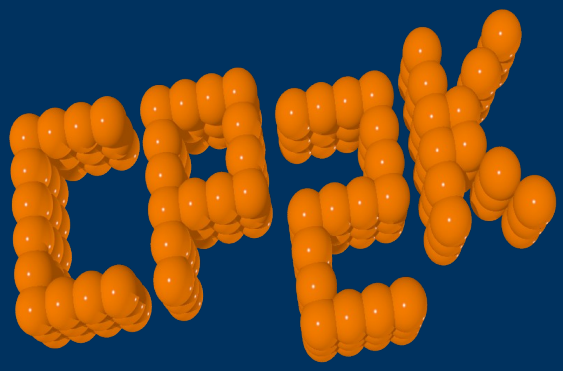
UNIVERSITY OF
LINCOLN

Protocols for geometry and cell optimization



Matt Watkins
mwatkins@lincoln.ac.uk

UNIVERSITY OF
LINCOLN

Optimization

- Optimize what? Positions of the atoms
  - typically the total energy of the system (potential energy + electron kinetic energy for DFT) at 0K
  - Free energy of the system at finite temperature
- PES – Born-Oppenheimer approximation
- Global vs local?

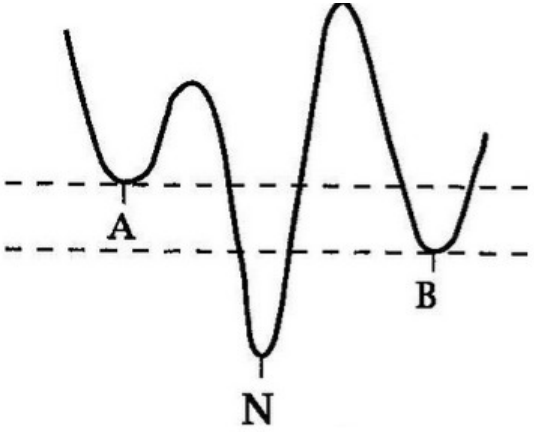


Fig. 3. Defining thermodynamic vs. kinetic reaction coordinates. Which state, A or B, is further "along the reaction coordinate" toward the native state N? State B is *energetically* closer to N (lower energy), but A is *kinetically* closer (smaller barrier to cross).

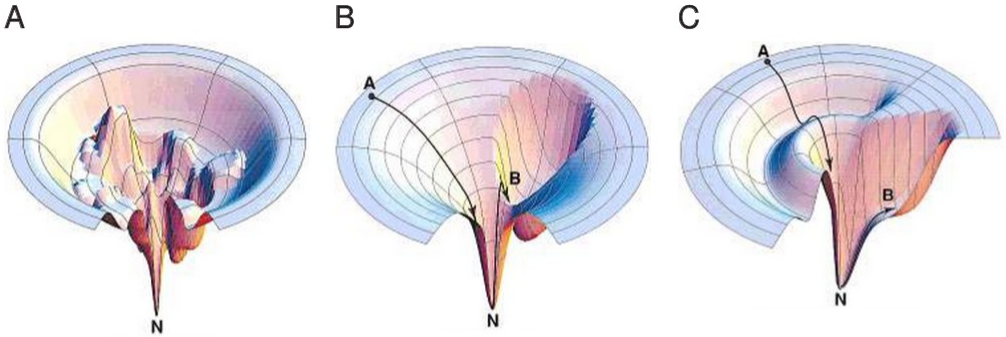


Fig. 12. Different folding scenarios. The vertical axis is internal free energy. Each conformation is represented as a point on the landscape. The two horizontal axes represent the many chain degrees of freedom. **a:** A rugged landscape with hills and traps, folding kinetics is likely multiple-exponential (from Ref. 8). **b:** A landscape in which folding is faster than unfolding. A is a through-way folding path, whereas unfolding chains (path B) must surmount a barrier to reach the most stable denatured conformations. **c:** A landscape in which folding is slower than unfolding. Most folding paths (path A) pass through a kinetic trap, whereas some low-lying denatured conformations are readily accessible from the native state during unfolding (path B).

Local minimizers

- BFGS
  - most efficient for small–medium size systems with a reasonable guess at the geometry
  - requires inversion/diagonalization of approximate Hessian matrix – Hessian matrix has dimension 3*N* where *N* is number of atoms being optimized
- L-BFGS
  - is a linear scaling version of the BFGS algorithm (Byrd, Richard H., et al. "A limited memory algorithm for bound constrained optimization. *SIAM Journal on Scientific Computing* 16.5 (1995): 1190-1208.")
- Conjugate gradients
  - Only uses gradients rather than approximation to curvature, should be more robust when far from minima
- Steepest descents
  - head downhill, line search to find how far – most robust far from minima

Global optimizers

- Brute force requires m^{3N} energy evaluations to get a grid of m points in each dimension
- Exponential growth with system size
- Shape of PES can be exploited to get methods that work in practice for modest size systems
  - Random search – just a script for initial coords
  - Genetic algorithms
  - Simulated annealing – ANNEALING keyword in MD section
  - Monte Carlo
  - Basin hopping
- Swarm methods in CP2K have basin hopping
  - $CP2K/tests/SWARM – Ole Schutt

- GEO_OPT – energy minimization allowing atomic coordinates to change
- CELL_OPT – allow cell parameters to vary when finding the local energy minima too
- CONSTRAINTS – there may be some variables that we want to fix – cell angles or positions of some atoms
- Collective variables – plot energy as function of some variable – a bond length or angle, for instance – reduce dimensionality

GEO_OPT

- RUN_TYPE GEO_OPT in global section
- OPTIMIZER in MOTION section
  - CG, use with poor intial guesses, noisy forces, rough optmization
  - (L)BFGS, for most QS calculations – consider switching to LBFGS above ~1000 atoms. Look for diagonalization routine timings at end of run to see relative cost
- MAX_ITER number of optimization steps
- CONSTRAINTS in MOTION section
  - &FIXED_ATOMS
    - COMPONENTS_TO_FIX X
    - LIST 1
  - &END
  - &FIXED_ATOMS
    - COMPONENTS_TO_FIX Y
    - LIST 2
  - &END

output

- grep for “Max. grad” in output file to see the progress of the optimization – this gives maximum energy gradient on atoms being optimized
- Below “Convergence check :” there is a summary of the progress. Output is like Gaussian for those familiar – convergence requires Max and RMS step size and Max and RMS gradients to be converged.
- Pressure extra criteria for CELL_OPT
- The convergence criteria can be set in the &MOTION section
- Default Max. grad is equal to 0.025 eV/A
- Good enough for most purposes

```
ENERGY| Total FORCE_EVAL ( QS ) energy (a.u.):           -2314.583506508490700


*******************************************************************************
***                       2PNT LINE SEARCH INFO                             ***
***                                                                         ***
***   DX (EVALUATED)=    0.025697          DX (THRESHOLD)=    0.250000      ***
***   DX (FITTED   )=    0.252697          DX (ACCEPTED )=    0.250000      ***
*******************************************************************************

--------  Informations at step =     4 ------------
 Optimization Method        =                   SD
 Total Energy               =     -2314.5828781392
 Internal Pressure [bar]    =     35379.3762511153
 Real energy change         =       -0.0099708467
 Decrease in energy         =                  YES
 Used time                  =              465.309

 Convergence check :
 Max. step size             =       0.1443379699
 Conv. limit for step size  =       0.0030000000
 Convergence in step size   =                  NO
 RMS step size              =       0.0456435465
 Conv. limit for RMS step   =       0.0015000000
 Convergence in RMS step    =                  NO
 Max. gradient              =       0.0148361109
 Conv. limit for gradients  =       0.0004500000
 Conv. for gradients        =                  NO
 RMS gradient               =       0.0046915771
 Conv. limit for RMS grad.  =       0.0003000000
 Conv. for gradients        =                  NO
 Pressure Deviation [bar]   =     35279.3762511153
 Pressure Tolerance [bar]   =      100.0000000000
 Conv. for  PRESSURE        =                  NO
-------------------------------------------------

CELL| Volume [angstrom^3]:                                           180.108
CELL| Vector a [angstrom]:       5.647     0.000     0.000    |a| =       5.647
CELL| Vector b [angstrom]:       0.000     5.647     0.000    |b| =       5.647
CELL| Vector c [angstrom]:       0.000     0.000     5.647    |c| =       5.647
CELL| Angle (b,c), alpha [degree]:                                       90.000
```

```
--------  Informations at step =     5 ------------
 Optimization Method        =                   CG
 Total Energy               =     -2314.5861062993
 Internal Pressure [bar]    =       44.8922780589
 Real energy change         =       -0.0032281601
 Decrease in energy         =                  YES
 Used time                  =              484.509

 Convergence check :
 Max. step size             =       0.0000008113
 Conv. limit for step size  =       0.0030000000
 Convergence in step size   =                  YES
 RMS step size              =       0.0000004766
 Conv. limit for RMS step   =       0.0015000000
 Convergence in RMS step    =                  YES
 Max. gradient              =       0.0000003312
 Conv. limit for gradients  =       0.0004500000
 Conv. in gradients         =                  YES
 RMS gradient               =       0.0000001945
 Conv. limit for RMS grad.  =       0.0003000000
 Conv. in RMS gradients     =                  YES
 Pressure Deviation [bar]   =       -55.1077219411
 Pressure Tolerance [bar]   =      100.0000000000
 Conv. for  PRESSURE        =                  YES
-------------------------------------------------

*******************************************************************************
***                   GEOMETRY OPTIMIZATION COMPLETED                       ***
*******************************************************************************

                  Reevaluating energy at the minimum
```

CELL_OPT

- Only place CP2K uses symmetry(?)
- Can set cell symmetry types in &CELL section

```
&SUBSYS
  &CELL
    ABC 9.167 9.167 11.808
    SYMMETRY ORTHORHOMBIC
    MULTIPLE_UNIT_CELL 2 2 2
  &END CELL
  &TOPOLOGY
    COORD_FILE_NAME tio2.xyz
    COORD_FILE_FORMAT XYZ
```

- Can place constraints on the cell optimization in the %CELL_OPT section

```
&MOTION
  &CELL_OPT
    KEEP_SYMMETRY .TRUE.
  &END CELL_OPT
&END MOTION
```

  - KEEP_ANGLES, KEEP_SYMMETRY

CELL_OPT

- Three algorithms in CP2K (version > 2.4) controlled by TYPE variable in $CELL_OPT
  - GEO_OPT: Original implementation.
    1. Inner cycle optimize atomic positions
    2. Outer cycle optimize cell vectors
  - DIRECT_CELL_OPT: New implementation from version 2.4 onwards – cell parameters (stresses) go into the optimizer along with atomic coordinates
  - MD: Optimize at finite temperature. Uses MD, so only of use if you have a cheap Hamiltonian
- DIRECT_CELL_OPT should be much more efficient – try for yourself
- Generally best to enforce symmetry / fix angles to start with to minimize number of degrees of freedom.

Transition states

Two methods for transition state

- Nudged Elastic Band method – covered by Teo later
- Dimer method – two images of the system (the dimer) calculate energy / force at each – then rotate to find steepest path up hill and head upwards.
  - Needs guess (vector) at the up-hill direction – can be generated from difference in coordinates between minima and guess at TS
  - No guarentee of converging (to the TS you want)
  - Should be a good method for refining TSs obtained form NEB

Optimizing other things

Gradient free optimization
Powell's Algorithm
https://en.wikipedia.org/wiki/Powell%27s_method"

- BASIS SETS – built in
- Pseudo – built in in ATOM package
- FORCE FIELDS – built in force matching algorithm
  - http://www.cp2k.org/exercises:2015_uzh_molsim:h2o_ff
- Anything – general scheme in ${CP2K_ROOT}/cp2k/tools/scriptmini