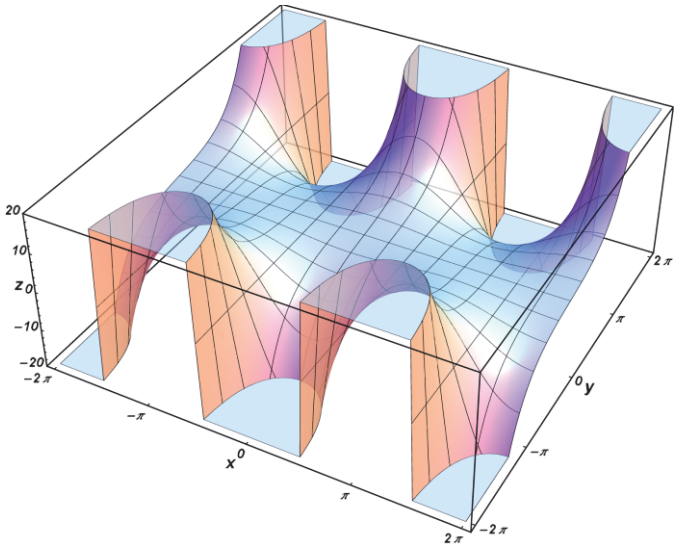



*"My dear fellow, who will let you?"*

*"That's not the point. The point is, who will stop me?"*

Ayn Rand

# Lecture 6: molecular geometry optimization

*energy minima, transition states, simplex, steepest descent, conjugate gradients, Newton, quasi-Newton, BFGS, TRM, RFO*

Dr Ilya Kuprov, University of Southampton, 2012

*(for all lecture notes and video records see http://spindynamics.org)*

# Molecular geometry optimization

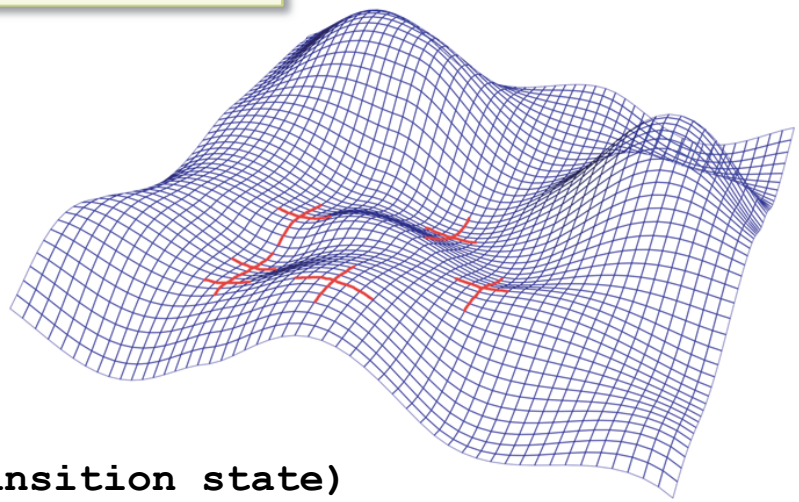
Within the Born-Oppenheimer approximation, the geometry of a molecule at zero absolute temperature corresponds to the minimum of the total energy:

$$\{x_i, y_i, z_i\}_{T=0} = \underset{\{x_i, y_i, z_i\}}{\arg\min}\left(\langle \Psi | \hat{H} | \Psi \rangle\right)$$

The process of finding the coordinates that minimize the energy is known in the trade as *geometry optimization*. It is usually the first step in the calculation.

We usually look for zero gradient (a *gradient* is the steepest ascent direction), but not all such points are minima:

$$(\nabla E)_i = \frac{\partial E}{\partial x_i}; \qquad H_{ij} = \frac{\partial^2 E}{\partial x_i \partial x_j}$$



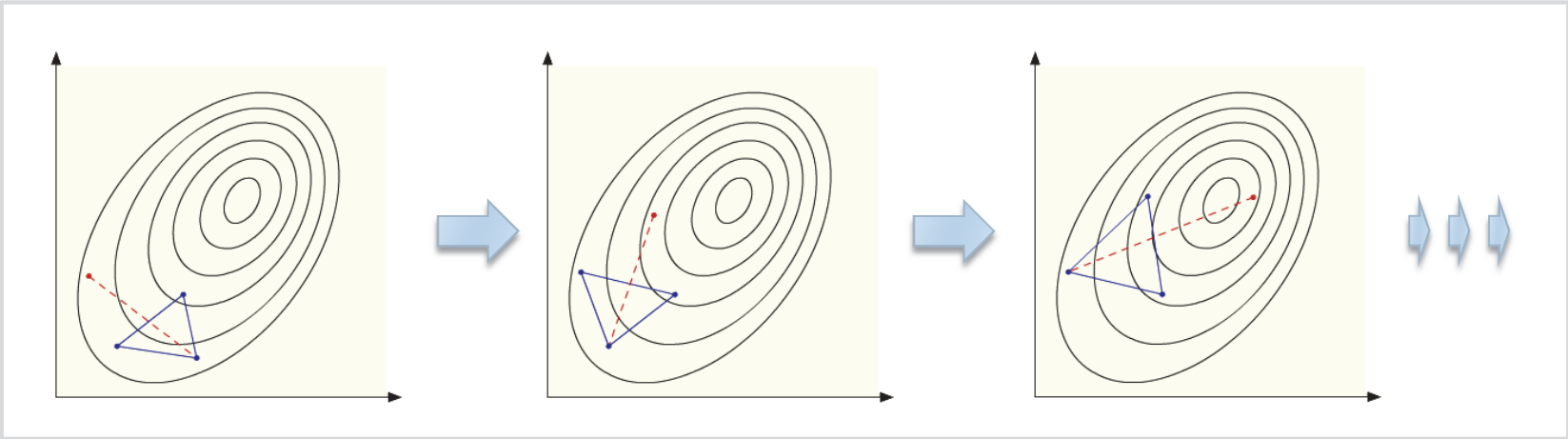
If the gradient is zero and

All **H** eigenvalues positive => **Minimum**
All **H** eigenvalues negative => **Maximum**
**H** eigenvalues of both signs => **Saddle (transition state)**

J.B. Foresman, A. Frisch, *Exploring Chemistry with Electronic Structure Methods*, Gaussian, 1996.

# Minimization methods: *simplex*

*Nelder-Mead simplex* method performs minimization by polygon node reflection, accompanied by expansion, contraction and line search where necessary.



| Pro | Contra |
|---|---|
| Does not require gradient information | Very slow – thousands of iterations even with simple problems. |
| Works for discontinuous, noisy and irreproducible functions. | Often gets stuck on ridges and in troughs. |
| Can jump out of a local minimum. | No guarantee of convergence. |

Simplex and other non-gradient methods are only used in desperate cases when gradient information is not available and when the function is very noisy or discontinuous.

J.A. Nelder and R. Mead, *Computer Journal*, 1965 (7) 308-313.

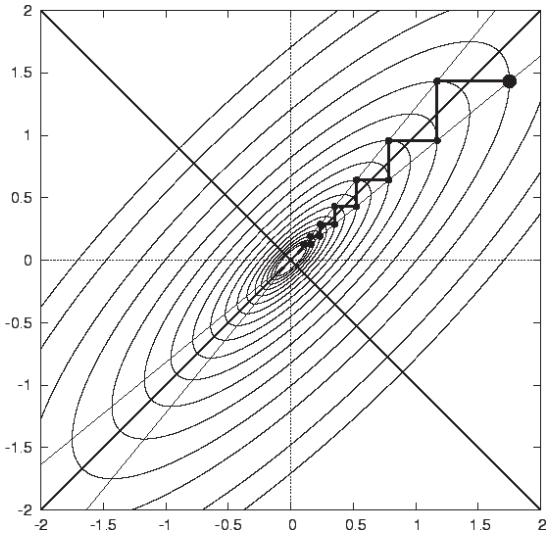
# Minimization methods: *gradient descent*

The basic *gradient descent* method repeatedly takes a fixed step in the direction opposite to the direction of the local gradient of the objective function:

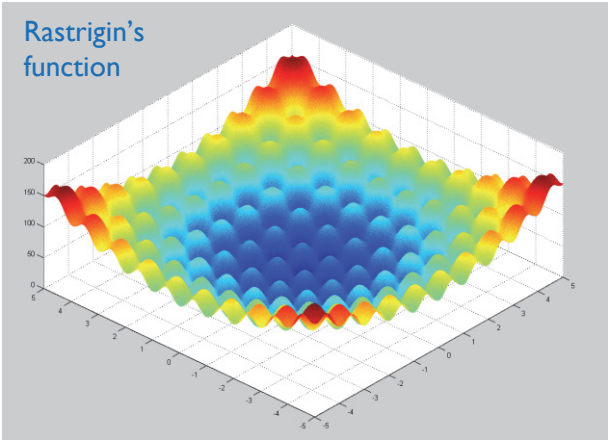
$$\vec{x}_{n+1} = \vec{x}_n - k_n \nabla f\left(\vec{x}_n\right)$$

The step size may be adjusted at each iteration using the *line search* procedure:

$$k_n = \arg\min_k \left( f\left(\vec{x}_n - k\nabla f\left(\vec{x}_n\right)\right)\right)$$




Characteristic zig-zagging pattern produced by the basic gradient descent method.




| Pro | Contra |
|---|---|
| Convergence is guaranteed. | Becomes slow close to the minimum. |
| Easy to implement and understand. | Not applicable to noisy functions |
| Relatively inexpensive. | Always converges to the nearest local minimum. |

J.A. Snyman, *Practical Mathematical Optimization*, Springer, 2005.

# Minimization methods: *conjugate gradients*

*Conjugate gradients* method makes use of the gradient history to decide a better direction for the next step:

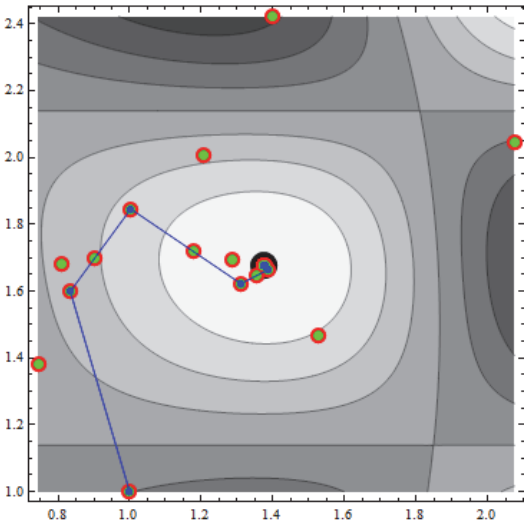
$$\vec{x}_{n+1} = \vec{x}_n - k_n \vec{h}_n \qquad \vec{h}_n = \nabla f\left(\vec{x}_n\right) + \gamma_n \vec{h}_{n-1}$$

Fletcher-Reeves method:

$$\gamma_n = \frac{\left|\nabla f\left(x_n\right)\right|^2}{\left|\nabla f\left(x_{n-1}\right)\right|^2}$$

Polak-Ribiere method:

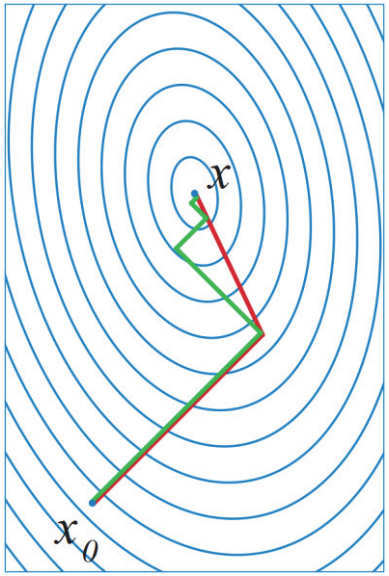
$$\gamma_n = \frac{\left(\nabla f\left(x_n\right) - \nabla f\left(x_{n-1}\right)\right)^{\mathrm{T}} \nabla f\left(x_n\right)}{\left|\nabla f\left(x_{n-1}\right)\right|^2}$$

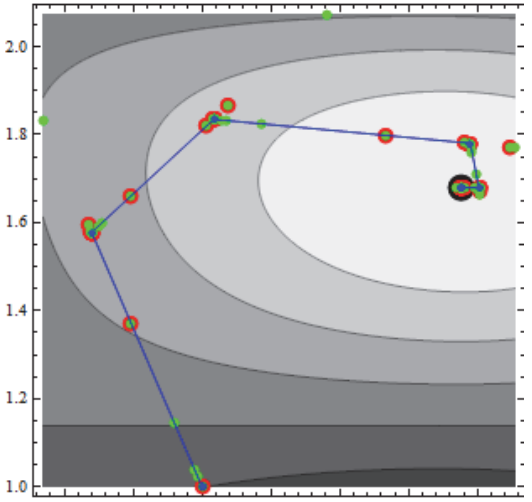


Conjugate gradients: 22 evaluations.

CG method does not exhibit the zig-zagging behaviour during convergence, but still tends to be quite slow in very non-linear cases.

CG is very useful for solving linear systems of equations, but has been superseded by more sophisticated quadratic step control methods for non-linear optimization.



Gradient descent: 63 evaluations.

J.A. Snyman, *Practical Mathematical Optimization*, Springer, 2005.

# Minimization methods: *Newton-Raphson method*

Newton-Raphson method approximates the objective function by a quadratic surface at each step and moves to the minimum of that surface:

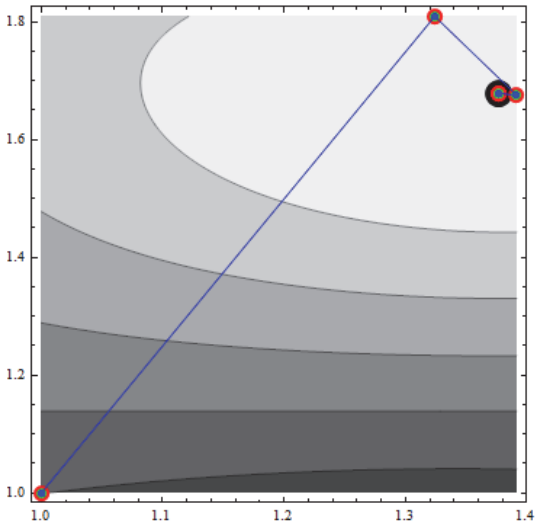
$$f\left(\vec{x}+\Delta\vec{x}\right) \approx f(\vec{x}) + \nabla f(\vec{x})^{\mathrm{T}} \Delta\vec{x} + \frac{1}{2}\Delta\vec{x}^{\mathrm{T}}\mathbf{H}\Delta\vec{x}$$

$$\nabla f\left(\vec{x}+\Delta\vec{x}\right) \approx \nabla f(\vec{x}) + \mathbf{H}\Delta\vec{x}$$

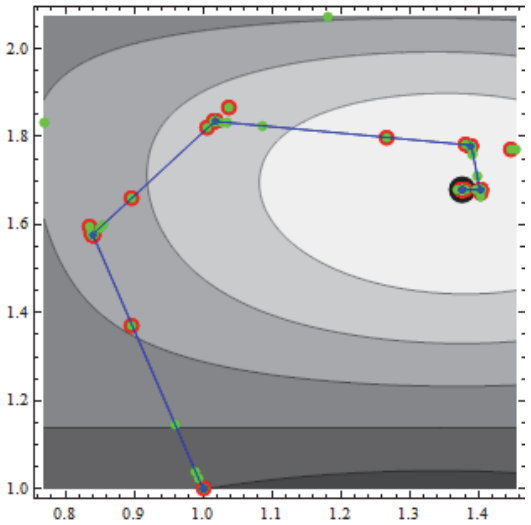
$$\Delta\vec{x} = -\mathbf{H}^{-1}\nabla f(\vec{x})$$

Variations of Newton-Raphson method are currently the primary geometry optimization algorithms in QC software packages.

| Pro | Contra |
|---|---|
| Very fast convergence. | Very expensive. |
| Well adapted for molecular geometry optimization. | Not applicable to noisy functions. |
| Useful if Hessians are cheap (*e.g.* in molecular dynamics). | Always converges to the nearest local *stationary point*. |



Newton-Raphson: 6 evaluations.



Gradient descent: 63 evaluations.

J. Nocedal, S.J. Wright, *Numerical Optimization*, Springer, 2006.

# Minimization methods: *quasi-Newton methods*

The most expensive part of Newton-Raphson method is the Hessian. It turns out that a good *approximate* Hessian may be extracted from the gradient history:

$$\mathbf{H}_{k+1} = \left( E - \frac{\vec{g}_k \vec{s}_k^T}{\vec{g}_k^T \vec{s}_k} \right) \mathbf{H}_k \left( E - \frac{\vec{s}_k \vec{g}_k^T}{\vec{g}_k^T \vec{s}_k} \right) + \frac{\vec{g}_k \vec{g}_k^T}{\vec{g}_k^T \vec{s}_k}$$

DFP update
(Davidon-Fletcher-Powell)

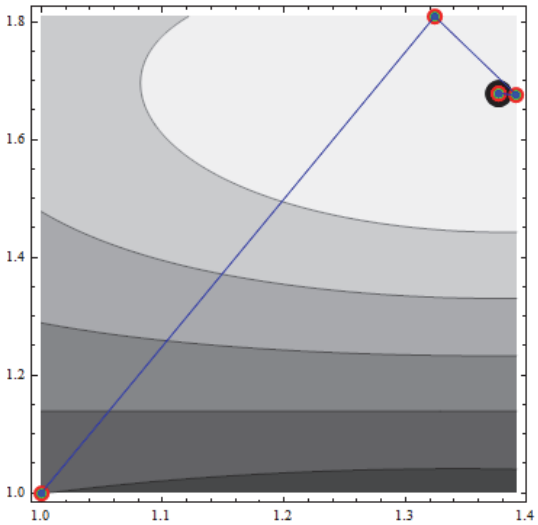
$$\mathbf{H}_{k+1} = \mathbf{H}_k + \frac{\vec{g}_k \vec{g}_k^T}{\vec{g}_k^T \vec{s}_k} - \frac{\mathbf{H}_k \vec{s}_k (\mathbf{H}_k \vec{s}_k)^T}{\vec{s}_k^T \mathbf{H}_k \vec{s}_k}$$

BFGS update
(Broyden-Fletcher-Goldfarb-Shanno)

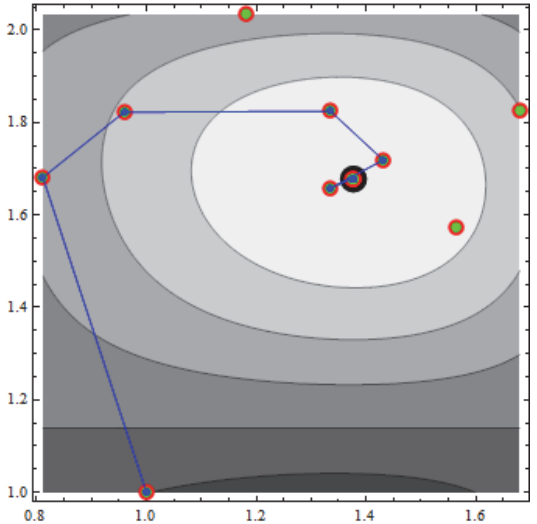
$$\vec{g}_k = \nabla f(x_{k+1}) - \nabla f(x_k), \qquad \vec{s}_k = \vec{x}_{k+1} - \vec{x}_k$$

The BFGS method is particularly good. Quasi-Newton methods provide super-linear convergence at effectively the cost of the gradient descent method.

Initial estimates for the Hessian are often computed using inexpensive methods, such as molecular mechanics or semi-empirics (this is what the `connectivity` data is for in Gaussian).



Newton-Raphson: 6 evaluations.



BFGS quasi-Newton: 13 evaluations.

J. Nocedal, S.J. Wright, *Numerical Optimization*, Springer, 2006.

# Minimization methods: *TRM and RFO*

The basic Newton-Raphson method requires the Hessian to be non-singular and tends to develop problems if any of its eigenvalues become negative. A simple fix for this is to add a regularization matrix (often a unit matrix):

$$\Delta\vec{x} = -\left(\mathbf{H} + \lambda\mathbf{S}\right)^{-1} \nabla f(\vec{x})$$

The two ways of doing this are known as *trust region method* and *rational function optimization method*. Both are implemented in Gaussian09. For TRM we have:

$$\lambda > \max\left\{0, -\min\left[\mathrm{eig}\left(\mathrm{H}\right)\right]\right\}$$

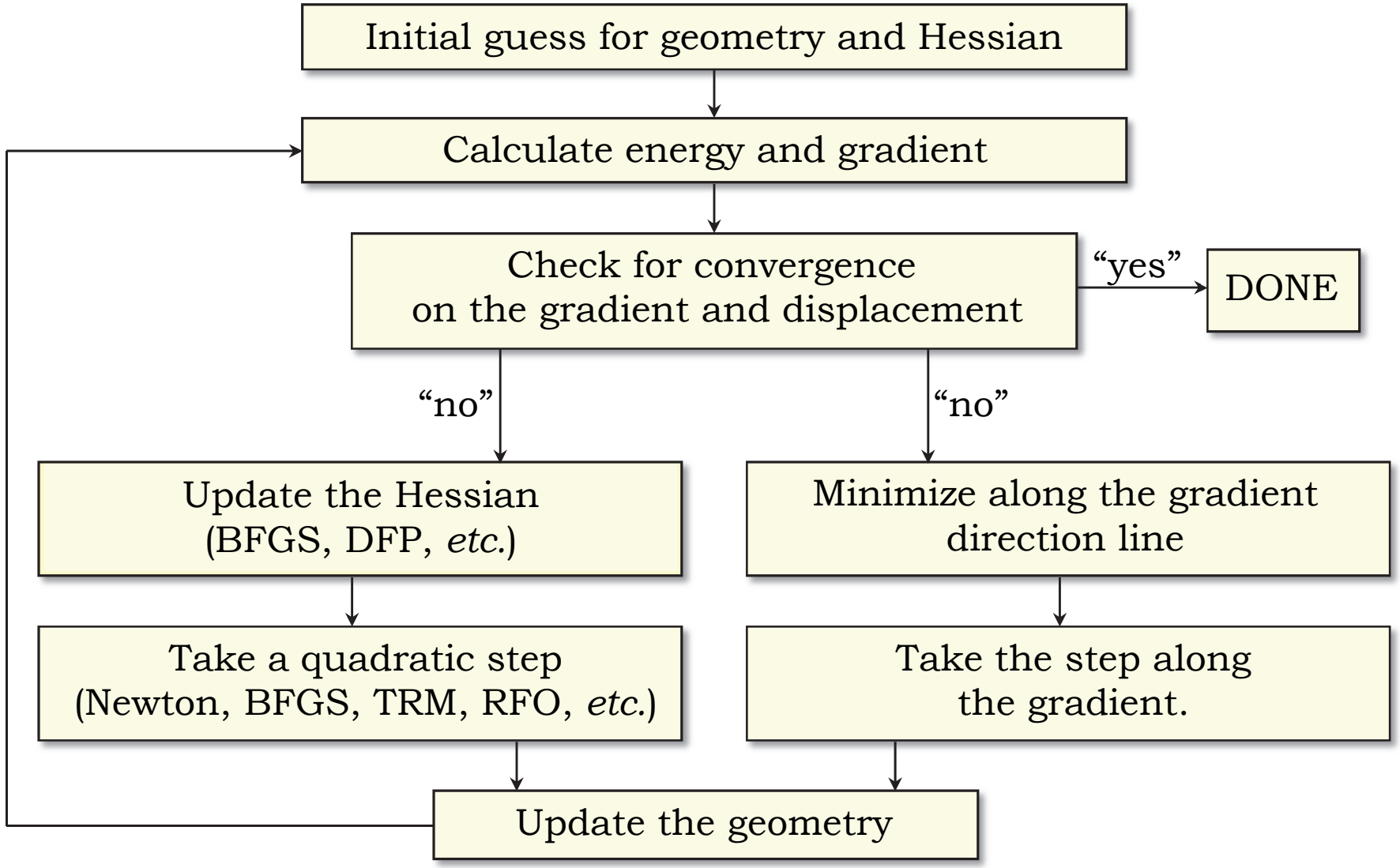
and the step is scaled to stay within the trust radius. Rational function optimization introduces a step size dependent denominator, which prevents the algorithm from taking large steps:

$$f\left(\vec{x}+\Delta\vec{x}\right) \approx f(\vec{x}) + \frac{\nabla f(\vec{x})^{\mathrm{T}}\Delta\vec{x} + \frac{1}{2}\Delta\vec{x}^{\mathrm{T}}\mathbf{H}\Delta\vec{x}}{1+\Delta\vec{x}^{\mathrm{T}}\mathbf{S}\Delta\vec{x}}$$

RFO behaves better than Newton-Raphson in the vicinity of inflection points.

R.H. Byrd, J.C. Gilbert, J. Nocedal, http://dx.doi.org/10.1007/PL00011391

# Geometry optimization flowchart



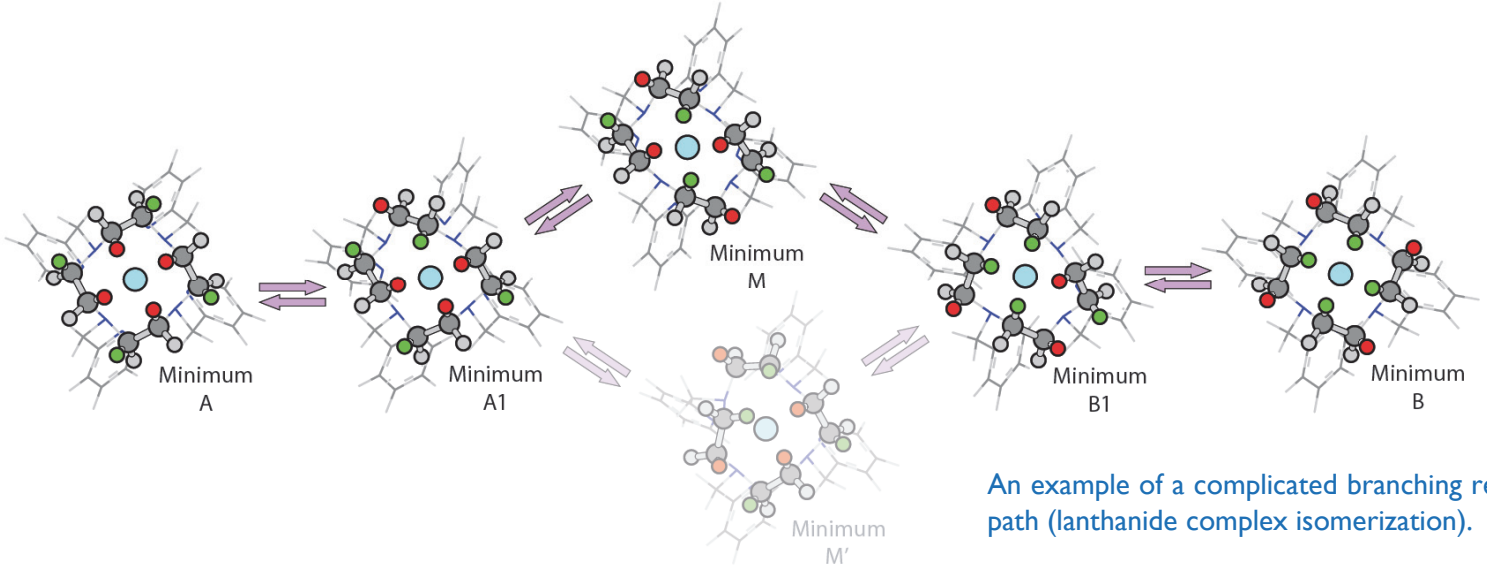

J.B. Foresman, A. Frisch, *Exploring Chemistry with Electronic Structure Methods*, Gaussian, 1996.

# Intrinsic reaction coordinate

IRC is defined as *the minimum-energy reaction path on a potential energy surface in mass-weighted coordinates, connecting reactants to products.*

$$\mathrm{IRC} = \left\{ \vec{x}(s) \left| \frac{\delta}{\delta \vec{x}(s)} \int_{s=0}^{s=1} E\left(\vec{x}(s)\right) ds = 0 \right. \right\}$$

Because the reactants and the products are energy minima, the IRC path necessarily passes through a saddle point, known as the *transition state*. Finding the transition state without system-specific knowledge is quite difficult (minima are usually easy).

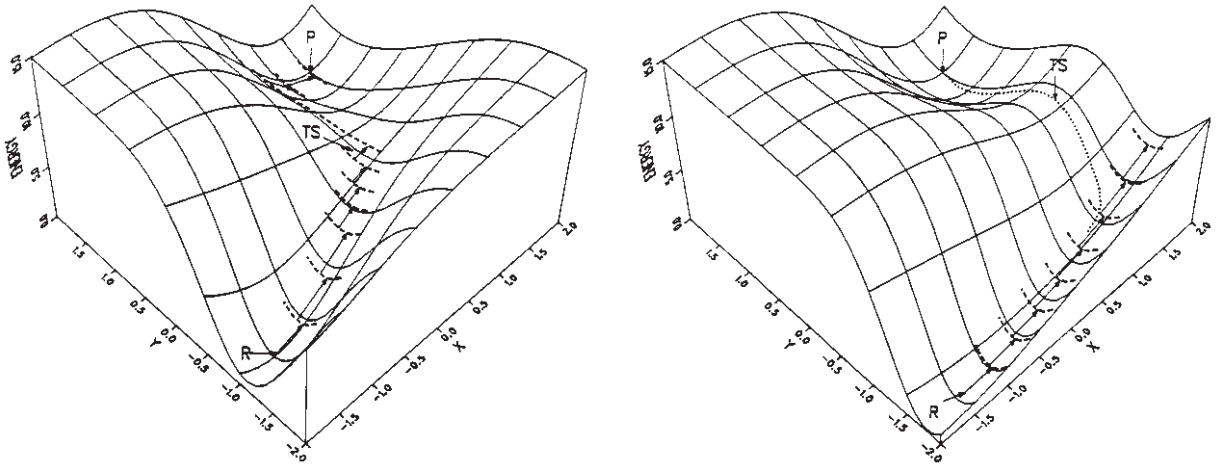



An example of a complicated branching reaction path (lanthanide complex isomerization).

F. Jensen, *Introduction to Computational Chemistry*, Wiley, 2007.

# Coordinate driving and hypersphere search

*Coordinate driving* is also known as *relaxed potential energy scan*. A particular internal coordinate (or a linear combination thereof) is systematically scanned and the structure is optimized to a minimum with respect to other coordinates at each step.



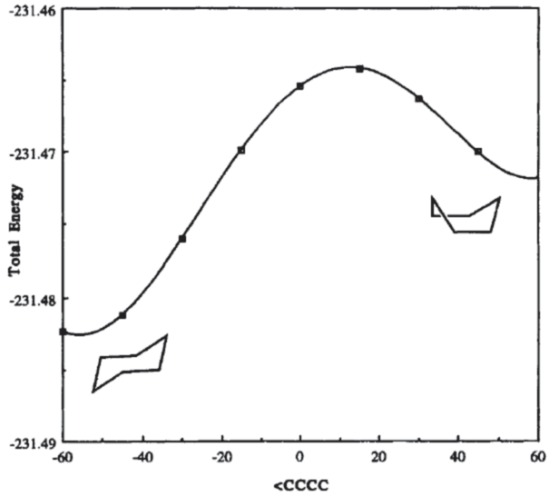
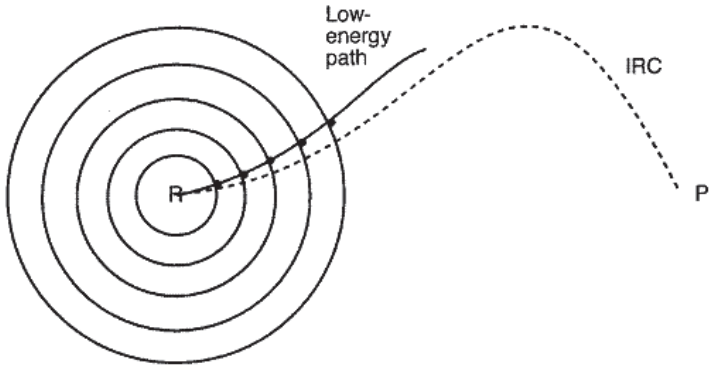




**FIGURE 1.** Relaxed potential energy surface scan of the conversion of chair cyclohexane to twist boat using redundant internal coordinates.

*Hypersphere search* proceeds by locating all energy minima on a hypersphere of a given radius in the coordinate space and tracing these minima as a function of the sphere radius.

Very expensive, but has the advantage of mapping *all* reaction paths from the given structure.

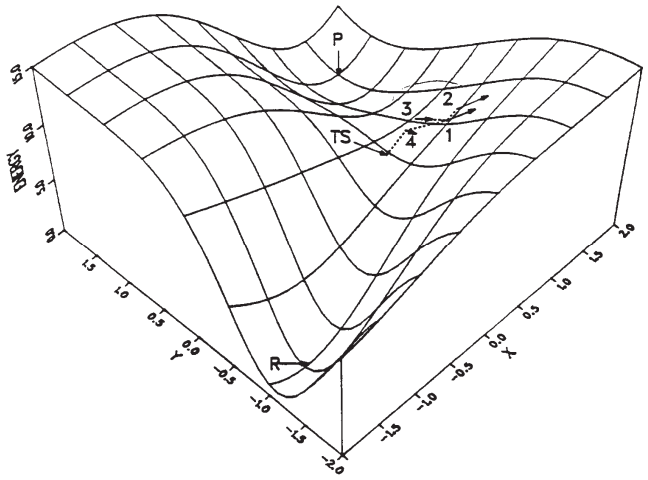



C. Peng, P.Y. Ayala, H.B. Schlegel, M.J. Frisch, *J. Comp. Chem.* 17 (1996) 49-56.

# Newton and quasi-Newton methods

The type of stationary point reached by quadratic algorithms depends on the definiteness of the Hessian. In the Hessian eigenframe we have:

$$\begin{pmatrix} \Delta x_1 \\ \Delta x_2 \\ \vdots \\ \Delta x_N \end{pmatrix} = -\begin{pmatrix} H_{11} & 0 & \cdots & 0 \\ 0 & H_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & H_{NN} \end{pmatrix} \begin{pmatrix} \nabla f_1 \\ \nabla f_2 \\ \vdots \\ \nabla f_N \end{pmatrix}$$




If any eigenvalue is negative, the step would be performed *not down, but up the gradient* and the corresponding coordinate would be optimized into a maximum. Importantly, for indefinite Hessians the BFGS quasi-Newton update scheme becomes ill-conditioned. It is commonly replaced by *Bofill update scheme*, which avoids that problem:

$$\Delta\mathbf{H}^{\mathrm{BOF}} = \varphi\Delta\mathbf{H}^{\mathrm{SR1}} + (1-\varphi)\Delta\mathbf{H}^{\mathrm{PSB}}, \qquad \Delta\mathbf{H}^{\mathrm{SR1}} = \frac{(\Delta g - \mathbf{H}\Delta x)(\Delta g - \mathbf{H}\Delta x)^{\mathrm{T}}}{(\Delta g - \mathbf{H}\Delta x)^{\mathrm{T}}\Delta x}$$

$$\Delta\mathbf{H}^{\mathrm{PSB}} = \frac{(\Delta g - \mathbf{H}\Delta x)\Delta x^{\mathrm{T}} + \Delta x(\Delta g - \mathbf{H}\Delta x)^{\mathrm{T}}}{\Delta x^{\mathrm{T}}\Delta x} - \frac{\Delta x^{\mathrm{T}}(\Delta g - \mathbf{H}\Delta x)\Delta x\Delta x^{\mathrm{T}}}{(\Delta x^{\mathrm{T}}\Delta x)^2}, \qquad \varphi = \frac{\left|\Delta x^{\mathrm{T}}(\Delta g - \mathbf{H}\Delta x)\right|^2}{|\Delta x|^2\,|\Delta g - \mathbf{H}\Delta x|^2}$$

K.P. Lawley, H.B. Schlegel, http://dx.doi.org/10.1002/9780470142936.ch4

# Eigenvector following method

In principle, one could follow a particular normal mode out of the minimum and into a maximum by deliberately stepping uphill on that mode:

1. Start at the minimum, compute and diagonalize the Hessian matrix.
2. Move in the direction of eigenvector of interest by a user-specified initial step.
3. Compute the gradient and the Hessian at the new point. Diagonalize the Hessian.
4. Take Newton steps in all directions except the direction of the smallest eigenvalue. Take an uphill step in that direction.
5. Repeat 2-4 until converged.

Advantages: very robust. Guarantees convergence into *some* saddle point. Good choice for small systems.

Downside: very expensive (a Hessian is computed at each step). No guarantee that the saddle point is the one wanted. Rarely a good choice for systems with 50+ coordinates.

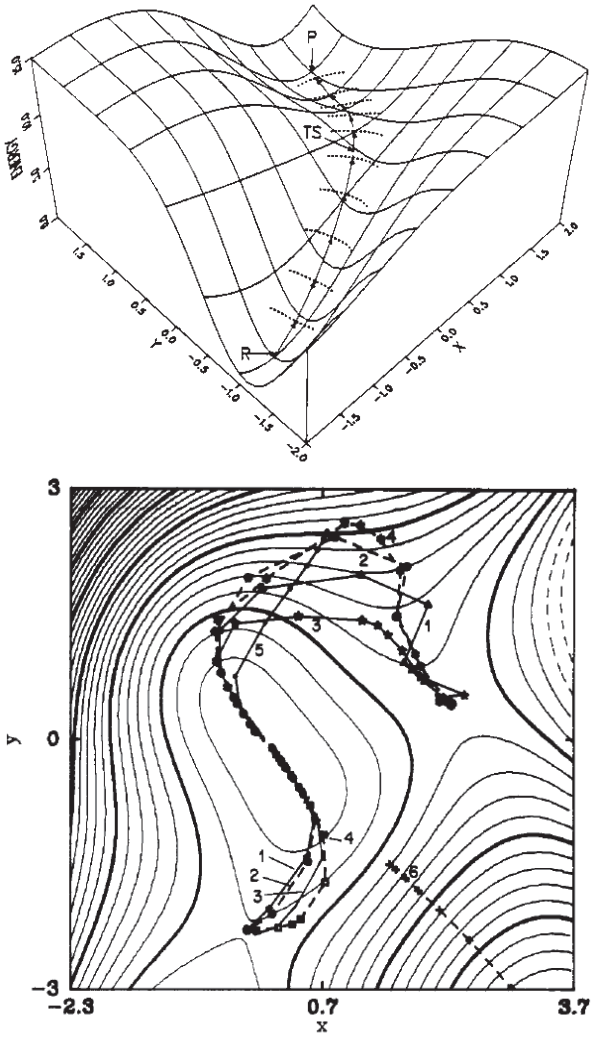



**Figure 4.** Walks on the Adams potential surface: (1–4) correspond to walks toward saddle points starting near the minimum at (0,0); (1) RFO; (2) P-RFO; (3) RFO with updated Hessian; (4) P-RFO with updated Hessian; (5) RFO walk to the minimum, (6) RFO walk toward the maximum.

A. Banerjee, N. Adams, J. Simons, R. Shepard, *J. Phys. Chem.* 89 (1985) 52-57.

# Gradient extremal following method

For every point on the *gradient extremal path* (GEP), the gradient is an eigenvector of the Hessian:

$$\mathrm{GEP} = \left\{ \vec{x} \,|\, H(\vec{x})\, g(\vec{x}) = \lambda(\vec{x})\, g(\vec{x}) \right\}$$

Equivalently, every point on a GEP is a minimum of the gradient modulus for a given energy:

$$\mathrm{GEP} = \left\{ \vec{x} \,|\, \frac{\partial}{\partial \vec{x}} \left( \vec{g}^{\mathrm{T}} g - \lambda \left( E(\vec{x}) - E_0 \right) \right) = 0 \right\}$$

GEPs are thus parameterized by energy and may be used for stationary point search. They have *no chemical meaning*, but they do connect the stationary points on the energy surface.

$$\vec{x}(s) = \vec{x}(0) - \left( \mathbf{I} - \vec{v}(s) \cdot \vec{v}(s)^{\mathrm{T}} \right) \mathbf{H}(s)^{-1}\, \vec{g}(s) + \sigma \vec{g}(0)$$

$$\mathbf{H}(0)\, \vec{g}(0) = \lambda(0)\, \vec{g}(0), \qquad \mathbf{H}(s)\, \vec{v}(s) = \lambda_{\vec{v}(s)}\, \vec{v}(s)$$

The update scheme may be formulated in terms of approximate as well as exact Hessians.

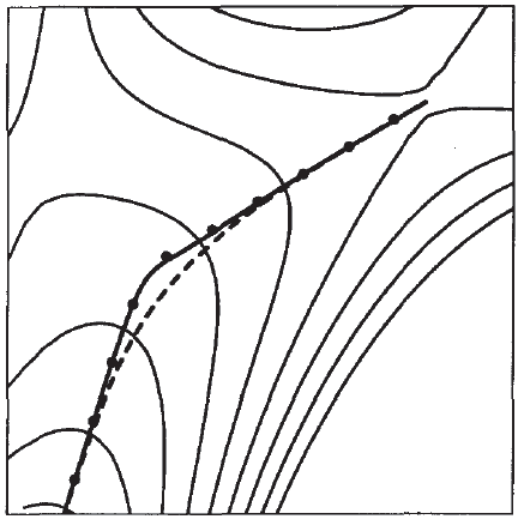


**Fig. 1.** Gradient extremal and minimum energy paths on the surface given by Eq. (17) ($-1 \leqslant x \leqslant 3$, $-2 \leqslant y \leqslant 2$). *Dashed line* – minimum energy path, *solid line* – gradient extremal path, *dots* – points on the gradient extremal path calculated using Eqs. (14) and (15)

*N.B.* Both MEPs and GEPs can bifurcate on high order stationary points. GEPs can also be quite convoluted on hilly energy surfaces.

F. Jensen, *Introduction to Computational Chemistry*, Wiley, 2007.

# Synchronous transit methods

Linear Synchronous Transit method searches for the energy maximum on the linear path (in internal coordinates) connecting the reactants and the products:

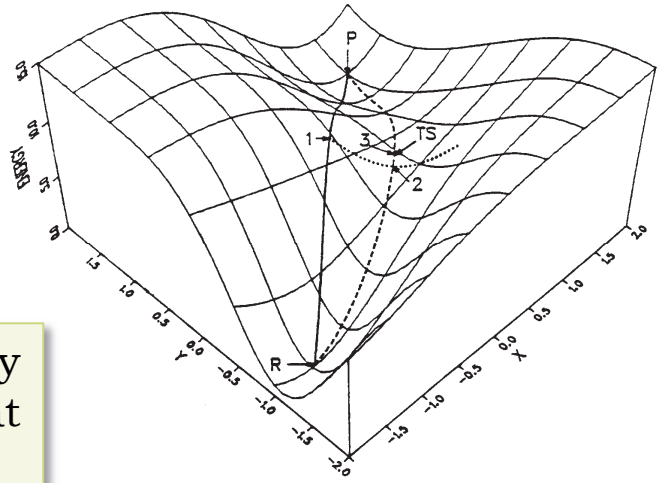
$$\vec{R}(p) = (1-p)\vec{R}^{\text{React}} + p\vec{R}^{\text{Prod}}$$

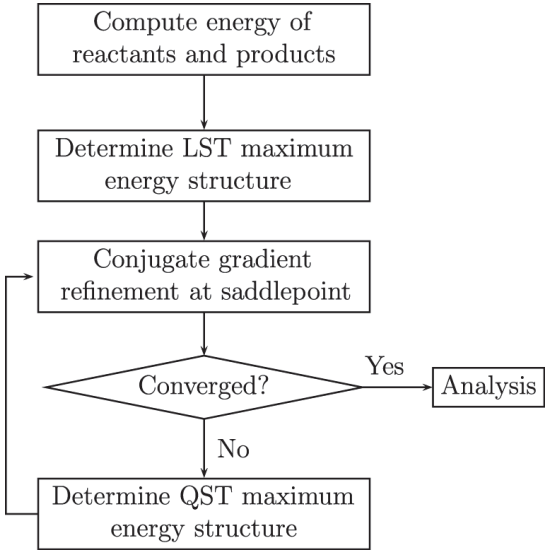
The energy is maximized as a function of $p$. This is a very crude approximation, which may be improved somewhat by performing a quadratic interpolation instead:

$$\vec{R}(p) = (1-p)\vec{R}^{\text{React}} + p\vec{R}^{\text{Prod}} + \vec{R}^{\text{M}} p(1-p)$$

The intermediate point is chosen in a direction perpendicular to the LST trajectory. This is known as Quadratic Synchronous Transit. The procedure may be further improved by CG refinement of the candidate saddle point.

The most sophisticated ST method is STQN (Synchronous Transit Guided Quasi-Newton), which uses the QST result as the initial guess for a local optimization.




Flowchart of LST/QST algorithm

K.P. Lawley, H.B. Schlegel, http://dx.doi.org/10.1002/9780470142936.ch4

# Plain elastic band method

The initial guess for the $N$-point *minimal energy path* (MEP) is obtained by linear interpolation of internal coordinates:

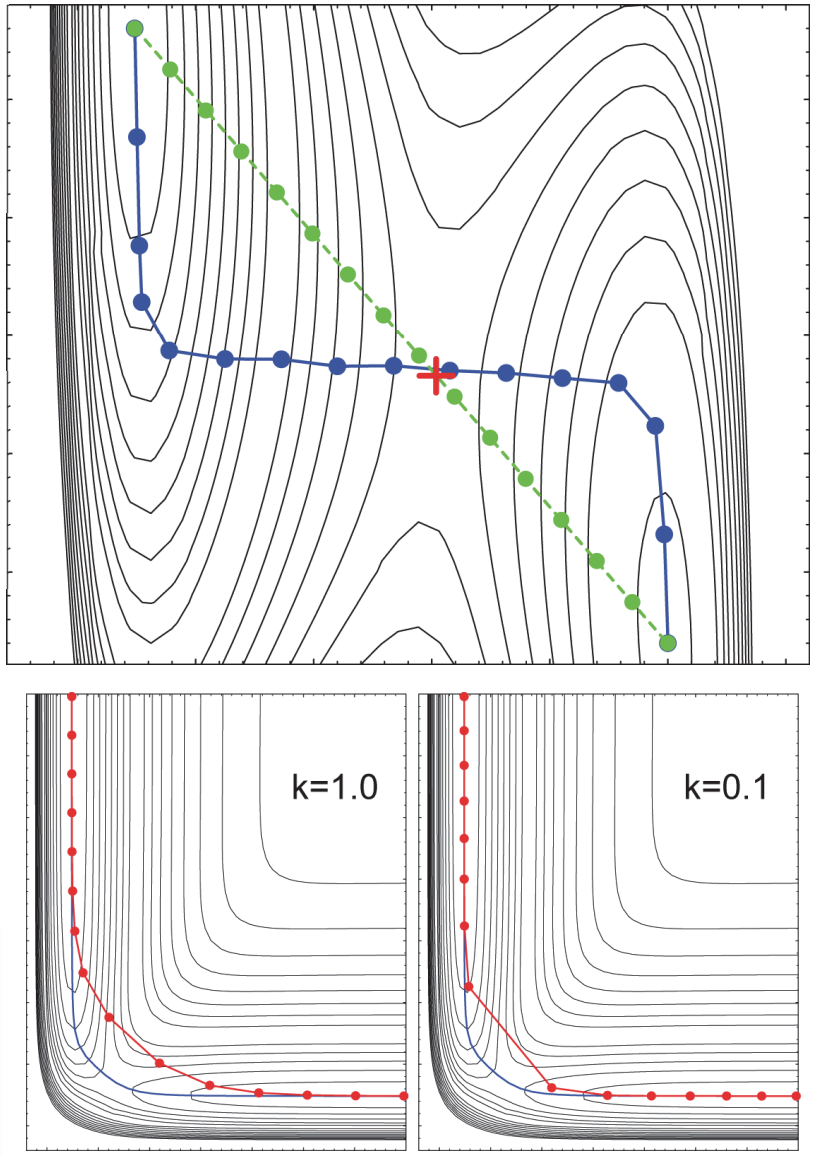
$$\vec{R}_n = \vec{R}_1 + \frac{n-1}{N-1}\left(\vec{R}_N - \vec{R}_1\right)$$

The individual images of the system are then connected by harmonic springs, so that the energy functional becomes:

$$E = \sum_{n=1}^{N} U\left(\vec{R}_n\right) + \sum_{n=2}^{N} \frac{k}{2}\left|\vec{R}_n - \vec{R}_{n-1}\right|^2$$

$$\vec{F}_k = \vec{\nabla} U\left(\vec{R}_k\right) + k\left(\vec{R}_{k+1} - \vec{R}_k\right) - k\left(\vec{R}_k - \vec{R}_{k-1}\right)$$

The path obtained by simple minimization of $E$ is not exactly MEP (although it can be close), because the spring term distorts the effective energy landscape.




H. Jonsson, G. Mills, K.W Jacobsen, http://theory.cm.utexas.edu/henkelman/pubs/jonsson98_385.pdf

# Nudged elastic band method

The shortcomings of PEB can be rectified if the perpendicular component of the spring force and the parallel component of the potential force are projected out of the gradient:

$$\vec{F}_k^{\mathrm{NEB}} = \left[\vec{\nabla} U\left(\vec{R}_k\right)\right]_{\perp} + \left[k\left(\vec{R}_{k+1} - \vec{R}_k\right) - k\left(\vec{R}_k - \vec{R}_{k-1}\right)\right]_{\parallel}$$

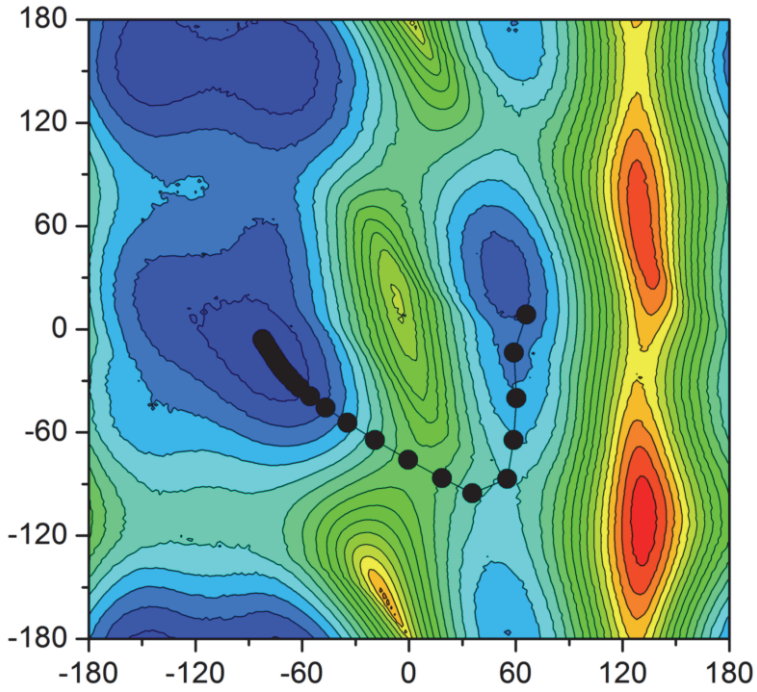
Energy minimization under this force is known as “nudging”. It decouples the dynamics of the point distribution on the path from the dynamics of the path itself.

Advantages of NEB:

1. Converges to MEP.
2. Does not require Hessian information.
3. Always produces a continuous MEP.
4. Easy to parallelize.

Caveats:

1. The number of images must be sufficient.
2. Multiple MEPs may exist.
3. The saddle point requires further refinement.
4. Slow convergence on hilly landscapes.




C. Bergonzo, C. Simmerling, R. Walker, *AMBER Tutorials*, 2010.

# Practical considerations

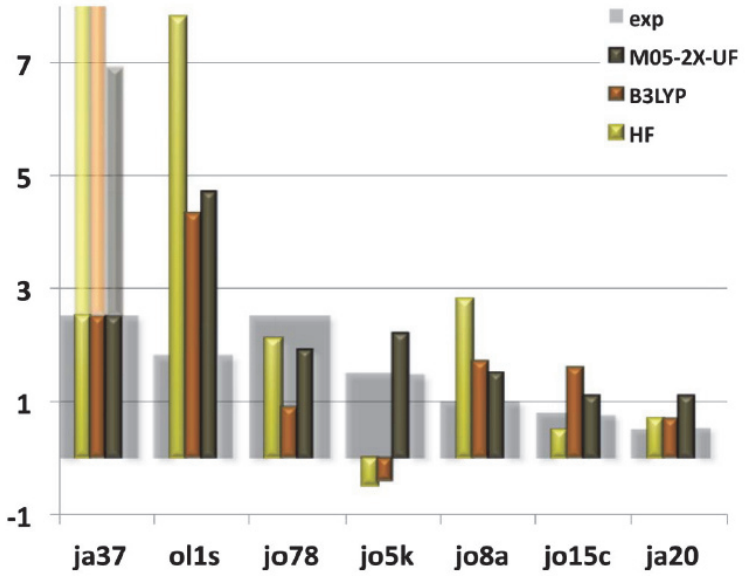



**Fig. 4** Comparison of $\Delta\Delta G^{\ddagger}$ calculated with the values obtained from the experiments.

Transition states often have highly correlated multi-reference wavefunctions. Even CCSD(T) results would in many cases be an estimate.

DFT calculations are often wrong by 20-50%.

Table 4
Calculated thermodynamic activation parameters at $T = 298.15$ K for both steps of hydrolysis

| Method | (1) | |
|---|---|---|
| | $\Delta H^{\ddagger}$298 kcal/mol) | $\Delta S^{\ddagger}$ (cal/K mol) |
| MP2/LANL2DZ/6-31G* | 22.9 | 5.1 |
| B3P86/LANL2DZ/6-31G* | 23.2 | 4.1 |
| BLYP-SCIPCM//B3P86[a] | 20.9 | 4.1 |
| BLYP-SCIPCM//LSDA[b] | 19.9 | 2.2 |
| Experiment—see Ref. [1] | 19.5–21.5 | 13.9 |

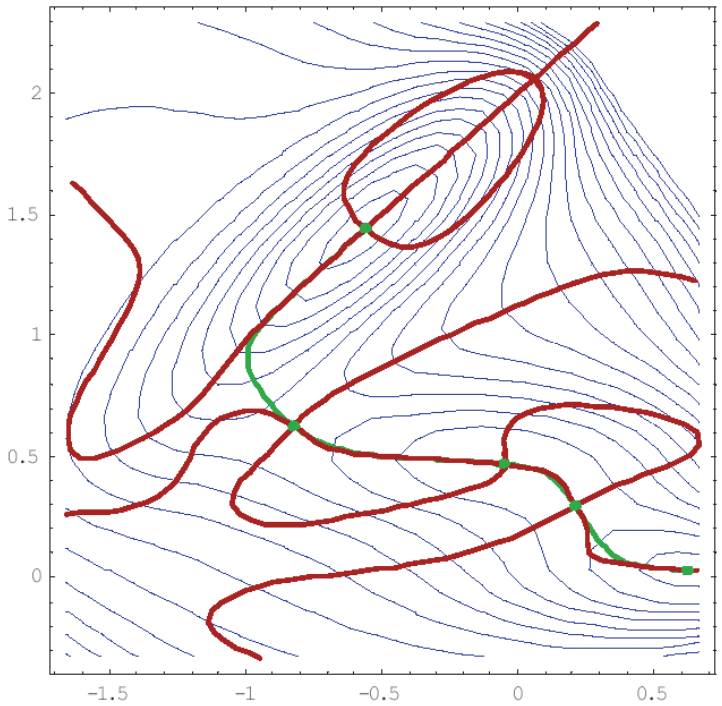
[a] B3P86/LANL2DZ/6-31G* 'vacuum phase' thermal corrections are used.
[b] LSDA/LANL2DZ/6-31G* 'vacuum phase' thermal corrections are used.

Table 1
Calculated rotational barrier for benzaldehyde

| | | |
|---|---|---|
| HF | STO-3G | 5.9 |
| | 3–21G | 11.3 |
| | 6–31G | 9.4 |
| | 6–311G//3–21G | 9.2 |
| | 6–311G** //6–31G* | 8.8 |
| post-HF | MP2/6–311G** //6–31G* | 8.3 |
| | MP2/D95V | 8.9 |
| | RCCSD (6–31G*) | 7.8 |
| MOLCAS | | |
| | ANO-II/CAS(8) | 5.0 |
| | ANO-II/RAS(14) | 4.3 |
| DFT | D-VWN/DZVP2(A2) | 10.5 |
| | BLYP/DZVP2(A2) | 9.5 |
| | D-VWN/TZVP(A2) | 10.2 |
| | B88-PW91/TZVP(A2) | 9.0 |
| | PW91-PW91/TZVP(A2) | 9.3 |
| | B88-P86/TZVP(A2) | 9.2 |
| | BLYP/TZVP(A2) | 9.1 |
| | FT-97/TZVP(A2) | 8.7 |
| | HCTH/TZVP(A2) | 8.5 |
| Experimental values | 4.9 (microwave spectr.) | |
| | 4.7 (gas-phase IR spectr.) | |

R.J. Meier, E. Koglin, http://dx.doi.org/10.1016/S0009-2614(02)00032-5

# Practical considerations



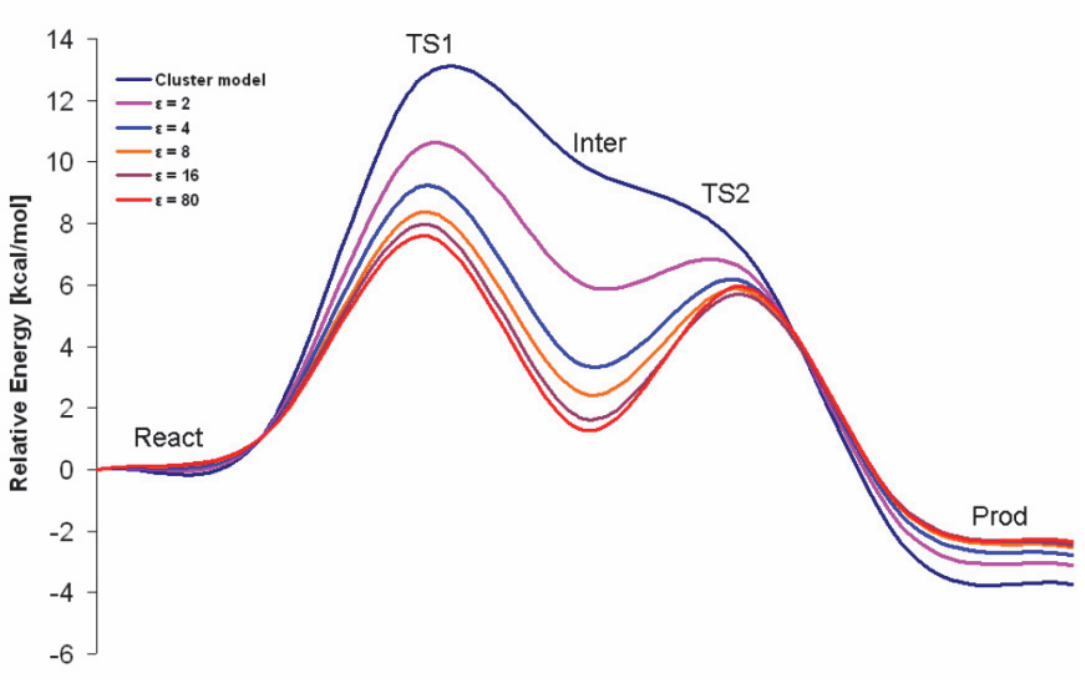




Figure 4. Gradient extremals (red) and steepest descent reaction paths (green) on the Müller-Brown surface.

1. Gradient extremal following gets the stationary points, but not the reaction paths.
2. Electrostatic environment can significantly alter the intermediate energy (intermediates are often charged or have a large dipole moment).
3. Convergence of saddle optimization is often slow – stock up on patience and tighten up all tolerances.

H.B. Schlegel, *J. Comp. Chem.* 24 (2003) 1514-1527.

# Practical considerations

| If you were looking for ... | And the frequency calculation found ... | It means ... | So you should ... |
|---|---|---|---|
| A minimum | 0 imaginary frequencies | The structure is a minimum. | Compare the energy to that of other isomers if you are looking for the global minimum. |
| A minimum | ≥ 1 imaginary frequencies | The structure is a saddle point, not a minimum. | Continue searching for a minimum (try unconstraining the molecular symmetry or distorting the molecule along the normal mode corresponding to the imaginary frequency). |
| A transition state | 0 imaginary frequencies | The structure is a minimum, not a saddle point. | Try using **Opt=QST2** or **QST3** to find the TS (see Chapter 3). |
| A transition state | 1 imaginary frequency | The structure is a true transition state. | Determine if the structure connects the correct reactants and products by examining the imaginary frequency's normal mode or by performing an IRC calculation. |
| A transition state | > 1 imaginary frequency | The structure is a higher-order saddle point, but is not a transition structure that connects two minima. | **QST2** may again be of use. Otherwise, examine the normal modes corresponding to the imaginary frequencies. One of them will (hopefully) point toward the reactants and products. Modify the geometry based on the displacements in the other mode(s), and rerun the optimization. |

If the Hessian has more than one negative eigenvalue, that is most probably not the stationary point you are looking for.

J.B. Foresman, A. Frisch, *Exploring Chemistry with Electronic Structure Methods*, Gaussian, 1996.

# Practical considerations

1. All gradient methods preserve molecular symmetry. It is your responsibility to check for non-symmetric (*e.g.* due to Jahn-Teller effect) minima.
2. Correctly chosen internal coordinates will significantly accelerate convergence. Cartesian coordinates are almost always a bad choice.
3. In multi-level optimizations, it is reasonable to retain the approximate Hessian and use it as a guess for the higher level method.

TABLE I. Comparison of geometry optimization performance using internal, cartesian, and mixed interal/cartesian coordinates.

| Molecule | Number of atoms | Symmetry | Number of variables | Number of optimization steps | | | |
|---|---|---|---|---|---|---|---|
| | | | | Internal | | Cartesian | Mixed |
| | | | | (a) | (b) | | |
| 2 fluoro furan | 9 | $C_s$ | 15 | 7 | 8 | 7 | 7 |
| norbornane | 19 | $C_{2v}$ | 15 | 7 | 6 | 5 | 5 |
| bicyclo[2.2.2]octane | 22 | $D_3$ | 11 | 11 | 25 | 19 | 14 |
| bicyclo[3.2.1]octane | 22 | $C_s$ | 33 | 6 | 5 | 6 | 7 |
| endo hydroxy bicyclopentane | 14 | $C_1$ | 36 | 8 | | 18 | 9 |
| exo hydroxy bicylopentane | 14 | $C_1$ | 36 | 10 | | 20 | 11 |
| ACTHCP | 16 | $C_1$ | 42 | 65 | | >81 | 72 |
| 1,4,5 trihydroxy anthroquinone | 27 | $C_s$ | 51 | 10 | | 11 | 17 |
| histamine $H^+$ | 18 | $C_1$ | 48 | 42 | | >100 | 47 |

H.B. Schlegel, *Intl. J. Quant. Chem. 26 (1992) 243-252.*

# Miscellaneous notes

1. Geometry optimization jobs should use tight convergence cut-offs at the SCF stage to avoid numerical noise in the gradients.
2. Invalid dihedrals (they appear in flat rings and linear systems) should be avoided – any optimization in such systems is best performed in Cartesians.
3. In very extended systems (e.g. carotene chains) small displacements of internal coordinates can translate to large displacements in Cartesians. Convergence criteria must be set appropriately.
4. The initial optimization of van der Waals complexes is best performed in Cartesian coordinates.
5. If a very good approximation to a minimum is known, it is best to start away from it – many numerical optimization algorithms crash if they cannot make the initial step.
6. Positive definiteness must be *strictly enforced* in the Hessian when looking for energy minima. Quasi-Newton optimizations with indefinite Hessians converge into saddle points.
7. Optimizations involving implicit solvent would not usually converge to the same accuracy as vacuum calculations due to the numerical noise arising from the discretization of the solvent cavity.

J.B. Foresman, A. Frisch, *Exploring Chemistry with Electronic Structure Methods*, Gaussian, 1996.