Potential Energy Surfaces and Optimization Methods
| 8 Ethane | C | 0.000000 | 0.000000 | 0.000000 | 1 | 2 | 3 | 4 | 5 |
| 1 C | 1 | 0.000000 | 0.000000 | 1.530000 | 1 | 1 | 6 | 7 | 8 |
| 2 H | 1.037098 | 0.000000 | -0.366645 | 5 | 1 | 1 | 6 | 7 | 8 |
| 3 H | -0.518502 | 0.898180 | -0.366645 | 5 | 1 | 1 | 6 | 7 | 8 |
| 4 H | -0.518502 | -0.898180 | -0.366645 | 5 | 1 | 1 | 6 | 7 | 8 |
| 5 H | 1.033151 | 0.090389 | 1.896645 | 5 | 2 | 1 | 6 | 7 | 8 |
| 6 H | -0.438247 | -0.939953 | 1.896645 | 5 | 2 | 1 | 6 | 7 | 8 |
| 7 H | -0.594811 | 0.849572 | 1.896645 | 5 | 2 | 1 | 6 | 7 | 8 |

| 8 Ethane | C | 1 |
| 1 C | 1 | 1.530000 |
| 2 C | 1 | 1.100000 |
| 3 H | 1.100000 | 2 | 109.4700 |
| 4 H | 1.100000 | 2 | 109.4700 |
| 5 H | 1.100000 | 2 | 109.4700 |
| 6 H | 2 | 1.100000 |
| 7 H | 109.4700 | 3 | -120.0000 |
| 8 H | 3 | -125.0000 | 0 |

| 8 Ethane | C | 1 |
| 1 C | 1 | 1.530000 |
| 2 C | 1 | 1.100000 |
| 3 H | 1.100000 | 2 | 109.4700 |
| 4 H | 1.100000 | 2 | 109.4700 |
| 5 H | 1.100000 | 2 | 109.4700 |
| 6 H | 2 | 1.100000 |
| 7 H | 109.4700 | 3 | -120.0000 |
| 8 H | 3 | -125.0000 | 0 |
Fig. 5.1: Variation in the energy of pentane with the two torsion angles indicated and represented as a contour diagram and isometric plot. Only the lowest-energy regions are shown.
FIG. 3 Model potentials for protein motions. a, Two-dimensional representation of harmonic potential (left) and multiminum (substate) potential (right). b, Representation of the r.m.s. difference criterion for different minima: 1, r.m.s. after the minimization is larger than the initial r.m.s., implying that the two conformations correspond to different minima; 2, r.m.s. after the minimization is smaller than the initial r.m.s., implying that the two conformations correspond to the same minimum (from ref. 13).
BOLTZMANN'S DISTRIBUTION

Two minima corresponding to states A and B

Area gives probability of state A or B

- Probability of system being at position $x$ is
  $$ P(x) = \frac{\exp(-U(x)/kT)}{Q} $$
  $U(x)$ is Potential Energy at position $x$.

- Find $Q$, the "Partition Function", so total probability is 1.
  $$ Q = \sum \exp(-U(x)/kT) $$
The actual transition from State A to B is very quick (a few picoseconds).

What takes time is the waiting. The average wait before going from A to B is:

$$
\tau_{A \rightarrow B} = \frac{h}{k_b T} \exp \left[ \frac{-\Delta G}{k_b T} \right], \text{ where } \Delta G = (G_T - G_A)
$$

$$
\frac{h}{k_b T} \sim 0.16 \text{ picoseconds at } T = 300^\circ K (27^\circ C)
$$

$h$ is Planck's constant, $k_b$ is Boltzmann's constant.
OPTIMIZATION: Local vs. Global

**LOCAL**
- Pattern Search
- Steepest Descent
- Newton's Method
- Variable Metric

**GLOBAL**
- Trajectory/Deterministic
- BREMMERMAN'S METHOD
- CLUSTERING (TORN)
- Controlled Random Search
- Stochastic Methods
Figure 1  Types of minima.
Types of stationary points.
Fig. 5.4: The three basic moves permitted to the simplex algorithm (reflection, and its close relation reflect-and-expand; contract in one dimension and contract around the lowest point). (Figure adapted from Press W H, B P Flannery, S A Teukolsky and W T Vetterling 1992. Numerical Recipes in Fortran. Cambridge, Cambridge University Press.)
The simplex search technique.
One at a time search.
Davies, Swann, and Campey's search.

1. Input data \( x_{0,1}, \delta x, \epsilon, K, f \). Set \( k = 0 \).
2. Set \( k = k + 1 \).
   Evaluate \( f \) at \( x_{0,k} \) and \( x_{1,k} \) to give \( E_{0,k} = f(x_{0,k}) \) and \( E_{1,k} = f(x_{1,k}) \).
3. Test for positive search direction.
   If \( E_{1,k} < E_{0,k} \), set \( p = +1 \) and go to Step 4.
   Otherwise go to Step 4.
4. Test for negative search direction.
   Evaluate \( f \) at \( x_{-1,k} \).
   If \( E_{1,k} < E_{0,k} \), set \( p = -1 \) and go to Step 5.
   Otherwise go to Step 8.
5. Search until the minimum is spanned.
   Either search the positive direction (for \( p = +1 \)),
   or search the negative direction (for \( p = -1 \)), until
   \( E_{0,k} > E_{-1,k} \).
6. Insert the extra function evaluation at the centre of the final interval.
   Evaluate \( f \) at \( x_{0,k} - 2^{-(k+1)} \delta x_k \).
7. Determine the minimum of the fitted quadratic.
   If \( E_{0,k} > E_{-1,k} \) use equation 2.33 to give \( x_{0,k+1} \).
   Otherwise use equation 2.34.
   If \( 2^{-(k+2)} \delta x_k \leq \epsilon \) go to Step 9.
   Otherwise, set \( \delta x_{k+1} = K \delta x_k \) and go to Step 2.
8. Determine the minimum of the quadratic fitted to
   \( x_{-1,k}, x_{0,k}, \) and \( x_{1,k} \).
   Set \( \delta x_{k+1} = \frac{\delta x_k (E_{1,k} - E_{-1,k})}{2(E_{1,k} - 2E_{0,k} + E_{-1,k})} \).
   If \( \delta x_k \leq \epsilon \) go to Step 9.
   Otherwise, set \( \delta x_{k+1} = K \delta x_k \) and go to Step 2.
9. Output data, either \( \delta x_k \) or \( 2^{-(k+1)} \delta x_k \).
   \( x_{\min} = x_{0,k+1} \).
   \( E_{\min} = E_{0,k+1} \).
Figure 9  A one-dimensional line search for $\tilde{f}(\lambda) = f(x + \lambda p)$ by cubic interpolation produces an approximate minimum along $p$, at $\lambda^*, f(\lambda^*)$. 
Figure 10: Possible situations in line search algorithms between the current and trial points: (a) The new slope is positive. (b) The new slope is negative but function value is greater. (c) The new slope is negative and function value is lower.
Steepest descent minimization.
Search by the method of conjugate directions.
Figure 12  Steepest descent and conjugate gradient quantities that affect the convergence rate for quadratic functions (see text for the distinct context of these functions).
Algorithm [A1]: Basic Descent

* Supply an initial guess \( x_0 \).
* For \( k = 0, 1, 2, \ldots \), until convergence
  1. Test \( x_k \) for convergence.
  2. Calculate a search direction \( p_k \).
  3. Determine an appropriate step length \( \lambda_k \) (or modified step \( s_k \)).
  4. Set \( x_{k+1} \) to \( x_k + \lambda_k p_k \) (or \( x_k + s_k \)).

Algorithm [A4]: Modified Newton

* For \( k = 0, 1, 2, \ldots \), until convergence, given \( x_0 \),
  1. Test \( x_k \) for convergence.
  2. Compute a descent direction \( p_k \) so that
     \[
     \|H_k p_k + g_k\| \leq \eta_k \|g_k\|,
     \]
     where \( \eta_k \) controls the accuracy of the solution and some symmetric matrix \( H \) may approximate \( H_k \).
  3. Compute a step length \( \lambda \) so that for \( x_{k+1} = x_k + \lambda p_k \),
     \[
     f(x_{k+1}) \leq f(x_k) + \alpha g_k^T p_k,
     \]
     \[
     |g_{k+1}^T p_k| \leq \beta |g_k^T p_k|,
     \]
     with \( 0 < \alpha < \beta < 1 \).
  4. Set \( x_{k+1} = x_k + \lambda p_k \).
Algorithm [A2]: CG Method to Solve $Ax = -b$

1. Set $r_0 = -(Ax_0 + b)$, $d_0 = r_0$.
2. For $k = 0, 1, 2, \ldots$, until $r$ is sufficiently small, compute

$$
\lambda_k = r_k^T r_k / d_k^T A d_k
$$

$$
x_{k+1} = x_k + \lambda_k d_k
$$

$$
r_{k+1} = r_k - \lambda_k A d_k
$$

$$
\beta_k = r_{k+1}^T r_{k+1} / r_k^T r_k
$$

$$
d_{k+1} = r_{k+1} + \beta_k d_k.
$$

Algorithm [A3]: PCG Method to Solve $Ax = -b$

1. Set $r_0 = -(Ax_0 + b)$, $d_0 = M^{-1} r_0$.
2. For $k = 0, 1, 2, \ldots$, until $r$ is sufficiently small, compute

$$
\lambda_k = r_k^T (M^{-1} r_k) / d_k^T A d_k
$$

$$
x_{k+1} = x_k + \lambda_k d_k
$$

$$
r_{k+1} = r_k - \lambda_k A d_k
$$

$$
\beta_k = r_{k+1}^T (M^{-1} r_{k+1}) / r_k^T (M^{-1} r_k)
$$

$$
d_{k+1} = (M^{-1} r_{k+1}) + \beta_k d_k.
$$
Figure 2  Sample matrix patterns for (a) block diagonal and (b–e) sparse unstructured. Pattern (b) corresponds to the Hessian approximation (preconditioner) for a potential energy model from the local energy terms (bond length, bond angle, and dihedral angle terms), and (c) is a reordered matrix pattern that reduces fill-in during the factorization. Pattern (d) comes from a molecular dynamics simulation of supercoiled DNA\textsuperscript{36} and describes pairs of points along a ribbonlike model of the duplex that come in close contact during the dynamics trajectory; pattern (e) is the associated reordered structure that reduces fill-in.
Numerical Example I: Rosenbrock Minimization

Rosenbrock's function is often used as a minimization test problem, because its minimum lies at the base of a "banana-shaped valley" and can be difficult to locate. This function is defined for even integers \( n \) as the sum

\[
f(x) = \sum_{i=1,3,5,...,n-1} [(1 - x_i)^2 + 100(x_{i+1} - x_i^2)^2].
\]

The contour plot of Rosenbrock's function for \( n = 2 \) is shown in Figure 14. The minimum point is \((1,1)\), where \( f(x) = 0 \). The gradient components of this function are given by

\[
\begin{align*}
g_{i+1} &= 200(x_{i+1} + x_i^2) \\
g_i &= -2[x_i g_{i+1} + (1 - x_i)] \\
\end{align*}
\]

and the Hessian is the \( 2 \times 2 \) block diagonal matrix with entries

\[
\begin{align*}
H_{i+1,i+1} &= 200 \\
H_{i+1,j} &= -400x_j \\
H_{ii} &= -2(x_i H_{i+1,j} + g_{i+1} - 1) \\
\end{align*}
\]

\( i, j = 1, 3, 5, ..., n - 1 \).
Fig. 1. Surface plot of the Rosenbrock function, $n = 2$. 
steepest-descent algorithm
conjugate-gradient algorithm
BFGS quasi-Newton algorithm
modified Newton algorithm
<table>
<thead>
<tr>
<th>Tasks per iteration</th>
<th>Classic Newton</th>
<th>Nonlinear Conjugate Gradient</th>
<th>Quasi-Newton (Full Memory)</th>
<th>Quasi-Newton (Limited Memory)</th>
<th>Truncated Newton</th>
</tr>
</thead>
<tbody>
<tr>
<td>Form g iteration</td>
<td>an</td>
<td>an</td>
<td>an</td>
<td>an</td>
<td>m(αn)</td>
</tr>
<tr>
<td>Form H</td>
<td>(α/2)n(n + 1)</td>
<td>an</td>
<td>an</td>
<td>an</td>
<td>O(l) − O(l^2)</td>
</tr>
<tr>
<td>Form M</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Φ(αn + 7n + O(l))</td>
</tr>
<tr>
<td>Factor M</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>&lt; 4an</td>
</tr>
<tr>
<td>Calculate p</td>
<td>6n^3</td>
<td>7n</td>
<td>O(n^2)</td>
<td>23n</td>
<td>O(n^3)</td>
</tr>
<tr>
<td>Perform line search</td>
<td></td>
<td>~ (2−4)αn</td>
<td>an</td>
<td>an</td>
<td></td>
</tr>
<tr>
<td>Storage</td>
<td>O(n^2)</td>
<td>3n−7n</td>
<td>O(n^2)</td>
<td>14n</td>
<td>O(max[n, m])</td>
</tr>
<tr>
<td>Advantages</td>
<td>Locally quadratically convergent</td>
<td>Easy to implement Modest storage and computation cost</td>
<td>Locally superlinearly convergent</td>
<td>Adapts well to available storage</td>
<td>Exploits problem structure to accelerate convergence (by preconditioning)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Can display local quadratic convergence</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Requires construction and factorization of preconditioner</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Performance may be slow for highly nonlinear functions when directions of negative curvature are detected repeatedly</td>
</tr>
<tr>
<td>Disadvantages</td>
<td>Requires O(n^2) storage and O(n^3) work</td>
<td>Convergence often slow Requires fairly accurate line search</td>
<td>Requires O(n^2) storage</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Fig. 5. The minimum-energy reaction path corresponds to an imaginary relaxation path of steepest descent, along which the molecular system glides with almost infinitely slow motion from a saddle point to an adjacent minimum.
Lagrange Multipliers

Lagrange multipliers can be used to find the stationary points of functions, subject to a set of constraints. Suppose we wish to find the stationary points of a function \( f(x, y) = 4x^2 + 3x + 2y^2 + 6y \) subject to the constraint \( y = 4x + 2 \). In the Lagrange method the constraint is written in the form \( g(x, y) = 0 \):

\[
g(x, y) = y - 4x - 2 = 0anumber{(1.40)}
\]

To find stationary points \( f(x, y) \) subject to \( g(x, y) = 0 \) we first determine the total derivative \( df \), which is set equal to zero:

\[
df = \frac{\partial f}{\partial x} dx + \frac{\partial f}{\partial y} dy = (8x + 3) dx + (4y + 6) dy = 0\]

(1.41)

Without the constraint the stationary points would be determined by setting the two partial derivatives \( \frac{\partial f}{\partial x} \) and \( \frac{\partial f}{\partial y} \) equal to zero, as \( x \) and \( y \) are independent. With the constraint, \( x \) and \( y \) are no longer independent but are related via the derivative of the constraint function \( g \):

\[
dg = \frac{\partial g}{\partial x} dx + \frac{\partial g}{\partial y} dy = -4dx + dy = 0
\]

(1.42)

The derivative of the constraint function, \( dg \), is multiplied by a parameter \( \lambda \) (the Lagrange multiplier) and added to the total derivative \( df \):

\[
\left( \frac{\partial f}{\partial x} + \lambda \frac{\partial g}{\partial x} \right) dx + \left( \frac{\partial f}{\partial y} + \lambda \frac{\partial g}{\partial y} \right) dy = 0
\]

(1.43)

The value of the Lagrange multiplier is obtained by setting each of the terms in parentheses to zero. Thus for our example we have:

\[
8x + 3 - 4\lambda = 0
\]

(1.44)

\[
4y + 6 + \lambda = 0
\]

(1.45)

From these two equations we can obtain a further equation linking \( x \) and \( y \):

\[
\lambda = 2x + 3/4 = -6 - 4y \quad \text{or} \quad x = -27/8 - 2y
\]

(1.46)

Combining this with the constraint equation enables us to identify the stationary point, which is at \((-59/72, -23/18)\).

This simple example could, of course, have been solved by simply substituting the constraint equation into the original function, to give a function of just one of the variables. However, in many cases this is not possible. The Lagrange multiplier method provides a powerful approach which is widely applicable to problems involving constraints such as in constraint dynamics (Section 7.5) and in quantum mechanics.
Fig. 6. Application of the reaction coordinate method to the two-parametric model potential.
Fig. 14. Localization of a saddle point on the two-parametric model potential surface by means of ascending valley points.
Fig. 5.30: The conjugate peak refinement method. $r$ and $p$ are the initial minima (the reactants and the products). A coarse step search along the line connecting $r$ and $p$ suggests that there is a maximum near the point $y_1^0$. Minimisation along the line connecting $r$ to $p$ gives the point $y_1^1$. A line minimisation is then performed along the conjugate vector to give the point $x_1$. In the second iteration the procedure is repeated for the lines $r - x_1$ and $x_1 - p$. A maximum is found at $y_1^2$, which after minimisation along the conjugate vector gives the saddle point $s_1$. Subsequent iterations of the algorithm enable the second saddle point $s_2$ to be identified. (Figure adapted in part from Fischer S and M Karplus 1992. Conjugate Peak Refinement: An Algorithm for Finding Reaction Paths and Accurate Transition States in Systems with Many Degrees of Freedom. Chemical Physical Letters 194:252–261.)
Fig. 5.27: Too large a step size may lead to the wrong saddle point or an inefficient algorithm.
Global Optimization Techniques

- Deterministic Methods
  - Space covering (grid search)
  - Systematic search
  - Trajectory & generalized descent
  - Smoothing / deformation

- Stochastic Methods
  - Multi–start from many trial points
  - Bayesian statistical models
  - Clustering techniques
  - Monte Carlo & Simulated Annealing

- Heuristic Methods from Chemistry
  - Distance geometry
  - Scheraga’s build–up procedure
  - Saunders’ random kick
  - Still’s torsional tree–search
GENERALIZED DESCENT GLOBAL OPTIMIZATION
Diffusion Equation Method

For a function $f(x)$, define $f^{(N)}(x)$ as the sum of $N$ even-power derivatives

$$f^{(N)}(x) = \left(1 + \beta \frac{d^2}{dx^2}\right)^N f(x)$$

Let

$$\beta = \frac{t}{N}$$

and evaluate the limit of the infinite expansion

$$F(x,t) = \lim_{N \to \infty} \left(1 + \frac{t}{N} \frac{d^2}{dx^2}\right)^N f(x)$$

$$= \exp\left(t \frac{d^2}{dx^2}\right) f(x)$$

where $F(x,t)$ satisfies the diffusion equation:

$$\frac{\partial^2 F}{\partial x^2} = \frac{\partial F}{\partial t}$$