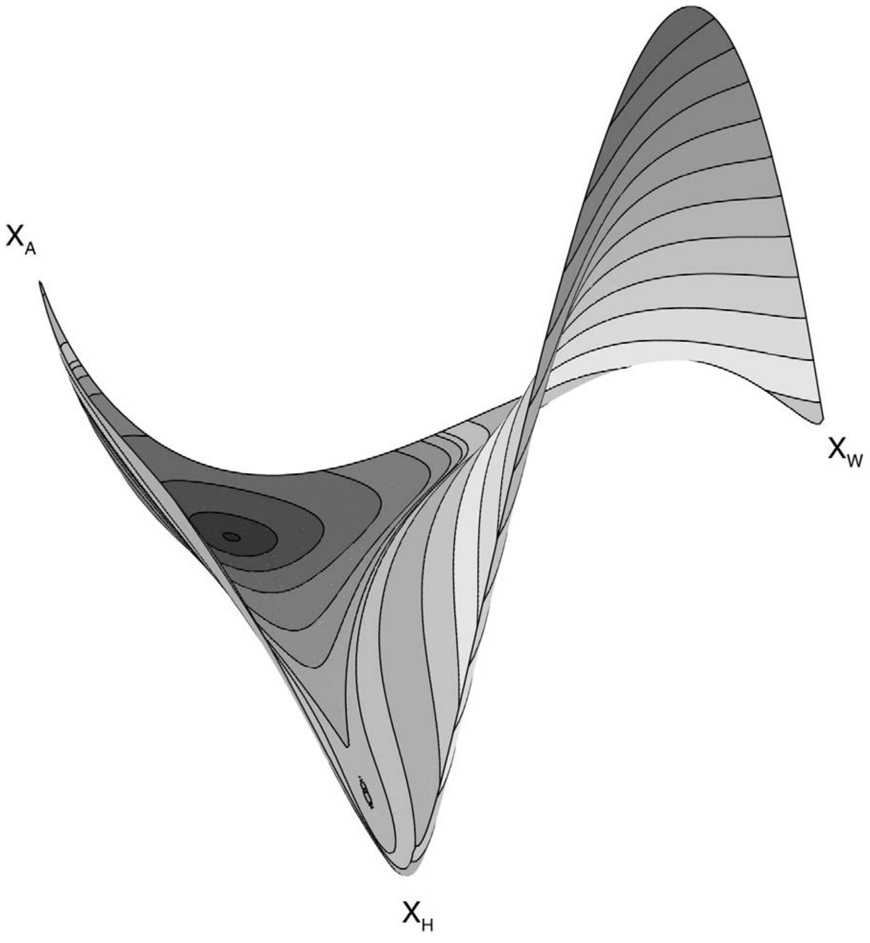

*Figure 1.* Gibbs energy of mixing for aniline-heptane-water at 313.15 K and 1.013 bar.

phase for this mixture at 313.15 K and $1.013 \times 10^6$ Pa. Note that the surface has three minima and two saddle points inside the feasible composition region. There are also six minima and three saddles on the boundaries of the feasible region. Therefore for any feed that contains nonzero amounts of all three components we are interested in finding the three minima (and perhaps the two saddle points) strictly inside the boundaries of the feasible region. Mathematically this problem

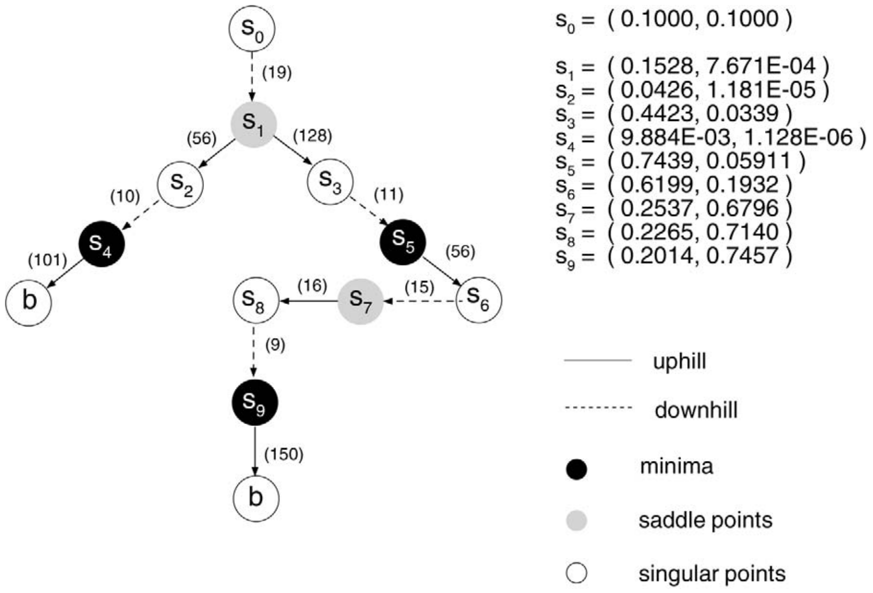

*Figure 2.* Computational tree for phase split calculation.

can be represented in the form

$$\min \Delta G/RT = \Sigma x_i(\ln x_i + \ln \gamma_i) \tag{9}$$

$$\text{subject to } \Sigma x_i - 1 = 0 \tag{10}$$

$$0 < x_i < 1 \tag{11}$$

where $x_i$ and $\gamma_i$ denote the composition and activity coefficient for the $i$th component and $\Delta G/RT$ denote the dimensionless Gibbs free energy of mixing.

Figure 2 shows the numerical results for this problem from a starting point of $x_0 = (0.1, 0.1, 0.8)$ reported in the form of a computational tree, where the component order is aniline, heptane, water. In this figure, dark gray, light gray and white nodes denote minima, saddle points and singular points respectively. The starting point and stationary points are reported using only the aniline and heptane compositions for brevity. Dashed branches denote downhill movement, solid branches denote uphill movement, and the number of function and gradient evaluations from stationary point to stationary point are shown in parenthesis along side the appropriate branch. Boundary points are also denoted by white nodes. Note that the terrain algorithm finds all five stationary points on the dimensionless Gibbs energy of mixing surface as well as four singular points in 571 function and gradient evaluations and 0.11 s of computer time. We also note that the principle eigenvalues and eigenvectors of M gave no conflicting information with regard to correctly identifying valleys or finding the next stationary point for this problem. Numerical details for these computations are quite lengthy and are available from the authors on request.

Now we turn to multiphase equilibria by Gibbs free energy minimization defined by

$$\min \mathrm{G}/RT = \Sigma\Sigma n_i^k(\mathrm{G}_i^{0k}/RT + \ln x_i^k + \ln \gamma_i^k) \tag{12}$$

$$\text{subject to } f_i - \Sigma n_i^k = 0, i = 1, \ldots, n_\mathrm{c} \tag{13}$$

$$0 < n_i^k < f_i \tag{14}$$

where $n_i^k$ is the molar flow for the $i$th component in the $k$th phase, $\mathrm{G}_i^{0k}$ is the $i$th component standard state Gibbs free energy, $x_i^k = n_i^k/\Sigma n_j^k$, $f_i$ is the feed molar flow of the $i$th component and $n_\mathrm{c}$ is the number of components in the mixture. Moreover, the double summation, $\Sigma\Sigma$, is over the number of phases, $n_p$, and number of components in that order.

From the minima shown in Figure 2 and elementary mass balance considerations, it is a relatively simple matter to determine the global minimum in the dimensionless Gibbs energy for any given feed mixture. For 1 mole of feed that contains 30 mol% aniline, 30 mol% heptane and 40 mol% water, the minima in Figure 2 and mass balance considerations suggest that the only appropriate choices of phases are either liquid–liquid equilibrium (LLE) or liquid–liquid–liquid equilibrium (LLLE). This is easily seen from Figure 1 by envisioning the possible points of double and triple-tangency that simultaneously pass through the given feed. Subsequent calculations give, in fact, four multiphase solutions—three liquid–liquid equilibria (LLE) and one liquid–liquid–liquid (LLLE) equilibrium. These solutions are shown in Table 2.

Solution 3, which required 13 function and gradient evaluations to find, is an unstable equilibrium and corresponds to a saddle point on the dimensionless Gibbs energy surface, G/RT, projected onto the plane defined by the component mass balances. Thus solution 3 can be ruled out immediately after it is computed. The two meta-stable liquid-liquid equilibrium solutions, on the other hand, were located in 9 and 12 function and gradient evaluations, have dimensionless Gibbs energies of $-0.100536$ and 0.090200 and are actually local constrained minima of G/RT. The

*Table 2.* Phase equilibria for aniline-heptane-water at 313.15 K and 1.013 × 106 Pa*

| | | | G/*RT* |
|---|---|---|---|
| 1 | LLE | $(2.00458\times10^{-3}, 2.49664\times10^{-7}, 0.299415)$<br>$(0.29799, 0.29999, 0.100585)$ | $-0.100536$ |
| 2 | LLE | $(0.27290, 1.39059\times10^{-2}, 0.38667)$<br>$(0.02710, 0.28609, 1.333\times10^{-2})$ | $-0.090200$ |
| 3 | LLE | $(0.19745, 0.29849, 7.70194\times10^{-2})$<br>$(0.10254, 1.5100\times10^{-3}, 0.32298)$ | $-0.059582$ |
| 4 | LLLE | $(2.99365\times10^{-2}, 0.271175, 1.03785\times10^{-2})$<br>$(1.68157\times10^{-3}, 2.03195\times10^{-7}, 0.274779)$<br>$(0.26838, 2.8825\times10^{-2}, 0.108425)$ | $-0.141625$ |

*Component molar flows in the order: (1) aniline; (2) heptane; (3) water.

intrinsically stable liquid–liquid–liquid equilibrium with a dimensionless Gibbs energy of $-0.141625$ required nine function and gradient evaluations to find. Each of these phase equilibrium computations takes less than 0.01 s of computer time. Clearly, the three-liquid phase solution is the global minimum of the Gibbs free energy for this feed at 313.15 K and $1.013 \times 10^6$ Pa. Note that the global minimum of the Gibbs free energy for any feed mixture can be obtained in a similar manner without re-calculating the minima in Figure 2.

Finally, we remark here that the Riedel equation and physical property data given in Prausnitz et al. [14] were used to calculate standard state Gibbs energies and that the numerical details of these Gibbs free energy minimization calculations are available from the authors by request.

### 4.2. MOLECULAR MODELING

One very important problem in molecular modeling is the determination of minima and saddle points of potential energy functions derived from either empirical force fields or ab initio quantum mechanical models. Results from these calculations can be used in a variety of ways. They can be used to help initialize molecular dynamic and Monte Carlo simulations, to find reaction pathways, or in transition state theory. Potential energy minimization can also be used for molecular conformational purposes (e.g., in protein folding).

Potential energy models derived from empirical force fields take the general form

$$E = E_b + E_{nb} \tag{15}$$

where $E_b$ represents the bonded or short-range energy effects such as bond lengths and bond and torsion angles and $E_{nb}$ represents the non-bonded or long-range energy effects from van der Waals and electrostatic forces. There are many, many models for bond length, bond angle, torsional, van der Waals and electrostatic effects. However, all of these energy effects can be expressed in terms of the Cartesian coordinates of the particles in the system. In this section we present two examples.

Consider the calculation of transition states and reaction pathways. One of the biggest disadvantages of some numerical techniques in molecular modeling like those that belong to the class of chain-of-states methods is that they require a priori knowledge of the coordinates of the reactants and products on the potential energy surface in order to locate transition states. That is, a global minimum (or product state) and at least one local minimum corresponding to a reactant state must be known ahead of time in order to calculate meaningful transition states (i.e., saddles) and the corresponding reaction pathway. In contrast, general methods like the $\alpha$BB method used by Westerberg and Floudas [17] and some methods specifically designed for molecular modeling such as Baker's method [3] do not require a priori knowledge of reactant and product states to find transition states. Terrain methods belong to the class of methods that do not require a priori knowledge of reactant and product states.

The first example is a slight modification of the classical Muller-Brown potential energy function [13], which can be viewed as something like a bond stretching Morse potential with cross terms to capture stretch-stretch bond length–bond angle interactions in a three-particle system with a fixed bond angle of 180°. The Muller-Brown potential energy function is, however, an empirical model and given by

$$E_{\mathrm{MB}} = \Sigma D_i \exp\{\mathrm{A}_i(r_{23} - x_{1i}{}^0)^2 + \mathrm{B}_i(r_{23} - x_{1i}{}^0)(r_{12} - x_{2i}{}^0) + \mathrm{C}_i(r_{12} - x_{2i}{}^0)^2\} \quad (16)$$

where the parameters $x_{1i}{}^0 = 3, 2, 1.5, 1$; $x_{2i}{}^0 = 1, 1.5, 2.5, 2$; $\mathrm{A}_i = -1, -1, -6.5, -0.7$; $\mathrm{B}_i = 0, 0, 11, 0.6$; $\mathrm{C}_i = -10, -10, -6.5, 0.7$; $\mathrm{D}_i = -200, -100, -170, 15$ for $i = 1, \ldots, 4$. All we have done here is modify the original constants of Muller and Brown so that the stationary points all lie in the positive orthant.

Figures 3 and 4 show the Muller-Brown function and the surface $g^T g$ for the Muller-Brown function for the feasible region $r_{23}$, $r_{12} \in [0, 3.5]$. Note that there are three minima and two saddle points on the surface $E_{MB}$ and 23 stationary

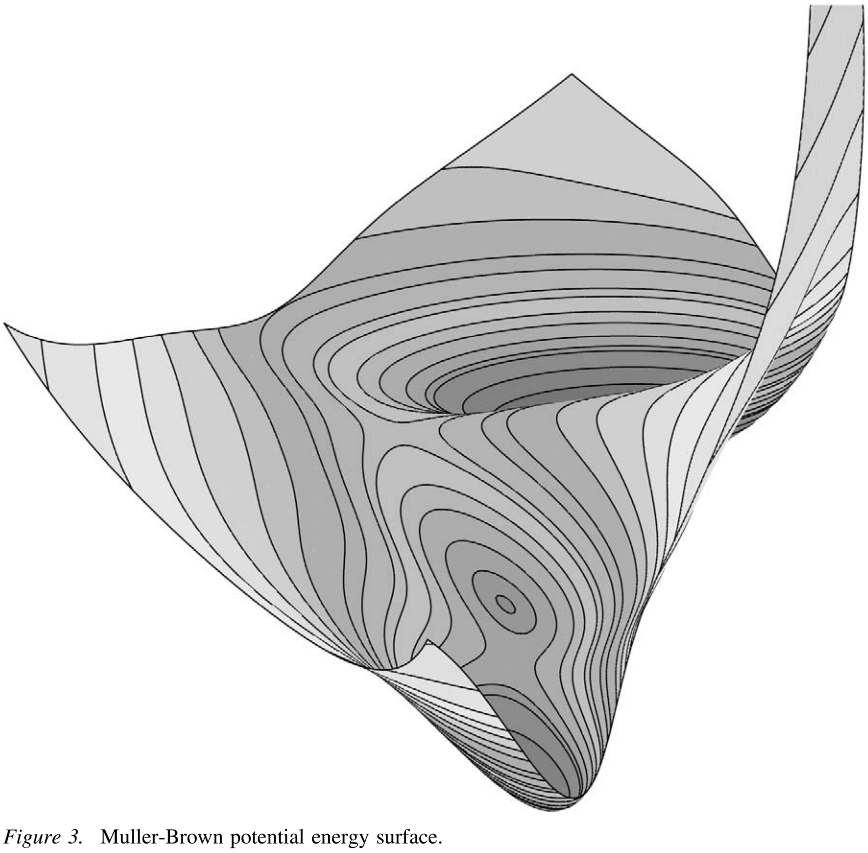

*Figure 3.* Muller-Brown potential energy surface.

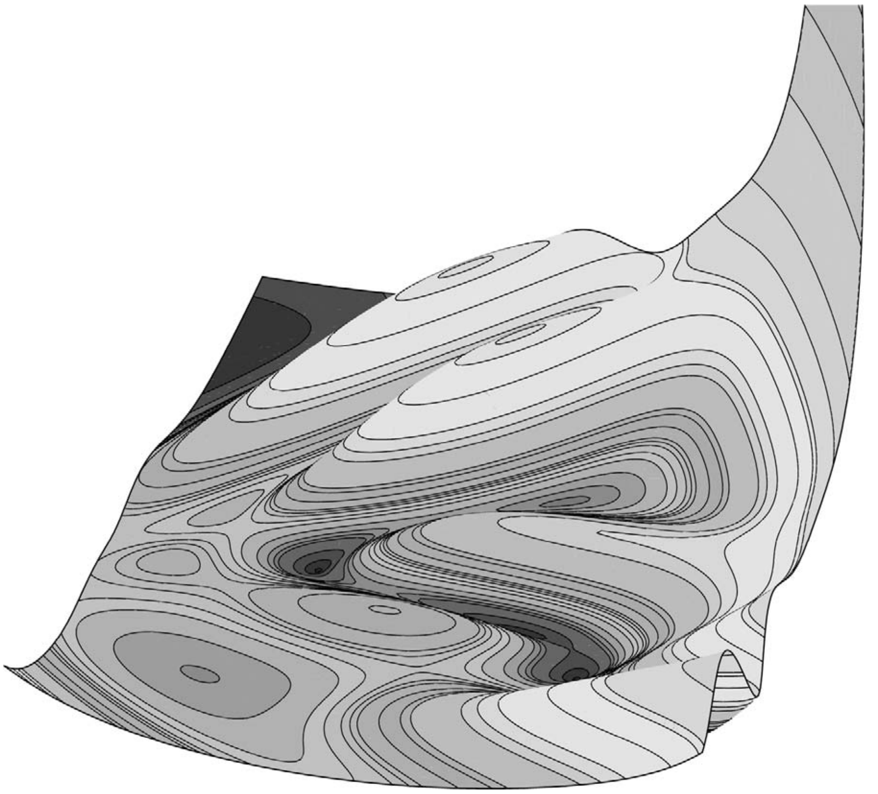

*Figure 4.* Gradient surface for the Muller-Brown function.

points on $g^T g$—nine minima, 10 saddle points and four maxima. The nine minima on $g^T g$ correspond to the three minima and two saddles on $E_{\mathrm{M}B}$ plus four other singular points of H, whereas the 10 saddles and four maxima on $g^T g$ are all singular points of H. Thus there are many singular points of H—18 in all—on the Muller-Brown potential energy surface. Nonetheless there are clear valleys on $E_{\mathrm{M}B}$ and $g^T g$ that contain stationary points of interest. Note, however, that the valleys of $g^T g$ are quite a bit more tortuous than the valleys of $E_{\mathrm{M}B}$.

One meaningful set of objectives that might be of interest here are the calculation of all minima and saddle points on $E_{\mathrm{M}B}$ and the determination of any reaction pathways from arbitrary starting points. We have, in fact, done this for a wide variety of starting points in the feasible region and have experienced no difficulties whatsoever in finding all stationary points on the Muller-Brown potential energy surface and uncovering the correct reaction pathway from any starting point. Figure 5 shows computational results for the starting point $(r_{23}, r_{12}) = (2.5, 1.4)$. In finding the 11 stationary points shown in Figure 5, 466 function and gradient evaluations and 0.06 s of computer time were required. Remember, for the purposes of the terrain following calculations, $g^T g$ is the primary objective function and $\phi = E_{\mathrm{M}B}$ is the secondary objective function so the results in Figure 5

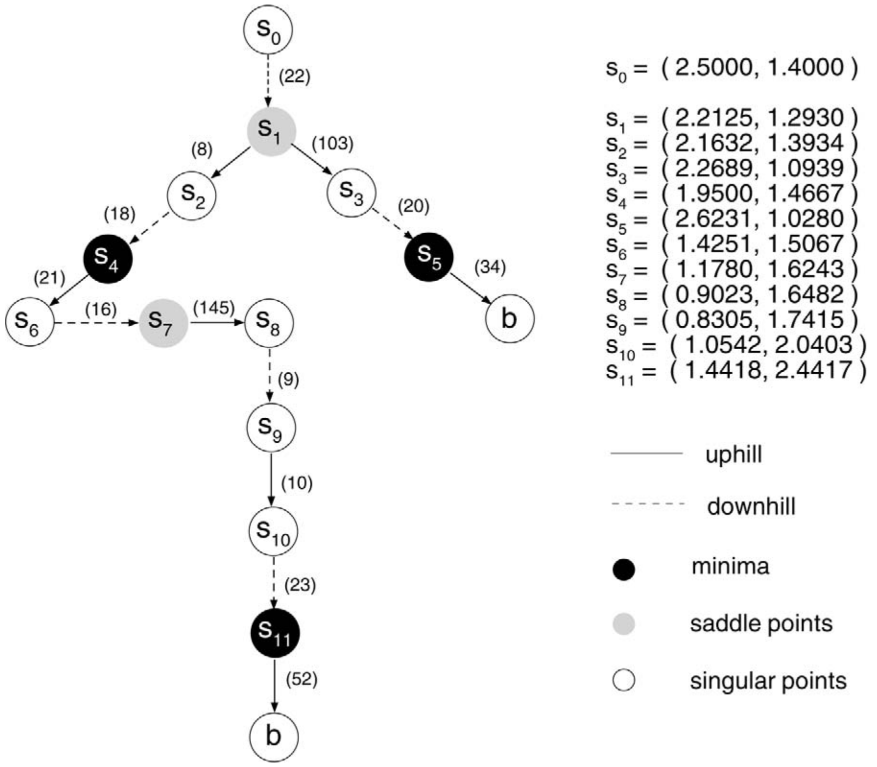

*Figure 5.* Computational tree for Muller-Brown potential.

also include the calculation of a number of singular points of H. Moreover, it is important to realize that the number of stationary points on $g^Tg$ that are found will depend on the starting point. In the case of the numerical results shown in Figure 5, our terrain method found 11 stationary points on $g^Tg$. For other starting points, more stationary points were found. For example from the starting point $(r_{23}, r_{12}) = (1.9, 2.2)$, 13 stationary points on $g^Tg$ were located in 580 function and gradient evaluations and 0.06 s of computer time. Note that because terrain methods follow integral curves, they automatically generate reaction pathway information on potential energy surfaces as a simple by-product of the calculations.

The Muller-Brown function also nicely illustrates the need for monitoring both the objective function and gradient surfaces in global optimization because of the sharp change in the orientation of the valley on $E_{MB}$ that occurs in the region containing the global minimum. Note that the portion of the valley containing the global minimum is essentially orthogonal to the portion of the valley containing all other stationary points on $E_{MB}$. During computations, difficulties arise at the stationary point $s_7 = (r_{23}, r_{12}) = (1.17800, 1.62431)$ in Figure 5. Eigenvalue–eigenvector calculations for the Hessian matrix of $g^Tg$ (i.e., the matrix M) give the smallest positive eigenvalue, $\lambda_M = 2.4003 \times 10^5$, and the associated normalized eigenvector, $c_M=(0.64829, 0.76140)$. However, this information is misleading. Here's why. Although this stationary point is a local minimum on $g^Tg$, it is a saddle point on $E_{MB}$; thus $g=0$, $M=H^TH$ and therefore $\lambda_M=\lambda_H^2$. However, the eigenvalues of H are 490.241 and −750.863.ecause $\lambda_M=\lambda_H^2$ and the positive eigenvalue of H is actually smaller in magnitude than the negative eigenvalue, if initial movement from this stationary point were based on eigenvector information from M, it would carry iterates uphill along a ridge on $E_{MB}$, which is wrong. In contrast, the largest negative eigenvalue and corresponding normalized eigendirection associated with of H are $\lambda_H=-750.863$ and $c_H=(-0.76140,\ 0.64829)$ and clearly indicate that the stationary point is a saddle. This eigen-information provides correct downhill movement toward the potential energy well containing the global minimum, keeping the terrain path in the right valley. All of this is accomplished rather easily and automatically with the aid of the test described in the second paragraph of Section 2.2.

The second example in the molecular modeling area is a Lennard-Jones fluid. The Lennard-Jones 6–12 potential energy function is a common pair-wise potential for estimating van der Waals or non-bonded forces between particles in an N-body system and typically has many stationary points, some of which are due to rotational symmetries. Moreover because of the number of unknown variables graphical representation of the potential energy surface is often not possible. The functionality of the Lennard-Jones potential is given by

$$E_{LJ}=\Sigma\Sigma\, 4\varepsilon_{ij}[(\sigma_{ij}/r_{ij})^{12}-(\sigma_{ij}/r_{ij})^{6}] \tag{17}$$

where the double summation is from $i=1,\ldots,N-1$ and $j=i+1,\ldots,N$ respectively, $\varepsilon_{ij}$ is an energy parameter, $\sigma_{ij}$ is a distance parameter, and $r_{ij}$ is the separation between particle $i$ and particle $j$ given by

$$r_{ij}=[(x_i-x_j)^2+(y_i-y_j)^2+(z_i-z_j)^2]^{1/2} \tag{18}$$

where $x_i$, $y_i$ and $z_i$ denote the Cartesian coordinates of the $i$th particle. There are also implicit constraints that are needed to avoid translational and rotational singularities. For any three-dimension configuration with three or more particles, there are six constraints and thus $3N$–6 degrees of freedom. These constraints are linear, most easily written in the form

$$x_1=y_1=z_1=0;\ y_2=z_2=0 \text{ and } z_3=0 \tag{19}$$

and are easy to handle. Note that these constraints arbitrarily tether the first particle to establish an origin and avoid translation of the entire system while the constraints on particles 2 and 3 prevent bond angle and torsion angle singularities.

The particular system studied is a classical system involving 100 Argon particles, where $\varepsilon_{ij}=1.66\times10^{-21}$ J and $\sigma_{ij}=3.4\times10^{-10}$ m. Thus there are 294 unknown particle coordinates from which bond angles and torsion angles are easily computed. The application of the terrain methodology to this problem resulted in the location of 73 stationary points in 3933 function and gradient calls and 3.05 s of computer time. The lowest value of the energy function obtained was $-2.73 \times 10^{-19}$ J. A number of molecular dynamics (MD) simulations were also run for the same exact N-body system from a variety of initial particle positions and velocities to verify that this was in fact the global minimum. For each MD simulation, we observed the appropriate Hamiltonian conservation with equilibrated trajectories that ‘orbited’ the potential energy well corresponding to $-2.73 \times 10^{-19}$ J. Again numerical results are available from the authors by request.

## 5. Conclusions

A geometric terrain methodology for the global optimization of general $C^3$ objective functions was presented. Novel features of this methodology include a rigorous characterization of valleys and ridges, the simultaneous use of two surfaces to guide exploration and attention to integral curve bifurcations and non-differentiabilities. A small collection of numerical examples was presented in this paper that show that the proposed terrain methodology represents a reliable and efficient way of finding minima, saddle points, singular points, changes in convexity and addressing other related goals in global optimization. However the problems presented here represent only a small fraction of the problem solving experience we have with our terrain methodology. We have in fact solved a large number of benchmark and engineering examples including reactor problems, flash and distillations, equations of state, data regression problems, phase transitions and polymer thermodynamics examples. Other recent work by Lucia and Yang [10] also shows that this methodology can be used to solve problems that have parametrically disconnected solutions, like reactor problems where some solutions lie on isola and other solutions lie on a disconnected S-shaped branch. All in all, we believe that the proposed terrain methodology is capable of solving global optimization problems in a reliable and efficient manner using only a modest amount of computer time.

Finally, we close with a remark that we have made in an earlier paper regarding numerical aspects of the NLP sub-problems that arise from our rigorous characterization of valleys and ridges (i.e., Equation (8). These constrained optimization problems can be quite ill conditioned and require a very reliable NLP algorithm to solve. We have tried readily available successive quadratic programming (SQP) software with very little success and presently use our own NLP solver for this task. Thus, it is important for anyone interested in constructing their own terrain method from various software components that considerable care be given to the robustness of the individual components (i.e., local equation solver, acceleration techniques, NLP solver, eigen routines, etc.).

## Acknowledgement

The authors would like to thank the National Science Foundation for support of this work under Grant No. CTS-0113091.

## References

1. Aluffi-Pentini, F., Parisi, V. and Zirilli, F. (1985), Global optimization and stochastic differential equations, *J. Opt. Theory and Appl.* 47, 1–15.
2. Bahren, J. and Protopopescu, V. (1996), Generalized TRUST algorithms for global optimization, *In State of the Art in Global Optimization*, Kluwer Academic Publishers, Dordrecht, pp. 163–180.
3. Baker, J. (1986), An algorithm for the location of transition states, *J. Comput. Chem.* 7, 385–395.
4. Bilbro, G.L. (1994), Fast stochastic global optimization, *IEEE Trans. Syst. Man. Cyber. SMC-24* 4, 684–689.
5. Bolhius, P.G., Chandler, D., Dellago, C. and Geissler, P.L. (2002), Transition path sampling: Throwing ropes over rough mountain passes, in the dark, *Annu. Rev. Phys. Chem.* 53, 291–318.
6. Hansen, E.R. (1980), Global optimization using interval analysis—the multidimensional case, *Numer. Math.* 34, 247–270.
7. Henkelman, G., Johannesson, G. and Jonsson, H. (2000), Methods for finding saddle points and minimum energy paths, In: *Progress in Theoretical Chemistry and Physics*, 5, pp. 269–300, Kluwer Academic Publishers, Dordrecht, The Netherlands.
8. Levy, A.V. and Montalvo, A. (1985), The tunneling algorithm for the global minimization of functions, *SIAM J. Sci. and Stat. Comp.* 6, 15–29.
9. Lucia, A. and Yang, F. (2002), Global terrain methods, *Comput. Chem. Engng.* 26, 529–546.
10. Lucia, A. and Yang, F. (2003), Multivariable terrain methods, *AIChE J.* in press.
11. Maranas, C.D. and Floudas, C.A. (1995), Finding all solutions to nonlinearly constrained systems of equations, *J. Global Optim.* 7, 143–153.
12. Michelsen, M.L. (1982), The isothermal flash problem. Part 1. stability, *Fluid Phase Equil.* 9, 1–19.
13. Muller, K. and Brown, L.D. (1979), Location of saddle points and minimum energy paths by a constrained simplex optimization procedure, *Theoret. Chim. Acta*, 53, 75–93.
14. Prausnitz, J.M., Anderson, T.F., Grens, E.A., Eckert, C.A., Hsieh, R. and O’Connell, J.P. (1980), *Computer Calculations for Multicomponent Vapor-Liquid and Liquid-Liquid Equilibria,* Prentice-Hall, Englewood Cliffs, NJ.
15. Schnepper, C.A. and Stadtherr, M.A. (1995), Robust process simulation using interval methods, *Comput. Chem. Engng.* 20, 187–199.
16. Sevick, E.M., Bell, A.T. and Theodorou, D.N. (1993), A chain of states method for investigating infrequent event processes occurring in multistate, multidimensional systems, *J. Chem. Phys.* 98, 3196–3212.
17. Westerberg, K.M. and Floudas, C.A. (1999), Locating all transition states and studying the reaction pathways of potential energy functions, *J. Chem. Phys.* 110, 9259–9295.