Lecture 6

Minimization

Summary Lecture 5

- Function calculation
  - Square roots :

$$q = -\frac{1}{2}\left(b + \mathrm{sgn}(b)\sqrt{b^2 - 4ac}\right)$$

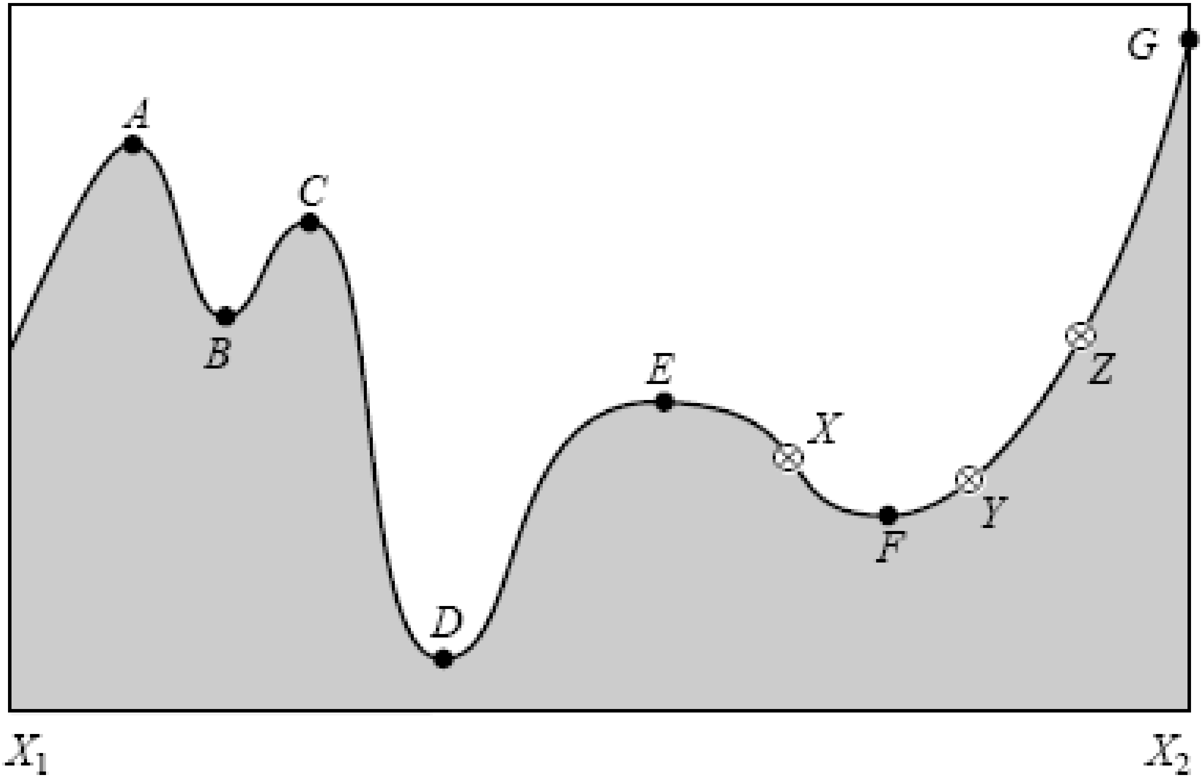
$$x_1 = \frac{q}{a}, \quad x_2 = \frac{c}{q}$$

  - Derivatives: choose stepsize eps (f(x+eps)-f(x))/eps correctly
    - Still looses half of significant digits!
    - When you need many derivatives, it is better to use Chebyshev
  - Roots
    - Bracket them
    - Bisection – failsafe
    - Secant/false position – more rapid convergence
    - Newton-Rhapson – when derivative is known analytically. Polish up.
    - Van Wijngaarden-Dekker-Brent – make quadratic interpolation with x-coordinates as function of the f(x) function values – interchange y and x
    - Polynomial : deflation (forward for small roots backwards for large roots), or use Laguerre
    - Multiple dimensions: tough

Minimization of functions

- Maximization: same as minimization of $g(x)=-f(x)$
- Choose between methods where $f'(x)$ can be calculated or not
  - (multi-dimensional case: all gradients of $f(x)$ needed)
- search for global minimum can be tough
  - e.g. traveling salesmen problem
- Analogy of bi-section method in root finding: Golden section search.
  - bracket minimum
  - (triplet of points $x_2<x_1$ and $x_2<x_3$)
  - choose new midpoint

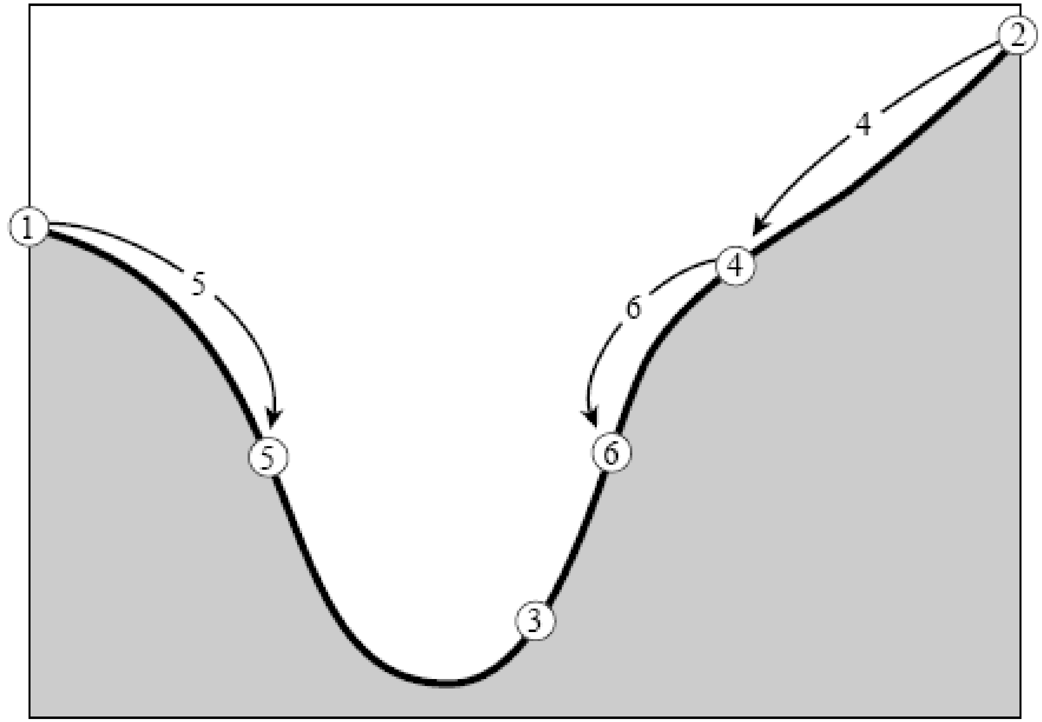
Bracketing



- X,Y,Z bracket local minimum F
- if $y_3<y_2<y_1$: replace x_2 to $x>x_3$ (e.g. $x=x_3+1.681*(x_3-x_1)$) until new sets of points bracket minimum

Golden section search

- Suppose minimum is bracketed, hunt it down.
- Where to choose next point?
  - points a<b<c. Choose point x e.g. between b and c.
  - if (f(b)<f(x), new sets of points is a<b<x
  - if (f(b)>f(x), new set is b,x,c



Minima

- Golden section: converges linearly, like bi-section. However, the interval is reduced not a factor of 2 but a factor of 1.618.. per step. About 4 steps gives additional digit of accuracy.
- In NR3: struct Golden, from include file mins.h
- How well can you determine the minimum?
  - $f(x) \sim f(b)+1/2\, f''(x)\, (x-b)^2$
  - $f(x)-f(x-b)$ precision eps.
  - $|x-b| \sim \mathrm{sqrt}(eps)$
  - relative precision of minimum: about 10^{-7}

Parabolic interpolation

- Brents method, analogous to root finding.
- make trial parabola through 3 bracketing points.
- jump to minimum of this parabola.
- only fails if points are on a line, you can safeguard against that by in such a case taking a golden section step instead.
- In NR3 : struct Brent in mins.h

```
Brent br;
br.bracket(a,b,func);
double min=br.minimize(func);
```

Derivatives

- Possibility: find the roots of the derivatives; they will equal local minima, maxima or bend points (x^3).
- I prefer that, in case of a numerically **exact** calculation of the derivative.
- Sometimes, you do not know the derivative as accurately as the function, (due to round-off errors or truncation errors, or when the function does not have a well-defined derivative).
- Method Dbrent in Numerical Recipes is very conservative:
  - Keep minimum bracketed
  - Derivative at midpoint b determines which interval will be intersected, a-b or b-c
  - Value of derivative and second-best point are interpolated to zero by secant method to get next trial point
  - If this point must be rejected: take bisection step instead

Multiple dimensions

- Problem: how to bracket the minimum?
- Self-contained strategy without derivatives: Downhill Simplex method.
- Simplex: N dimensions, N+1 points plus interconnecting line segments, surfaces etc. (e.g. 2 dim triangle, 3 dim tetrahedron).
- If nondegenerate, you can take a point as the origin and the other N points define N direction vectors spanning the N-dimensional space
- Starting simplex : e.g. take P_0 and N points $P_i=P_0+c^*e_i$, where e_i is the i-th unit vector.
- Downhill simplex method now takes a number of steps, reflecting the largest value through its opposing surface
- Followed by expansion steps and contraction steps.

Downhill simplex method

- steps: reflection of the highest point, replace if new point is lower
- If a “valley floor” is reached: contract towards lowest point.
- Simplex method will zoom in to a local minimum

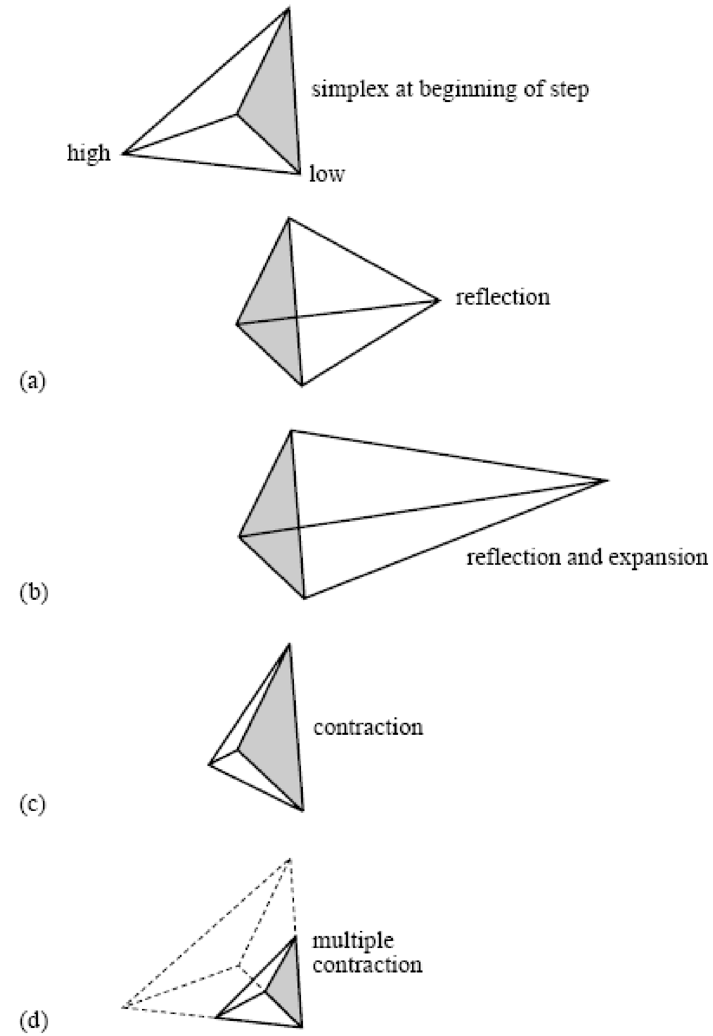


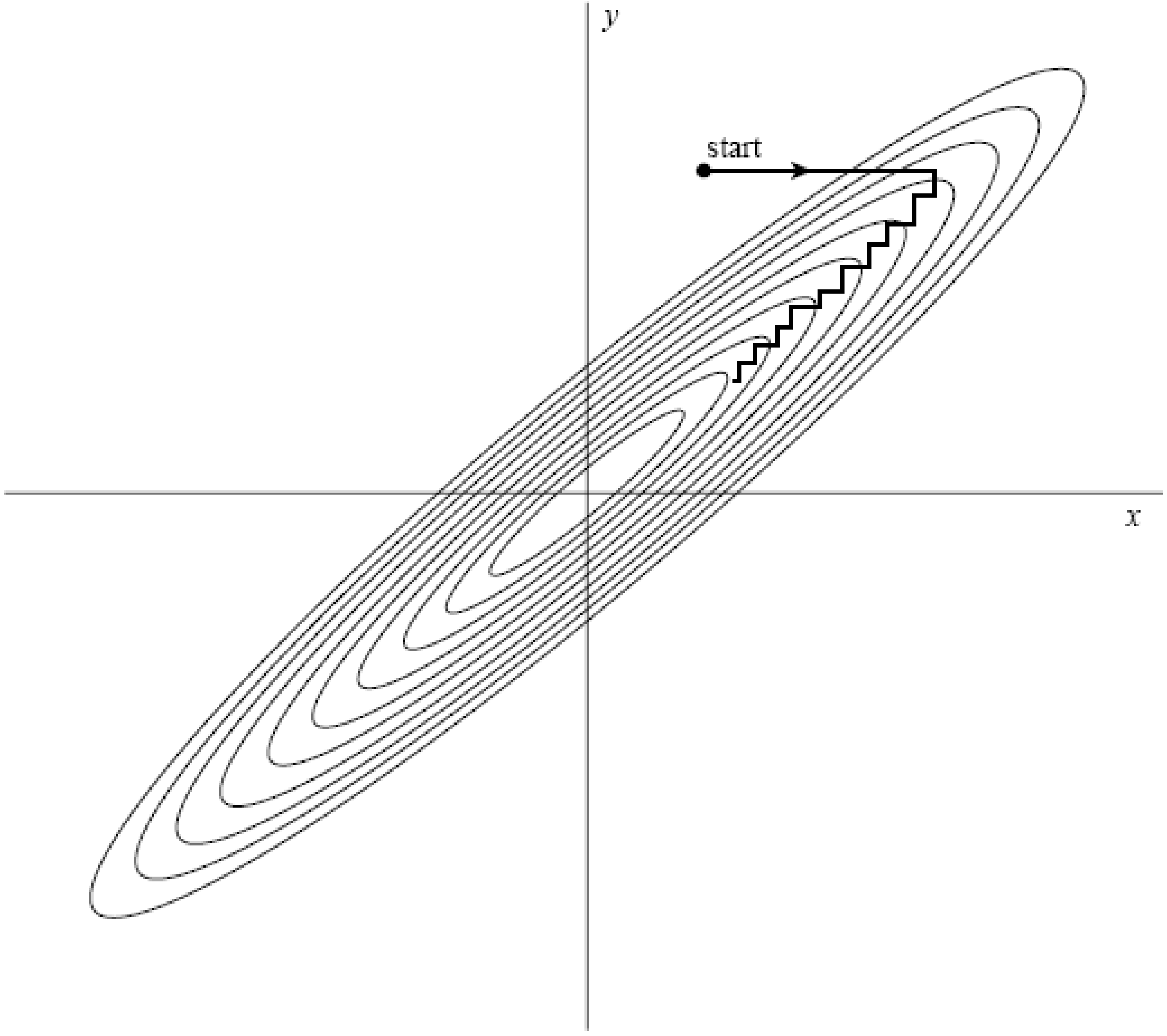
Figure 10.4.1. Possible outcomes for a step in the downhill simplex method. The simplex at the beginning of the step, here a tetrahedron, is shown, top. The simplex at the end of the step can be any one of (a) a reflection away from the high point, (b) a reflection and expansion away from the high point, (c) a contraction along one dimension from the high point, or (d) a contraction along all dimensions towards the low point. An appropriate sequence of such steps will always converge to a minimum of the function.

Powell’s Method

- Black-box routine (linmin) minimizes a function along 1 direction

$$\begin{aligned} &linmin: \quad Give \quad input \quad \vec{P} \quad , \quad \vec{n} \\ &minimize \quad f\left(\vec{P}+\lambda\vec{n}\right) \\ &new\,vector: \quad \vec{P}'=\vec{P}+\lambda\vec{n} \end{aligned}$$

- methods: differ in choice of next direction for next step. Simplest approach: minimize first along direction e1, then e2, etc.
  - Linmin in NR3 : in mins_ndim.h
- Might fail: conjugate directions


y
start
x

Conjugate directions

$$f(\vec{x}) = f(\vec{P}) + \sum_i \frac{df}{dx_i} x_i + \frac{1}{2} \sum_{i,j} \frac{d^2 f}{dx_i dx_j} x_i x_j + \ldots$$

$$f(\vec{x}) \approx c - \vec{b} \cdot \vec{x} + \frac{1}{2} \vec{x}^T \cdot A \cdot \vec{x}$$

$$c = f(\vec{P}), \quad b = -\nabla f|_P \quad , \quad [A]_{ij} = \frac{d^2 f}{dx_i dx_j}|_P$$

- Taylor expansion around P. For the mininum, the first order derivatives (the gradient) disappears:

$$\nabla f = A \cdot \vec{x} - \vec{b}$$
$$solve \quad A \cdot \vec{x} = \vec{b}$$

- Gradient along direction u changes as:

$$\delta(\nabla f) = A \cdot (\delta \vec{x})$$

Conjugate directions

- After minimization along u, choose new direction v.
- In order to retain the minimization along the direction u,

  the vector v should be perpendicular to the gradient used in the previous step:

$$0=\vec{u}\cdot\delta(\nabla f)=\vec{u}\cdot A\cdot\vec{v}$$

- If this condition holds for the vectors u and v, they are called **conjugate**. Powell's method to arrive at a conjugate base:
- start P_0. for i=0..N-1, move P_i along direction u_i for minimum. Call this point P_{i+1}
- for i=0..N-2, replace unit vector u_i by u_{i+1}
- $u_{n-1} = P_N - P_0$
- move P_N to minimum along u_{N-1}, call this point P_0.
- Repeating this procedure N times gives a conjugate base, that exactly minimizes a quadratic function.
- Different algorithms for choosing next direction v

Using derivative information

- Instead of N^2 line minimizations as in Powells method, when the gradients are known, you can try to do it in N steps.
- steepest downhill method: step in direction of local downhill gradient –grad(f(P_i)).
- not optimal: next step is necessarily perpendicular to previous step.
- method that steps not in direction of new gradient, but in directions conjugate to the previous directions, are called conjugate gradient methods.
  - NR::frprmn (Fletcher-Reeves-Polak-Ribiere minimization), struct Frprmn

Steepest downhill method

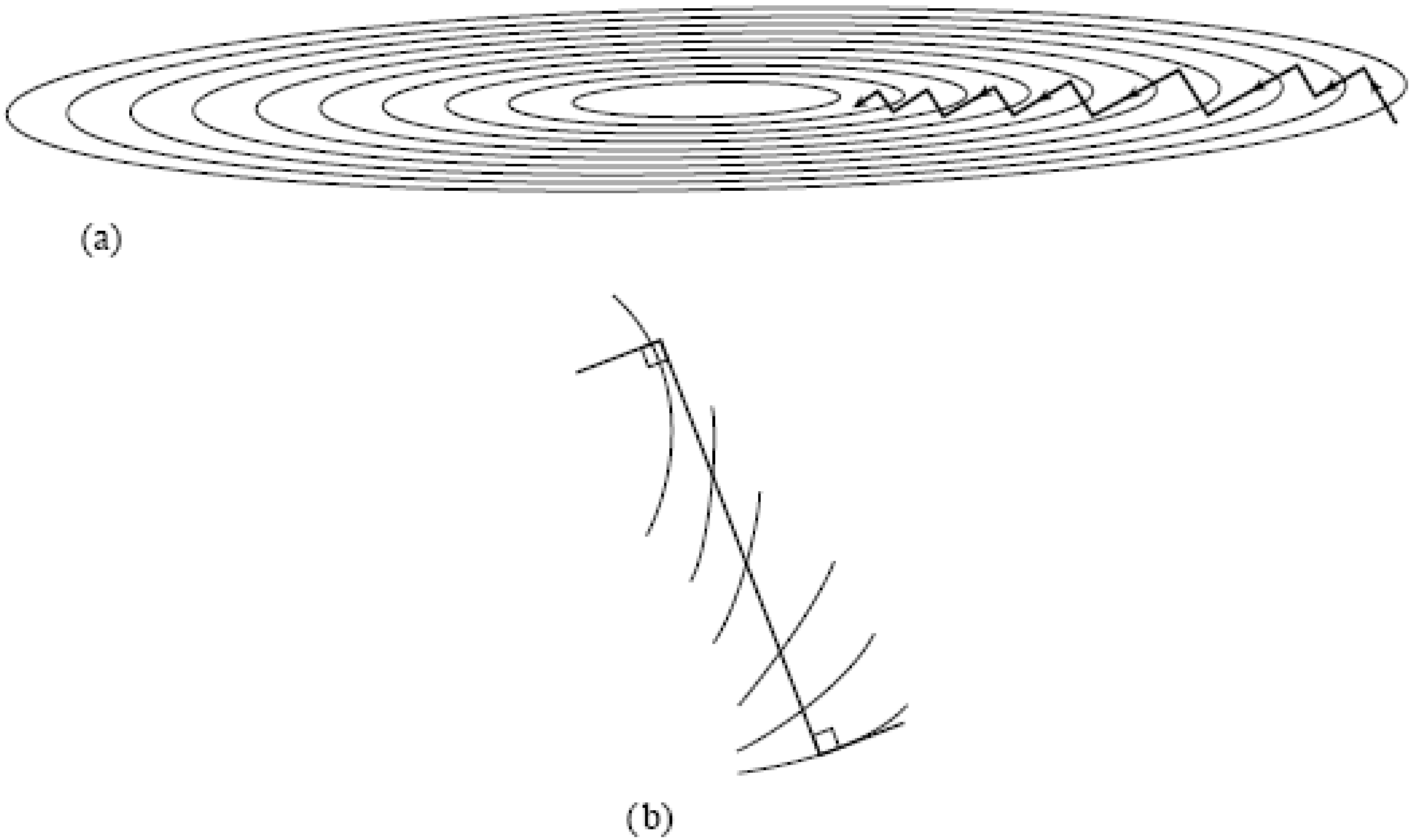


Figure 10.6.1. (a) Steepest descent method in a long, narrow "valley." While more efficient than the strategy of Figure 10.5.1, steepest descent is nonetheless an inefficient strategy, taking many steps to reach the valley floor. (b) Magnified view of one step: A step starts off in the local gradient direction, perpendicular to the contour lines, and traverses a straight line until a local minimum is reached, where the traverse is parallel to the local contour lines.

Simulated annealing

- e.g. traveling salesman problem: find the shortest route between n points, visiting each point only once. (exact solution: increases as exp(cN)).
- Objective function E describing the routes has many local minima
  - Annealing in solid states: crystallization. Rearrangements possible via Prob(E)~exp(E/kT) with k Boltzmanns constant.
  - Nature arrives at lowest energy state when cooling is slow. During this process, rearrangements of neighbouring atoms are possible. When you quench (rapidly cool), you obtain polycrystallline/amorphous states – higher energy.
- 1) starting configuration: order of N cities and their coordinates (x_i,y_i). The cost function E that is minimized is the total path length
- 2) rearrangements:
  - section is removed and replaced by same cities in opposite order
  - section is removed and inserted between other, randomly chosen, path

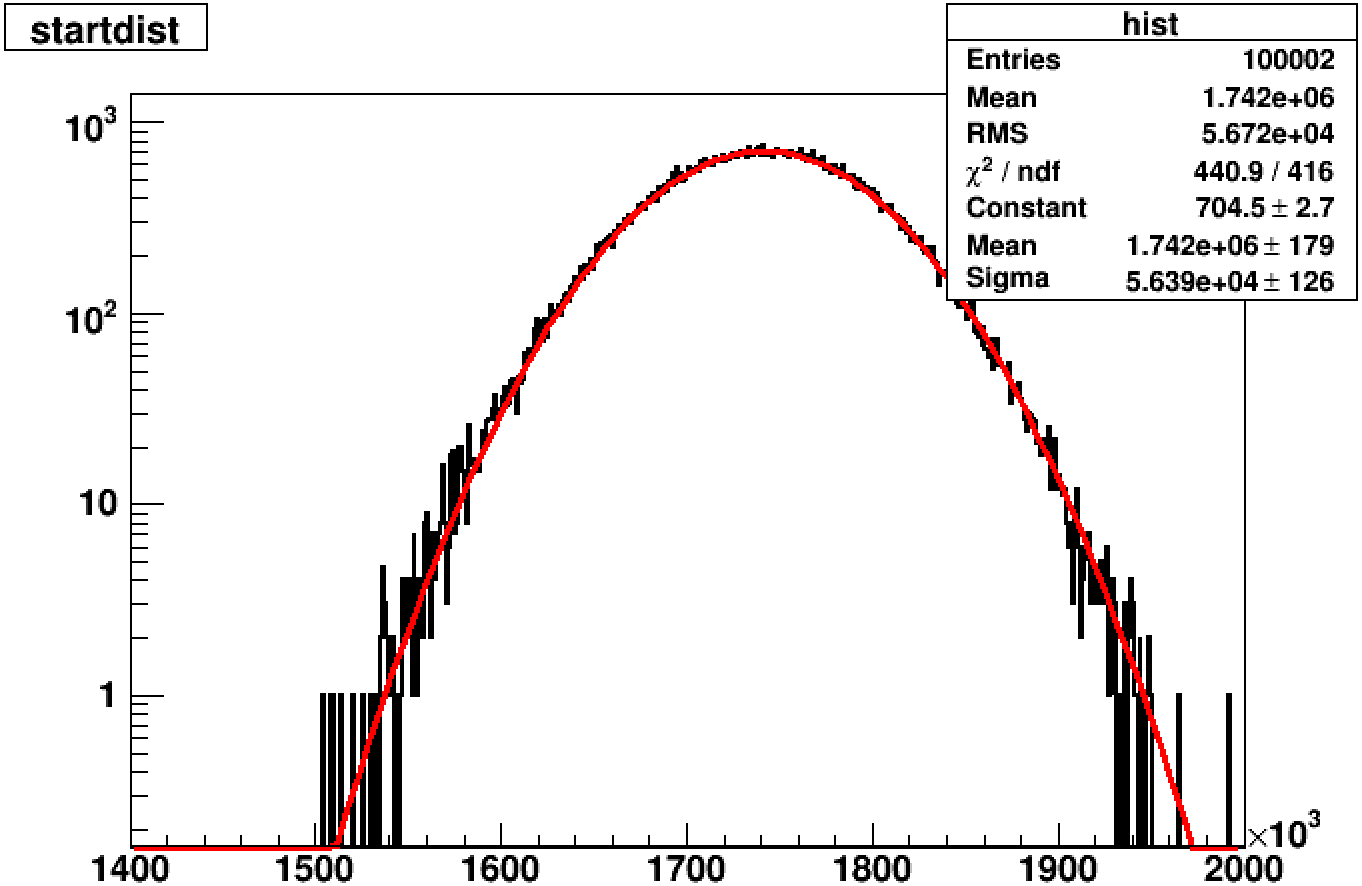
Example: shortest path 211 cities

- Find shortest roundtrip 211 cities
  - (422 equivalent paths, starting city arbitrary, direction arbitrary)
- Typical spread random start configuration?

Mean 1.74e6 km,
RMS 56e3 km

Probability finding a path at <1e6 km by random trial: negligible.

Typical distance between 2cities about 9000 km,
Start temperature of a few 1000 km seems reasonable



strategies

- Many strategies possible. E.g.
  - 0: swap 2 cities in the path (1-2-3-4-5-6-7 → 1-6-3-4-5-2-7-
  - 1: change order between 2 cities (1-2-3-4-5 → 1-4-3-2-5)
  - 2: take out a stretch of cities and reinsert it (in normal or inverted order)
  - 3: do combinations of these

- Annealing:
  - Make a random start configuration
  - Choose a start temperature T (a distance that is used to compare the distances between the paths after taking steps as above).
  - Draw random numbers to select which section of the path will be changed

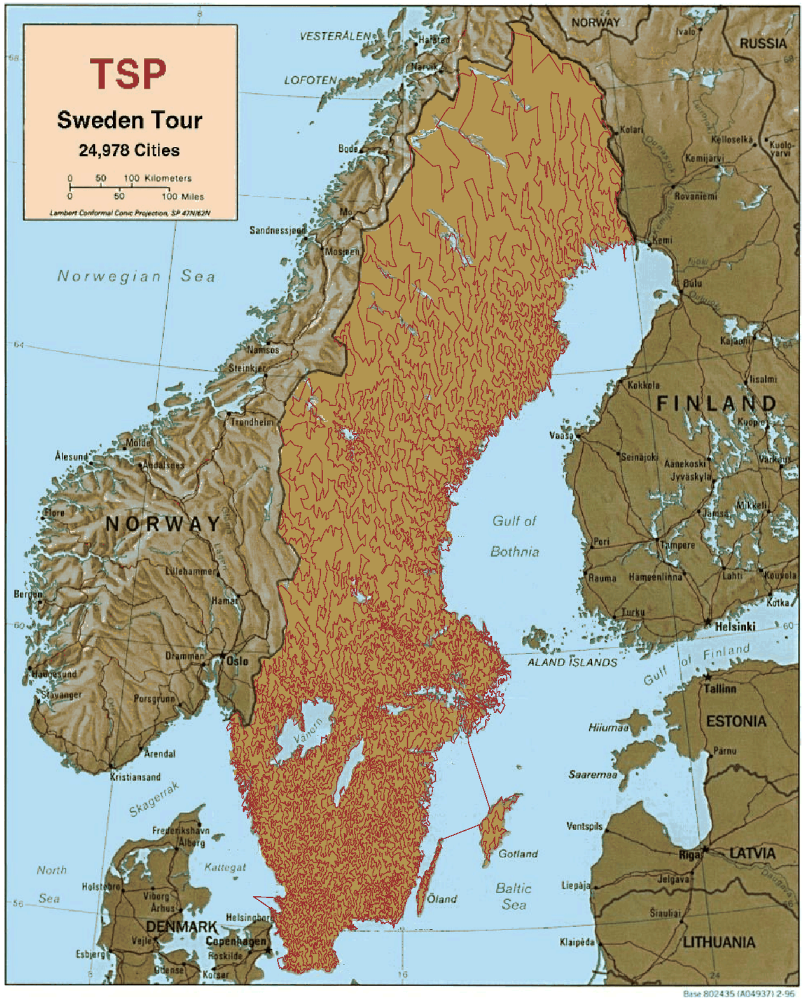
    Calculate ΔE, the difference in total path length between the old and new configuration. If $\Delta E < 0$, keep the new (shorter) path.
  - If new path is longer, then draw random number x between 0 and 1.
    - If $\exp(-\Delta E/T) < x$ then keep the longer path as the new sequence, else toss it and try a new step.

Annealing

- After a sufficient number of steps, reduce the temperature and repeat the procedure. Keep on reducing the temperature until it is much shorter than the average distance in 2 paths.
- Always store the shortest path; your end result is the shortest path you found, not the last path.
- Sometimes it is good to try a complete new start configuration
- At higher temperatures, large sections of the path can be updated. At low temperatures, the main changes will be in reordering of close-by cities.
- How many steps to take, and how to choose/update the temperature, is a matter of experimenting. The optimal strategy is not a priori known.

Traveling Salesmen Problem

- Better optimizations possible (TSP codes)
  - Write as a set of constraints: linear programming
    - Cutting plane method
  - Cities in Sweden
  - Mininum distance of a city to the next is given by the distance to the nearest city
  - For some short ranges you can construct optimal paths
  - Millions of conditions to be satisfied simultaneously
  - branch cutting



Linear programming

- Many large packages, used a lot in for instance econometrics. For n independent variables, minimize the function

$$\zeta = c_1 x_1 + \dots + c_n x_n$$

- With conditions

$$x_i \geq 0$$

- And satisfying m constraints

$$a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n \leq b_i \quad or$$
$$a_{i1}x_1 + a_{i2}x_2 + \dots + a_{in}x_n = b_i$$

- If an independent variable follows a condition y<b, replace it with x=b-y.
- A vector satisfying the constraints is called a feasible vector. The function to be minimized is the objective function. The vector that minimizes it is called the optimal feasible vector.
- There may not be any feasible vector (incompatible constraints) or there may be no minimum (e.g. variable i-> inf)

Linear programming

10.10 Linear Programming: The Simplex Method 527

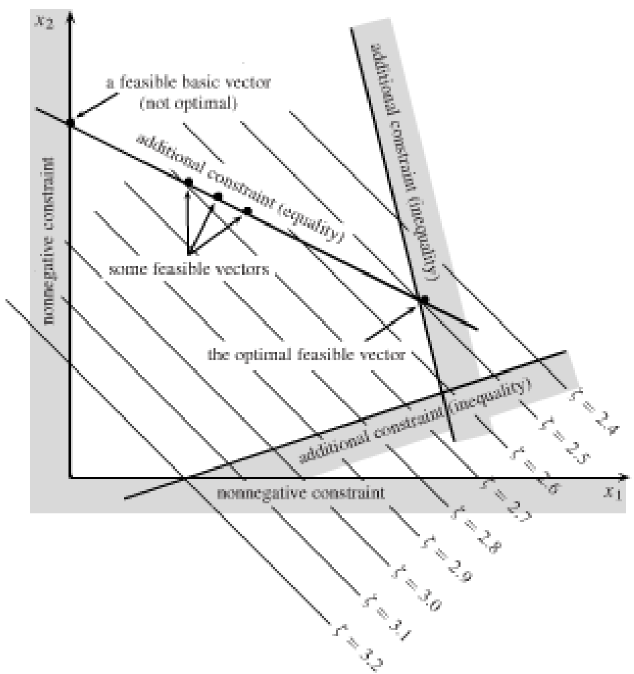


Figure 10.10.1. Basic concepts of linear programming. The case of only two independent variables, x_1, x_2, is shown. The linear function ζ, to be minimized, is represented by its contour lines. Nonnegativity constraints require x_1 and x_2 to be positive. Additional constraints may restrict the solution to regions (inequality constraints) or to surfaces of lower dimensionality (equality constraints). Feasible vectors satisfy all constraints. Feasible basic vectors also lie on the boundary of the allowed region. The simplex method steps among feasible basic vectors until the optimal feasible vector is found.

- Linear programming:

  example with 2 variables. The volume containing feasible vectors is given by a 2-dimensional area. Constraints form the boundaries of this surface(1-dimensional lines). The problem is linear: the minimum will be located at the surface (gradients are constant)

- Start from origin, minimizing x_2 for 0 x_1 leads to first feasible vector. Sliding down the boundary until the second constraint is hit leads to the optimal feasible vector, the solution.

- For higher dimensions, slide down the N-1 dimensional surface spanned by the boundary conditions

- Large matrices involved, specialized code necessary.

Linear programming

- Example: minimize $\zeta = -40x_1 - 60x_2$

  subject to:
  $$2x_1 + x_2 \leq 70$$
  $$x_1 + x_2 \geq 40$$
  $$x_1 + 3x_2 = 90$$

  Introduce slack variables to write <= inequalities in form of equalities:
  $$2x_1 + x_2 + x_3 = 70$$
  $$-x_1 - x_2 + x_4 = -40$$

  Find start feasible vector: introduce another parameter to write the third equality as a <= condition.

  $$x_1 + 3x_2 + x_5 = 90$$
  $$x_1 = x_2 = 0, \quad x_3 = 70, \quad x_4 = -40, \quad x_5 = 90$$

Linear programming

- Problem: x_4 is negative: add auxiliary objective function

  $\zeta' = -x_4$

- In phase 1, we start with the base $x_1 = x_2 = 0$ and minimize this auxiliary objective function, leading to $x_4 = 0$.
- If phase 1 fails (not all x are >=0) it signifies that there is no feasible vector: the constraints are conflicting.
- Else: write in matrix form and slide down the surfaces until the optimal vector is found:

Example linear programming

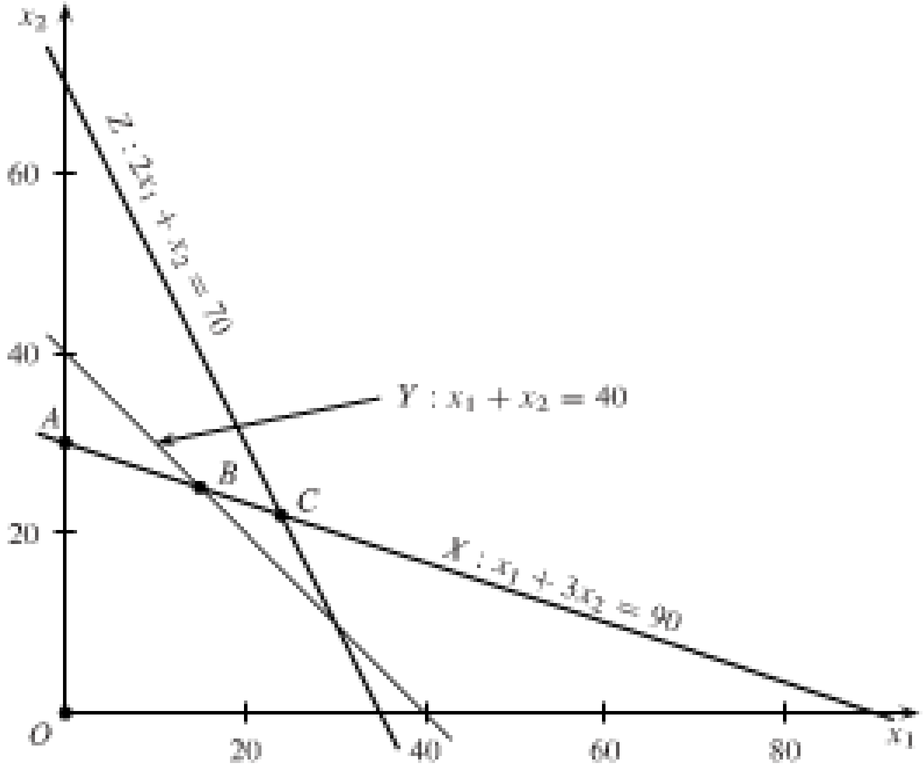


Figure 10.10.2. Graphical interpretation of the simplex solution of the problem (10.10.5) – (10.10.8). The initial basic vector is at the origin O. To satisfy the equality (10.10.8), the first step moves to A, on the line X. This is not yet a feasible point, since it is on the wrong side of the line Y. The next move is to B, which is feasible. We enter phase two, and find that we can reduce the objective function by moving to C. No further moves are possible, so we are done. Note that the figure is really the projection of a five-dimensional simplex onto the x_1-x_2 plane.