Computational Chemistry - part II: Optimisation methods

Tomas Öberg

University of Kalmar, Department of Biology and Environmental Science

2003-02-07

Content of the three seminars

- Previous meeting
  - Some concepts in molecular modelling
  - A brief introduction to *ab initio* modelling
  - Empirical force field models: Molecular mechanics
- Today
  - Energy minimisation
  - Optimisation methods
- Last Meeting
  - Chemoinformatics
  - Molecular descriptors
  - QSAR/QSPR, chemometrics

Literature

- Andrew R. Leach (2001): *Molecular modelling - principles and applications*, 2nd edition, Prentice-Hall, 744 pp.
- Walters, F. H. et al (1991): *Sequential simplex optimization*, CRC Press, 325 pp.
- Öberg, T.; Deming, S. N. *Find optimum operating conditions fast*, Chemical Engineering Progress **96**, 53-59, 2000.

2003-02-07

Potential energy of a molecule

- Last time we reviewed the Born-Oppenheimer approximation, what did it state?
  - Electronic and nuclear motions can be separated
- What does this imply for the energy of a molecule?
  - It is only a function of the nuclear coordinates (in the ground state)

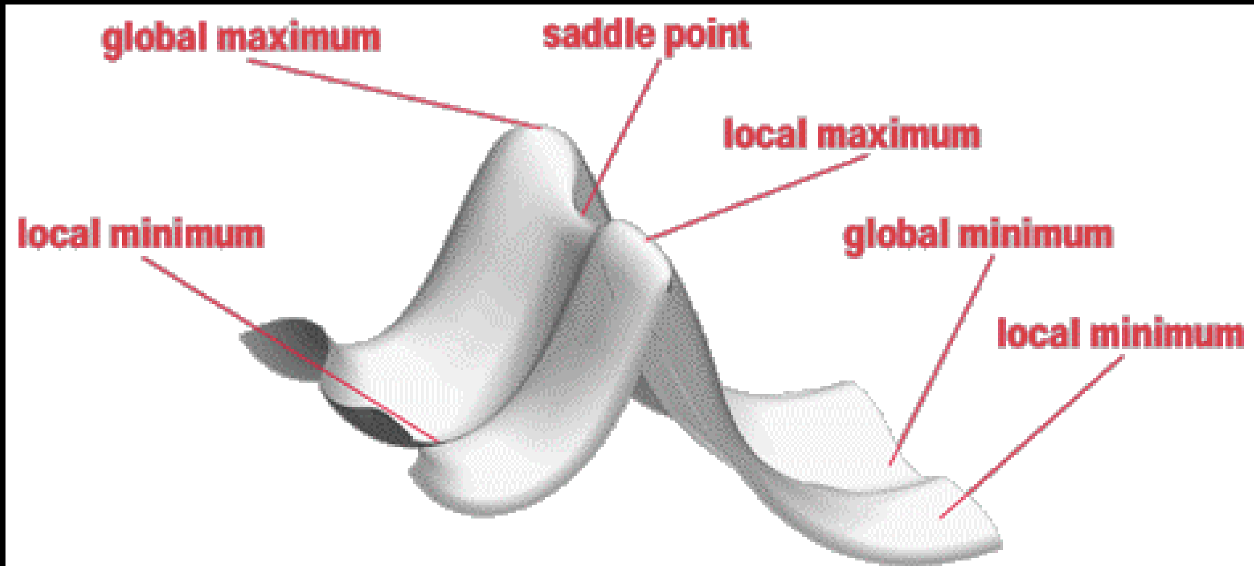
The potential energy surface (PES)

- The changes in energy of a system (molecule) can be viewed as movements on a multidimensional surface, why?
  - In a molecular system consisting of *N* atomic nuclei, the number of the independent coordinates that fully determine the PES is equal to *3N-6* internal or *3N* Cartesian coordinates.

Minimum points

- Why are we interested in locating the minimum points of the potential energy surface?
  - The minimum correspond to a stable state of the system
- Are all molecules at the minimum?
  - No, there can be several minima and the molecule conformations are not static, but are statistical properties i.e. some parts of PES are more densely populated than others (the Boltzmann distribution)

Local minima, saddle points and the global minimum



- At stationary points the first derivative of the energy function is zero, which are those?
  - In fact all are, although our interest mostly focus on minima and saddle points

The quest is to find the minima

- How can we find the minima?
  - Calculus methods ca be applied tor analytical functions

  but

  - this is generally to complicated for molecular systems
- Alternative?
  - Numerical methods, an iterative process that gradually change coordinates lower states of potential energy until the minimum is found

Optimisation algorithms

- We have already mentioned one function property that can be used to identify a minimum, which one?
  - derivative based (classifies one important group of optimisation algorithms)

  the other being
  - non-derivative based

Derivative optimisation methods

- It is often said that derivative based methods are more efficient
- Why would this be the case?
- Does the derivative provide anything besides identification of the stationary points?
  - Yes, the derivative also provide information about the shape of the energy surface
  - It tell us about the size (“steepness”) and direction of the gradient

First-order methods: Steepest descent

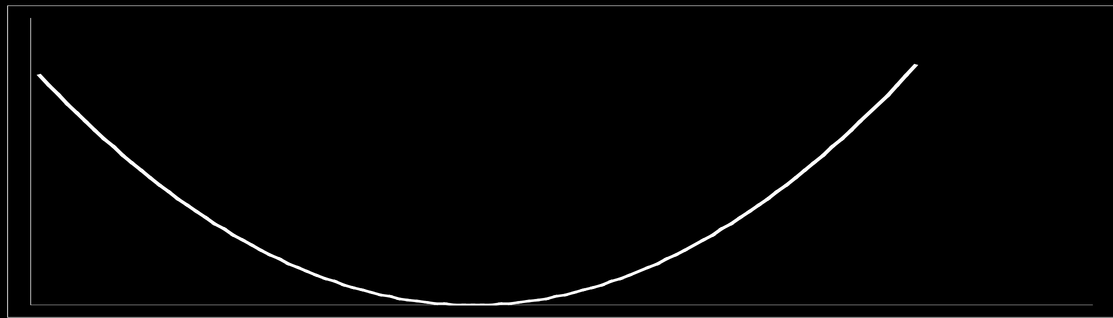
- Steepest descent is one of the first-order methods, i.e. only the first derivative is used
- Two first-order methods will be described
  – Line search in one dimension

  and

  – Steepest descent in a multiple dimensions

2003-02-07

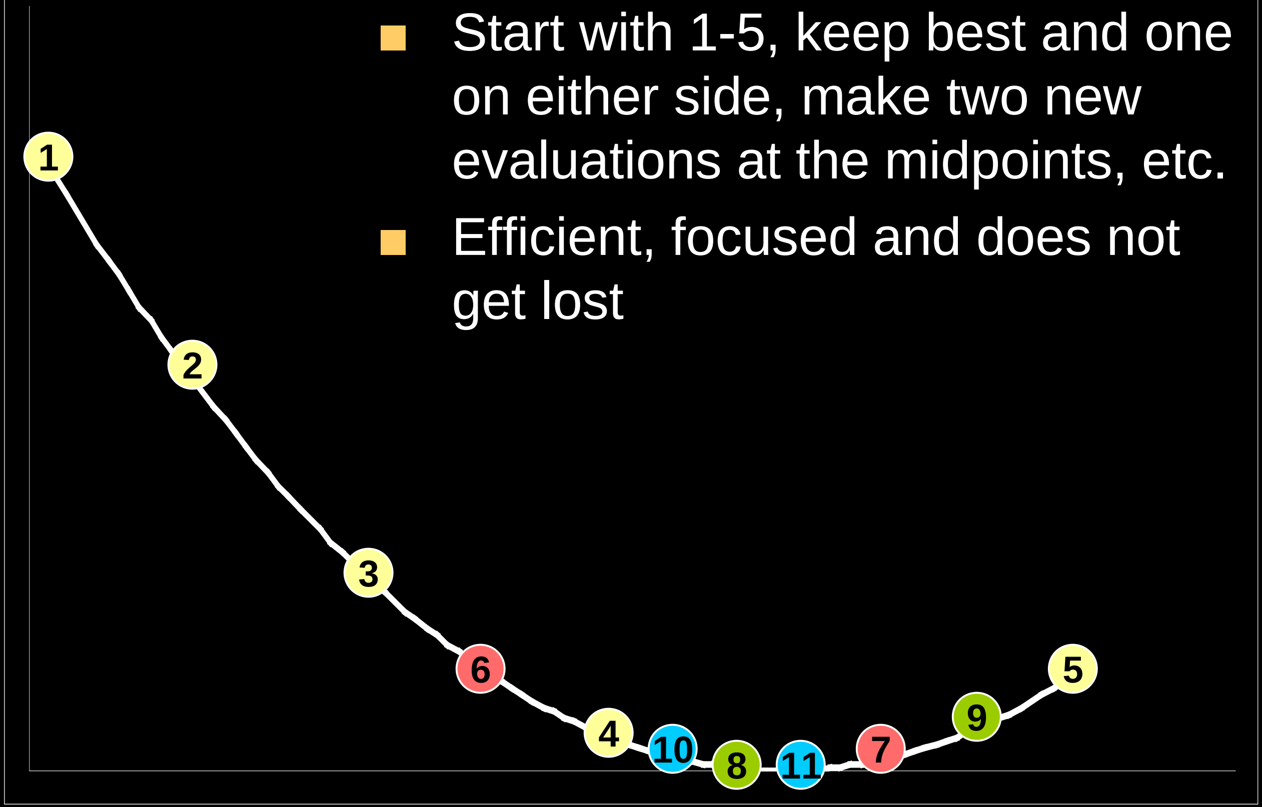
Line search in one dimension



- If we have an unknown analytical function as shown, what would be the best strategy to locate the optimum?
  - We obviously need information, and must start with evaluating at least three points
  - We can then continue with an iterative procedure or fit a function e.g. quadratic

A Fibonacci-type search

- Start with 1-5, keep best and one on either side, make two new evaluations at the midpoints, etc.
- Efficient, focused and does not get lost

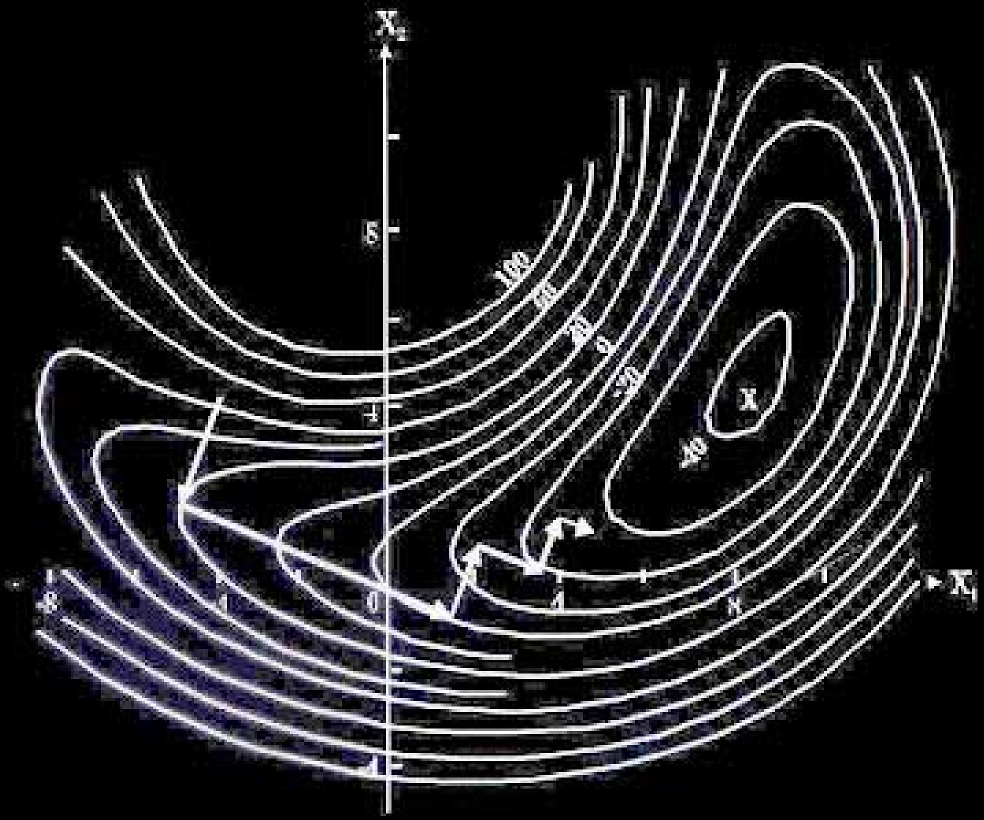


2003-02-07

Steepest descent in multiple dimensions

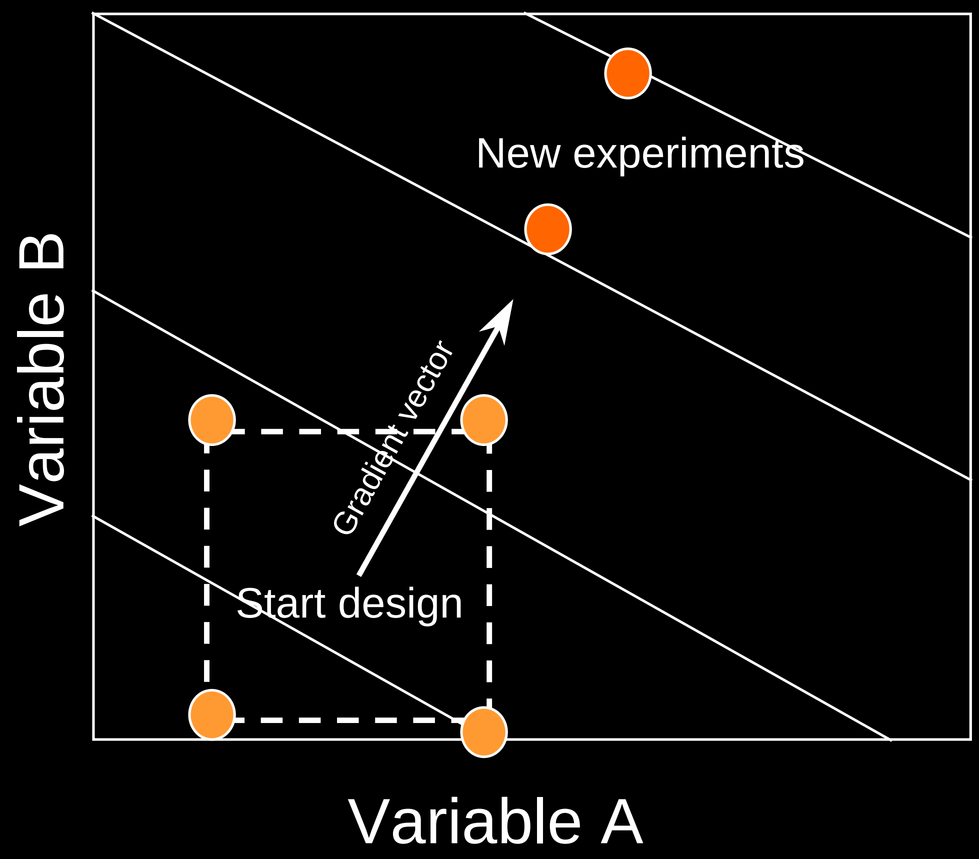
- How can this approach then be generalised to multiple dimensions?
  - The gradient vector for the initial configuration, i.e. first-order derivative, is computed
  - A line search is performed along the gradient vector
  - When the response does not improve, a new gradient of steepest descent estimated orthogonal (perpendicular) to the previous one
- Steepest descent methods are most efficient in locating the optimal region, but less so in locating the exact optimum

Steepest descent illustrated



2003-02-07

Steepest descent “footnote”: Experimental optimisation

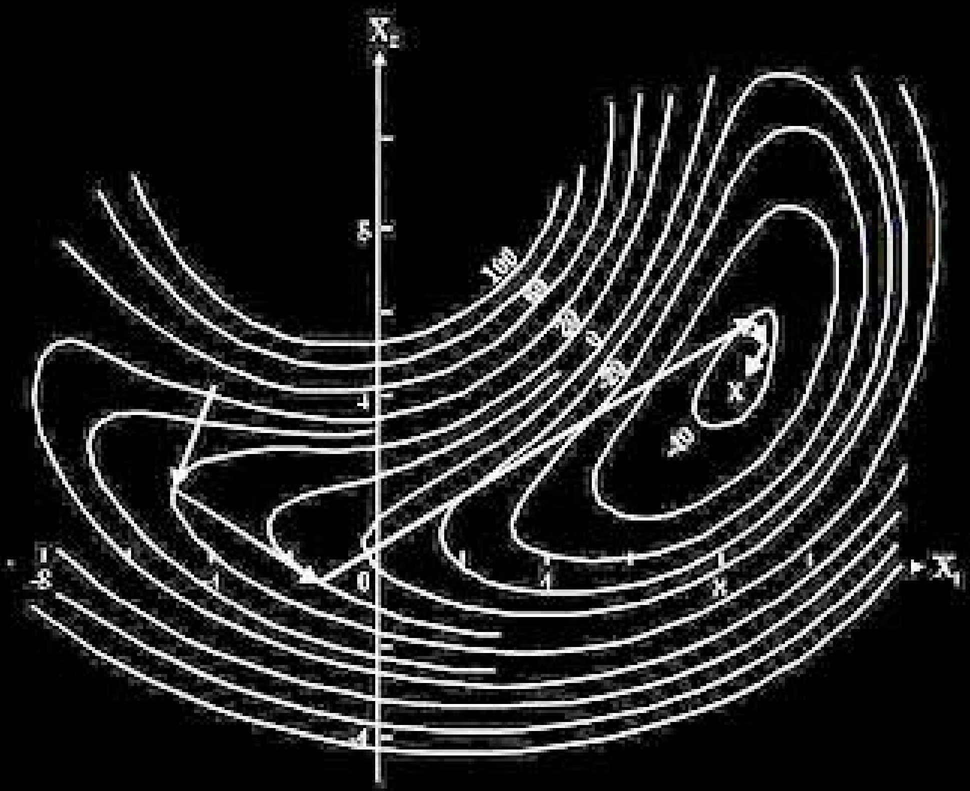


2003-02-07

Conjugated gradient minimisation

- Can you suggest any approach to improve the steepest descent method, i.e. make it faster?
  - Uses not only the first derivative information in each step, but also information from previous steps
  - The conjugated gradient method use a weighted average of the current gradient and the previous step direction.
  - The weight factor is calculated from the ratio of the previous and current steps
  - The Fletcher-Reeves and Polak-Ribiere methods are examples of conjugated gradient algorithms

Conjugated gradient illustrated



2003-02-07

Second derivative methods

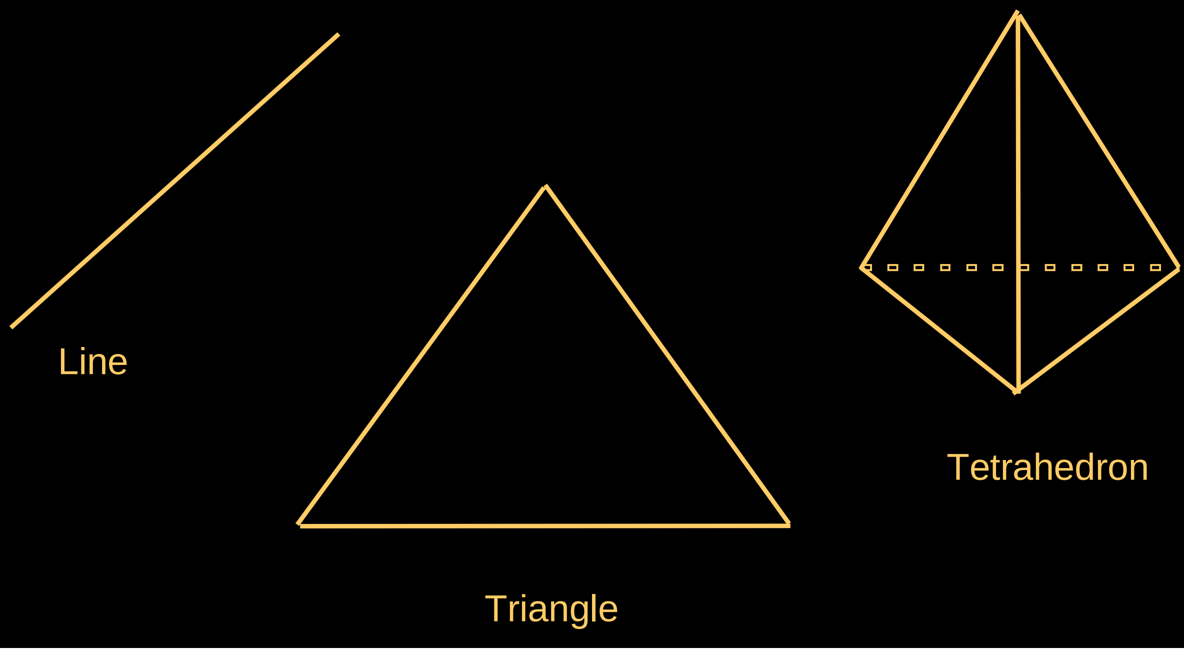
- What does the second derivative tell us?
  - It provide information about the curvature of the function, i.e. the second derivative is positive at a minimum
  - The Newton-Raphson method is the “simplest” second-order method
  - However NR is more complicated than the previous first-order methods, since the inverse of a Hessian (curvature) matrix must be calculated at each step
  - NR is accurate and converges well, but computationally expensive

Non-derivative methods

- When are non-derivative methods useful?
  - When it is difficult or impossible to obtain the derivative, examples?
  - Complex nonlinear response functions
  - An extreme number of dimensions, i.e large molecules
- Sequential simplex, simulated annealing, genetic algorithms, random and grid search

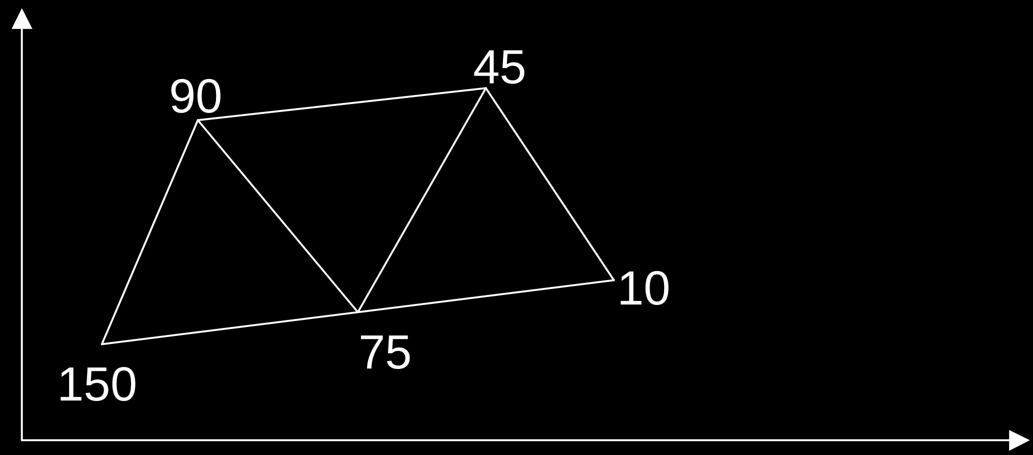
The sequential simplex method

- What is a simplex?
  – A geometric figure with M+1 interconnected vertices in a M dimensional space



2003-02-07

The basic simplex algorithm



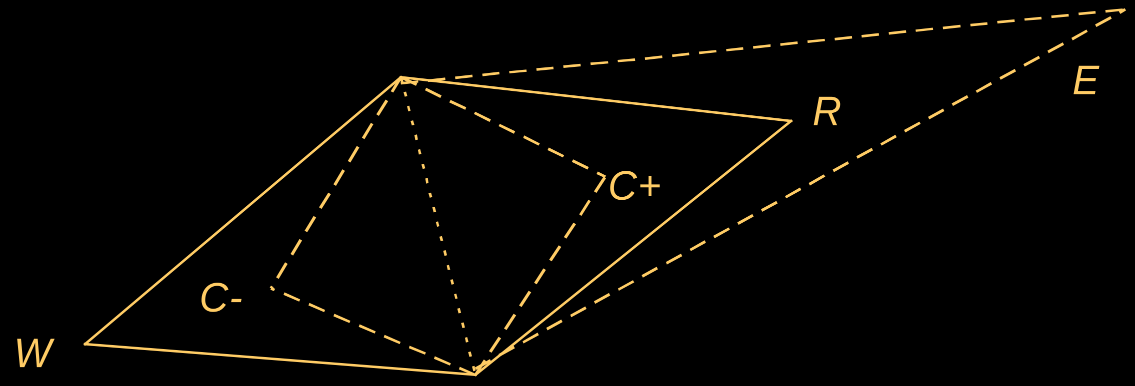
- What is the next move if we want to find the minimum?
  - Reflect (projection through the centroid) to move away from the unfavourable setting

Possible modifications to the basic simplex algorithm

- The progress of the basic simplex depend on the step-size
  - It was developed for sequential process optimisation, as an alternative to statistical experimental design
  - The inventors originally suggested the sequential application of smaller simplexes
- Can it be modified to work more efficiently?
  - What can be the modification if the response improve?
  - What would be the modification if the response get worse?

The modified simplex algorithm (Nelder-Mead)

- Nelder and Mead (1965) modified the simplex algorithm to make expansions and contractions possible
- Why not expand more?
  - The simplex would degenerate, i.e. collapse into a search space of less dimensionality



2003-02-07

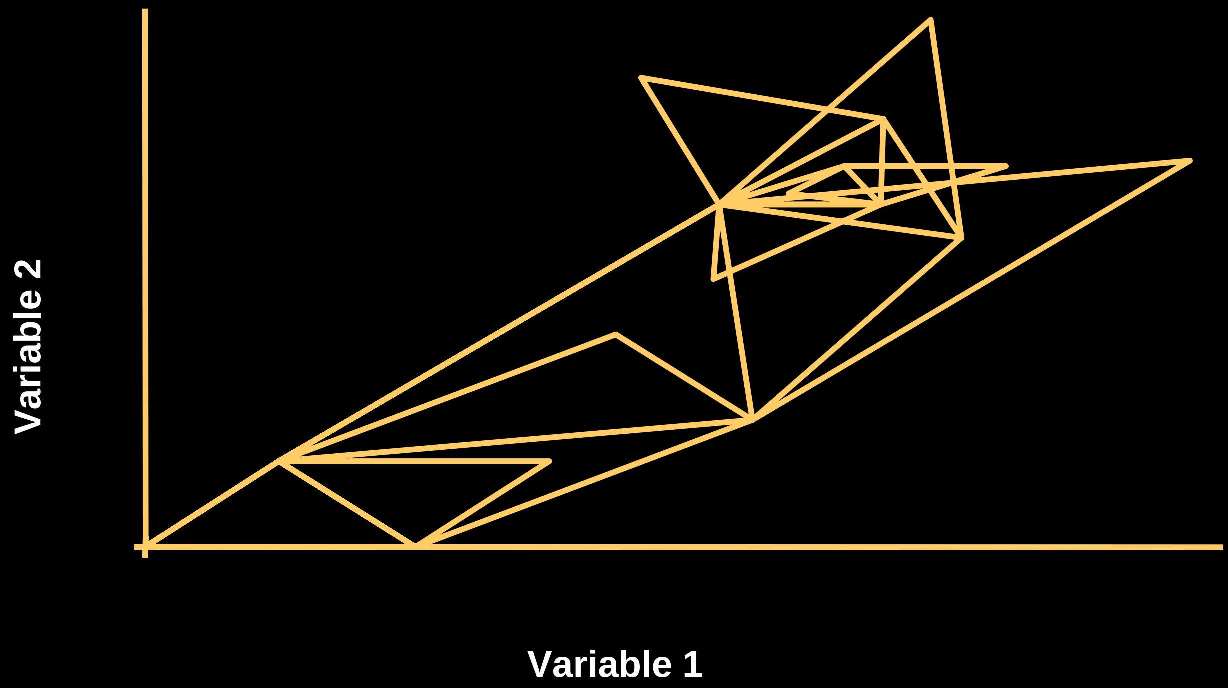
Modified “rules” (implemented by Deming)

- Evaluate R after reflection:
  - If Nw<R=<B, use simplex B..NwR
  - If R>B, evaluate E:
    - If E >R, use simplex B..NwE
    - If E=<R, use simplex B..NwR
  - If R=<Nw:
    - If R>W, evaluate C+, use simplex B..NwC+.
    - Om R=<W, evaluate C-, use simplex B..NwC-

Simple calculations

Rank	***Var 1***	***Var 2***
Nw	37.60	65.90
B	51.75	51.75
C (medelv.)	44.68	58.83
W	20.18	34.32
C-W	24.50	24.51
R=C+(C-W)	69.17	83.33
(C-W)/2	12.25	12.25
C-=C-(C-W)/2	32.43	46.57
C+=C+(C-W)/2	56.92	71.08
E=R+(C-W)	93.67	107.84

The modified simplex in action

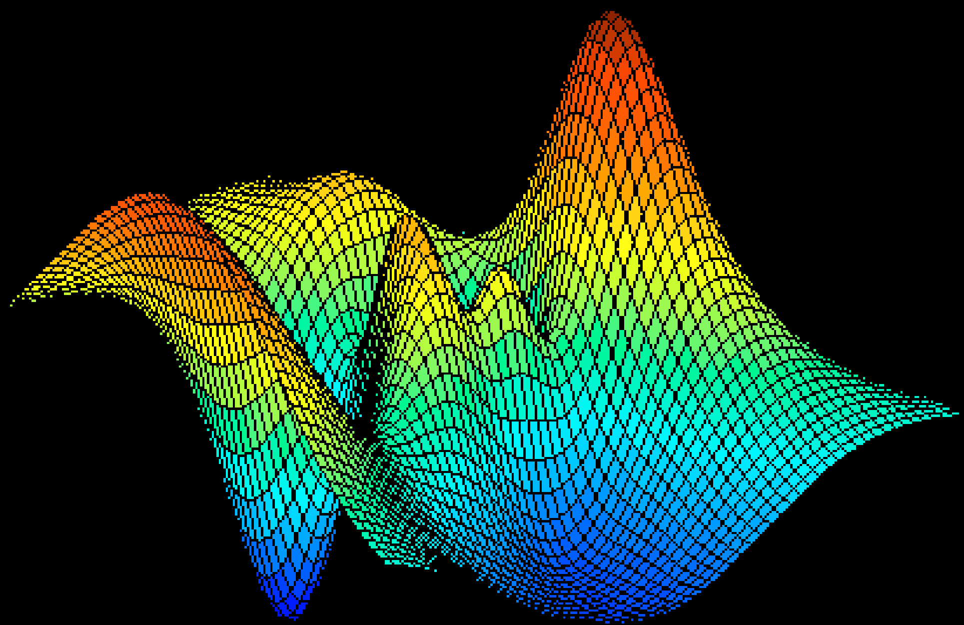


2003-02-07

Convergence

- When should the optimisation stop?
- There are several options for convergence criteria:
  - Numerical precision limit, stop before limit is reached
  - Monitor the change in energy between steps, and stop when a suitable threshold is reached
  - Monitor changes in the coordinates, and stop when changes in configuration is negligible

The global optimum - how can we find it?



2003-02-07

Local optima - a challenge for optimisation

- How can one deal with a situation as shown on the previous slide? How can we be sure to find the global optimum?
  - A separate algorithm to generate different starting structures

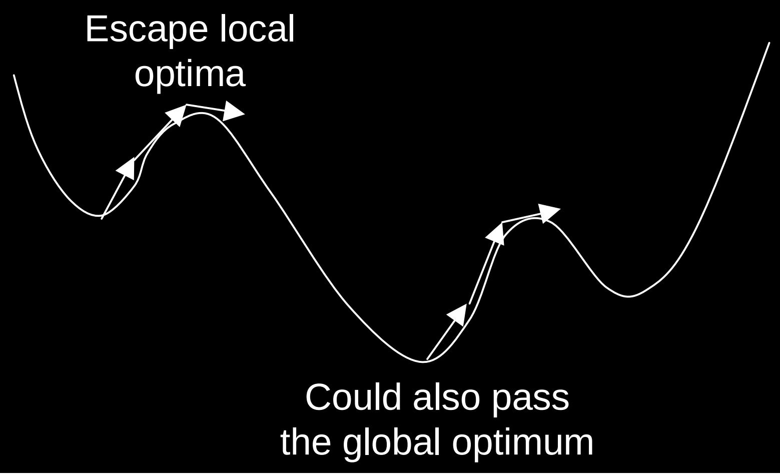
  or

  - an algorithm that move across local minima
- Consequences for the minimisation process?
  - Uses a lot of computational power, and thus time consuming

2003-02-07

Simulated annealing

- Analogy with annealing of heated solids
- Allow occasional ascents in the search process, to escape the trap of local minima
- Slowly ‘cooling’, with less acceptance of ascending steps



2003-02-07

A combined scheme

- Direct search strategies are usually faster than a stochastic search algorithm like ‘simulated annealing’
- Is it possible to combine ‘simulated annealing’ with any of the pervious schemes?
  - The simplex algorithm can easily be modified to accept occasional ascending steps, with a cooling schedule
  - Press WH, Teukolsky SA (1991): *Simulated annealing optimization over continuous* spaces. Computers in Physics **5**, 426-429 (included in “Numerical Recipes”)

2003-02-07

Genetic algorithms - another stochastic approach

- Name due to analogy with evolution in biology
- Create a starting ‘population’ of possible solutions, i.e. conformations
- Score against a ‘fitness function’, i.e. potential energy
  - Selection with bias towards better ‘fitness’
- Generate new solutions ‘breeding’
  - Crossover (recombination)
  - Mutation

Conformational search

- The purpose with a conformational search is to identify the preferred conformations (3D-structures)
- Energy minimisation and optimisation is then crucial, and the methods already described are used
- Are there other approaches to energy optimisation, more or less systematic?
  - Random search
  - Grid search

Random search

- The opposite of systematic search
  - The Cartesian method adds random changes to *x*,*y*,*z* coordinates
  - The dihedral method make random changes to rotatable bonds
- Each time a new random structure is generated, it is subsequently optimised in the traditional way
- When do we stop?
  - There is actually no natural endpoint

2003-02-07

Grid search

- Grid search is a systematic exploration of the conformational space making regular changes to the conformation
- How to implement a simple approach?
  - Identify each rotatable bond
  - Keep bond lengths and angles fixed
  - Rotate each bond systematically through 360º using a fixed increment
  - Each conformation is then subject to energy minimisation as before
- When do we stop?
  - When each torsion angle has been explored

2003-02-07

When do we apply the last two methods

- Random search?
  - Large molecules, that cannot be handled systematically
- Grid search?
  - Small molecules, otherwise the ‘combinatorial explosion’ put an end to it
  - Example, five bonds varied in increments of 30º, will generate 12^5 = 248832 structures

$$No. = \prod_{i=1}^{N} \frac{360}{\theta_i}$$

2003-02-07

In summary

- The minimum points of potential energy surface correspond to stable conformations
- It is therefore of interest to locate these minima and the global minimum
- Derivative optimisation methods are generally fast and accurate, e.g. line search, steepest ascent and conjugated gradient
- Non-derivative methods are useful when the response function is complex or the number of dimensions is high, e.g. sequential simplex, simulated annealing, genetic algorithms and random search

Demonstrations

- With HyperChem®
  - Optimisation of structural properties (geometry) of cyclohexane
  - Conformation searching of 2,2’,3,3’,4,6’-hexabromodiphenylether
- Java-applet
  - The basic and modified simplex algorithms