**FIFTH INTERNATIONAL CONFERENCE ON HYDROINFORMATICS**

1- 5 July 2002, Cardiff, UK

**C05 - Evolutionary algorithms in hydroinformatics**

# An evolutionary annealing-simplex algorithm for global optimisation of water resource systems

**Andreas Efstratiadis and Demetris Koutsoyiannis**
Department of Water Resources
National Technical University of Athens

## Parts of the presentation

1. Outline of the global optimisation problem
2. Overview of global optimisation techniques
3. The evolutionary annealing-simplex algorithm
4. Evaluation of the algorithm
   - Mathematical applications
   - Real-world applications
5. Conclusions

## The global optimisation problem

Posing the problem

Find a real vector $\mathbf{x}^*$, defined in the n-dimensional continuous space D = [**a**, **b**] $\subset R^n$, that minimises a real, nonlinear function f, i.e.:

$$f(\mathbf{x}^*) = \min f(\mathbf{x}),\ \mathbf{a} < \mathbf{x} < \mathbf{b}$$

Main assumptions

1. The objective function is non-convex
   - *Due to non-convexity, the search space is rough and multimodal*
2. No external constraints are imposed to the problem
   - *The mathematical constraints are handled either through penalty methods or via simulation*
3. The analytical expression of the objective function is unknown
   - *Apparently, the analytical expression of the partial derivatives is also unknown*



## Troubles encountered

1. Convergence to a local optimum

   Generally, it is relatively easy to locate a local optimum, but very difficult or even impossible to get out of it
2. Extremely large number of trials to locate the global optimum

   To avoid getting trapped by local optima, a detailed exploration of the search space may be required
3. The curse of dimensionality

   The theoretical time to solve a nonlinear problem increases even exponentially with its dimension
4. The practical aspect of real-world applications

   In real-world problems, a highly accurate solution is neither **possible**, because of uncertainties and inaccuracies in the underlying model or data, nor **feasible**, due to the unacceptable high computational effort required to attain it when the function evaluation is time consuming

## Local vs. global optimisation

| | Local methods | Global methods |
|---|---|---|
| Evolving "pattern" | Single point (usually) | Random sample (population) of points |
| Transition rules | Deterministic (subsequent line minimisations) | Deterministic and stochastic |
| Location of the global optimum | Not guaranteed | Asympotically guaranteed |
| Typical categories of algorithms | Gradient-based methods (steepest descend, conjugate gradient, Newton, quasi-Newton)<br>Direct search methods (downhill simplex, rotating directions) | Set covering, pure, adaptive & controlled random search, two-phase algorithms (multistart), evolutionary & genetic algorithms, simulated annealing, tabu search, heuristics |



## Effectiveness vs. efficiency

| | Effectiveness | Efficiency |
|---|---|---|
| Definition | Probability of locating the global optimum, starting from any random initial point (or population of points) | Convergence speed |
| Performance measure | Number of successes out of a predefined number of independent runs of the algorithm | Average number of function evaluations to converge |
| Examples and counter-examples | The exhaustive character of grid search methods ensures high probability of locating the global optimum; however this usually requires an extremely large number of function evaluations | The gradient-based concept of local search methods ensures quick convergence to the nearest optimum, without guaranteeing that this is the global one |

## Nonlinear optimisation: The story so far

| | |
|---|---|
| after 1990 | The era of integration: old methods become popular again, diverse approaches are combined into heuristic schemes |
| 1980–1990 | A variety of global optimisation methods are developed, following the expansion of "artificial intelligence" |
| 1970–1980 | Direct search methods are gradually abandoned<br>Evolutionary algorithms come to the fore |
| 1960–1970 | Dominance of direct search (derivative-free) methods<br>Development of primitive stochastic optimisation methods |
| 1950–1960 | Foundation of the modern nonlinear optimisation theory<br>Development of gradient-based algorithms |
| before 1950 | Classical approach, based on methods of calculus |



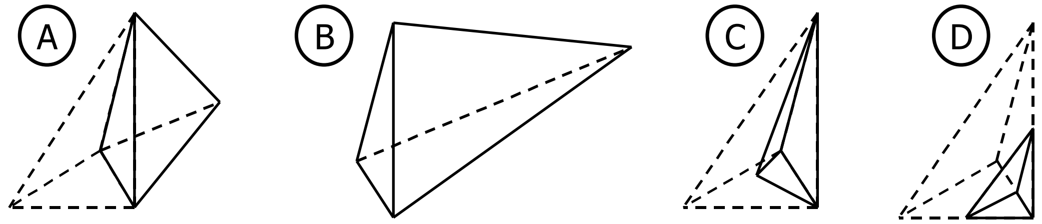
## The downhill simplex method (*)

General concept

Instead of improving the objective function by calculating or approximating its gradient, the direction of improvement is selected just by comparing the function values at adjacent points, using an evolving "pattern" of n+1 points that span the n-dimensional space

Possible simplex operations

- Reflection (A) through the actual worst vertex
- Expansion (B) or contraction (C) along the direction of reflection
- Shrinkage (D) towards the actual best vertex




(*) Nelder, J. A., and R. Mead, A simplex method for function minimization, *Computer Journal*, 7(4), 308-313, 1965

## Stochastic optimisation techniques

Pure random search

A pre-specified number of points are randomly generated into the feasible space, and the best of them is taken as an estimator of the global optimum

Adaptive random search

Each next point is generated as a random perturbation around the current one, and it is accepted only if it improves the function

Multistart strategy

A local optimisation algorithm is run several times, starting from different, randomly selected locations (in an ideal case, we wish to start at every region of attraction of local optima)

Due to their almost exclusively random character, stochastic methods attain to escape from local optima and effectively cope with rough spaces, but, on the other hand, the lack of determinism within the searching procedure leads to a very slow convergence rate



## Evolutionary algorithms

Inspiration

Modelling the search process of natural evolution

Main concepts

- Representation of control variables on a chromosome-like (usually binary string) format
- Search through a population of points, not a single point
- Application of genetic operators to create new generations

Genetic operators

- The **selection** operator aims at improving the average genetic characteristics of the population, by providing higher probability of surviving to the better of its individuals
- Through the **crossover** operator, two "parents" exchange part of their characteristics, to generate more powerful "offsprings"
- The **mutation** operator aims at introducing new characteristics into the population, in order to enhance its diversity

## The simulated annealing strategy

The annealing process in thermodynamics

For *slowly* cooled thermodymnamical systems (e.g., metals), nature is able to find the minimum energy state, while the system may end in an amorphous state, having a higher energy, if it is cooled quickly

Energy minimisation strategy

The system is allowed sometimes to go "uphill", so as to get out of a local energy minimum in favor of finding a better, more global one

Mathematical formulation

The system's energy state E depends on the actual temperature T and it is distributed according to the Boltzmann probability function:

$$\text{Prob}(E) \sim \exp(-E / k\,T)$$

Simulated annealing and global optimisation

The annealing strategy was transferred into optimisation algorithms by introducing a control parameter T (analogue of temperature), an annealing cooling schedule, describing the gradual reduction of T, and a stochastic criterion for accepting non-optimal (uphill) moves



## The shuffled complex evolution method (*)

General concept

Instead of using multiple simplexes starting from random locations and "working" fully independently, it would be more efficient to let them both evolving individually and exchanging information (an analogous with a research project, where many scientists, even working independently or in small groups, they organise frequent meetings to discuss their progress)

Description of the algorithm

- A random set of points (a "population") is sampled and partitioned into a number of **complexes**
- Each of the complexes is allowed to evolve in the direction of global improvement, using competitive **evolution** techniques that are based on the downhill simplex method
- At periodic stages, the entire set of points is **shuffled** and reassigned to new complexes, to enable information sharing

(*) Duan, Q., S. Sorooshian, and V. Gupta, Effective and efficient global optimization for conceptual rainfall-runoff models, *Water Resources Research*, 28(4), 1015-1031, 1992

## The concept of annealing-simplex methods

**Simulated annealing**

Advantage: Capability of escaping from local optima, by accepting some uphill moves, according to probabilistic criteria (**effectiveness**)

Disadvantage: Too slow convergence rate (the slowest the convergence, the most probable to locate the global optimum)

**Downhill simplex method**

Advantage: Quick and easy location of the local optimum, in the region of attraction of which the starting point is found (**efficiency**)

Disadvantage: Failing of getting out of a local optimum after converging to it – Bad performance in case of rough or ill-posed search spaces

Incorporation of a simulated annealing strategy within a downhill simplex searching scheme



## The evolutionary annealing-simplex algorithm

The motivation

Formulation of a probabilistic heuristic algorithm that joins concepts from different methodological approaches, in order to ensure both effectiveness and efficiency

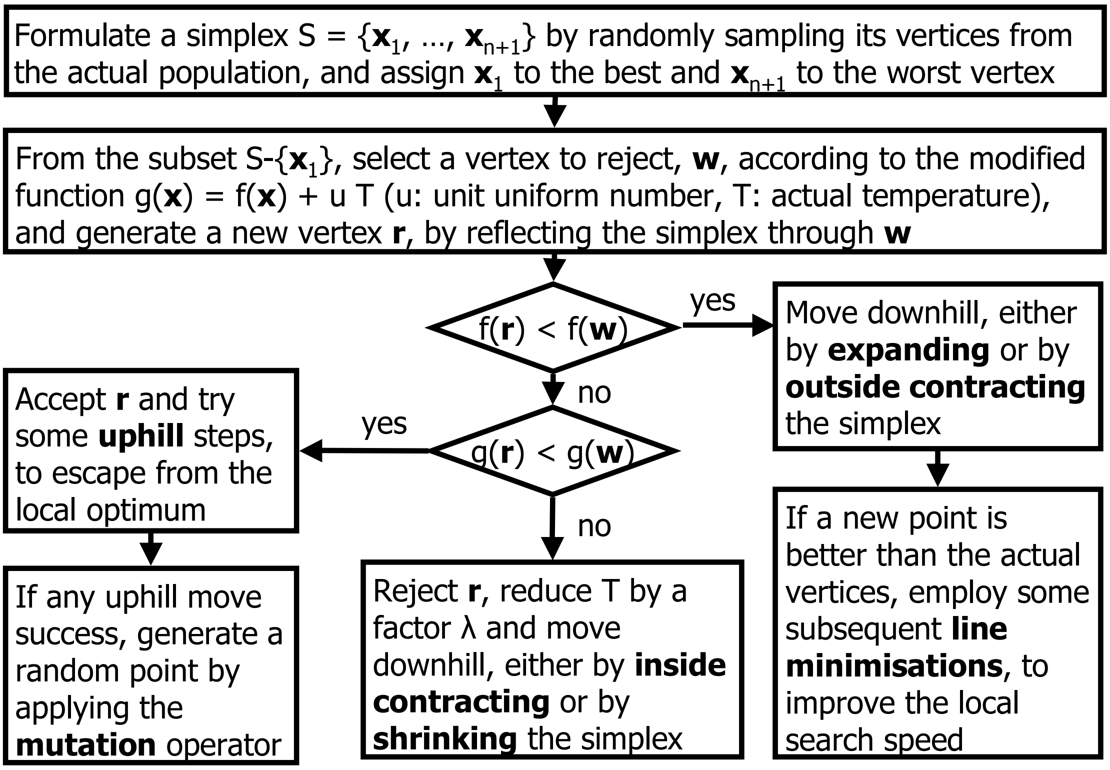
Main concepts

- An **evolutionary** searching technique is introduced
- The evolution is made according to a variety of combined (deterministic and stochastic) transition rules, most of them based on the downhill **simplex** scheme
- An adaptive **annealing** cooling schedule regulates the “temperature”, which determines the degree of randomness through the evolution procedure

Input arguments

- the size of the population, m
- the parameters of the annealing schedule
- the mutation probability, $p_m$

# Flowchart of a typical iteration step






# Further investigation (1)

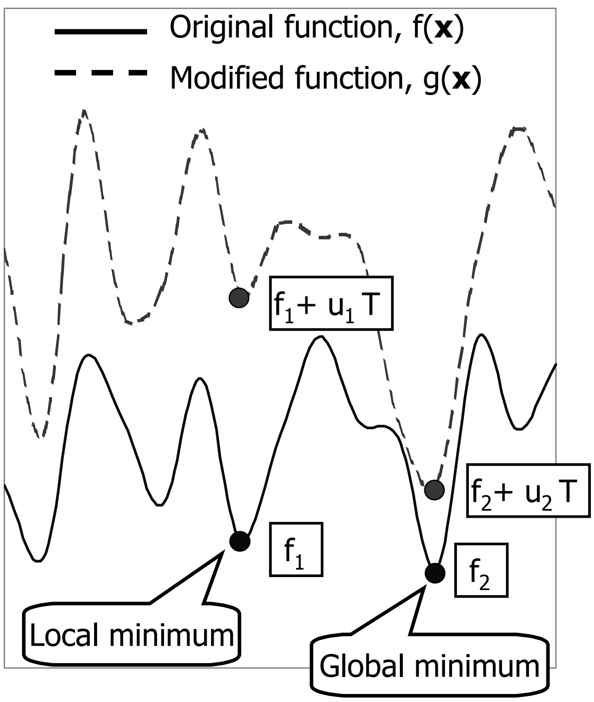
## The modified objective function

“Coupling” of determinism and randomness

By adding a stochastic component to the objective function (relative to the actual temperature), the algorithm behaves as between random and downhill search

Transforming the search space

The use of the modified function as the criterion to accept or to reject a newly generated point, provides more flexibility, either by “smoothing” a rough search space or by “aggravating” a flat surface



# Further investigation (2)

## The generalised downhill simplex procedure

The quasi-stochastic operator

In order to increase randomness, the generation of new vertices within the reflection, expansion and contraction steps is implemented according to a transition rule of the form:

$$\mathbf{x}_{new} = \mathbf{g} + (a + b\,u)\,\mathbf{d}$$

where **g** is the centroid of the simplex, a and b are appropriate scale parameters, **d** is the direction of improvement, which is specified according to the original Nelder-Mead formula, and u is a unit uniform number (for u = 0.5, the generalised and the original Nelder-Mead methods become identical)

A deeper insight

The implementation of simplex movements, the lengths of which are randomised, prohibits "recycling" of the simplex to the same vertices



# Further investigation (3)

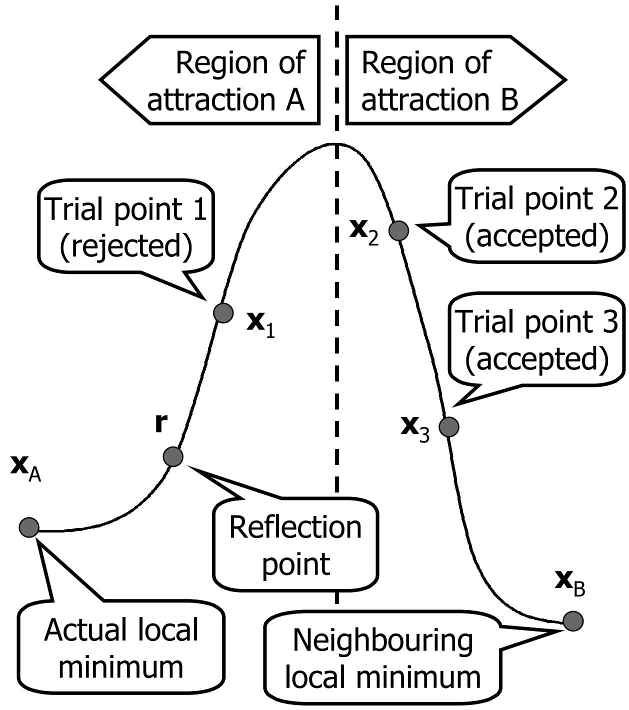
## Escaping from local optima by hill-climbing

The hill-climbing strategy

According to a probabilistic criterion, a pre-specified number of subsequent uphill moves may be taken along the direction of reflection, in order to surpass the "ridge" and explore a neighbouring "valley", where another local optimum is located

The acceptance criterion

To accept a new point, it just suffices to check if it is better than the previous one, which guarantees the crossing of the ridge



# Further investigation (4)

## The adaptive annealing cooling schedule

The role of temperature

At the initial stages, we prefer high values of T, in order to easily escape from local optima, whereas, after the global optimum is approached, we prefer low values to accelerate convergence

The importance of an appropriate annealing schedule

Very large values of T reduce drastically the efficiency of the algorithm, while very low ones reduce its effectiveness (the algorithm becomes too deterministic)

The adaptive strategy

To avoid extremely high temperatures, at the beginning of each iteration cycle, T is regulated according to the rule:

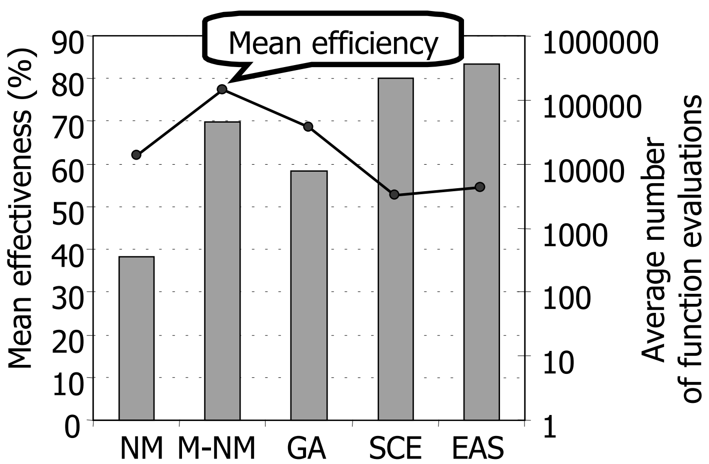
$$T \leq \xi\,(f_{max} - f_{min}),\ \xi \geq 1$$

On the other hand, whenever a local optimum is reached (this is recognised by the fact that the simplex volume decreases), T is "slightly" reduced by a factor λ (typically λ = 0.90-0.99)



# Mathematical applications

- 8 typical benchmark functions were tested, by implementing 100 independent runs for each one
- The performance of 5 optimisation techniques was evaluated (in terms of effectiveness and efficiency) via Monte Carlo simulation
- The influence of several algorithmic parameters was examined
- The best performance was indicated by the SCE and the EAS methods




Abbreviations

NM: Nelder-Mead method

M-NM: Multistarts (20) of the Nelder-Mead method

GA: Simple, binary-coded, genetic algorithm

SCE: Shuffled complex evolution method

EAS: Evolutionary annealing-simplex algorithm

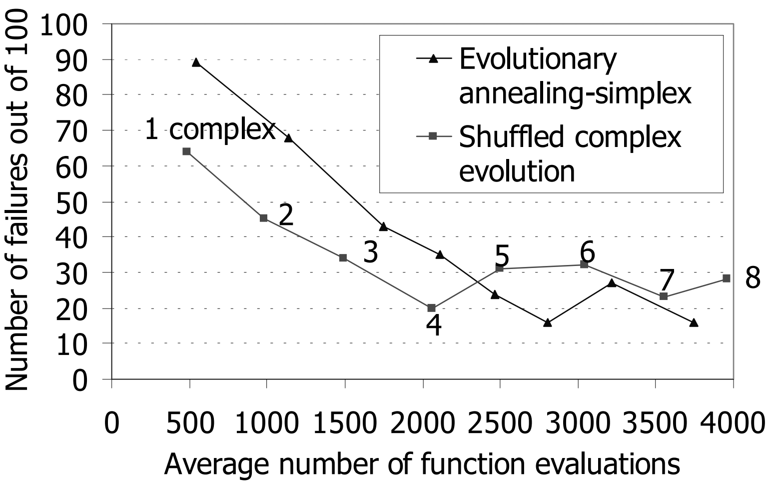
## Real world applications: Brief description

| | Problem A | Problem B |
|---|---|---|
| Objective | Calibration of a simple water balance model, based on a Thornthwaite approach | Maximisation of the mean annual energy profit of a hypothetical system of two parallel reservoirs |
| Control variables | 6 (= 4 parameters of the model and 2 initial conditions) | 384 (= the target releases from each reservoir, for a monthly simulation period of 16 years) |
| Main difficulties | The conceptual character of the model, the quality of data, the interdependence between parameters, the introduction of bias, due to the use of a single "measure" for calibration | The flat-type response surface, due to the use of desirable and not real magnitudes as control variables, which makes extremely difficult the location of the gradient |



## Real world applications: Main conclusions

- The SCE method and the evolutionary annealing-simplex algorithm faced with success both real-world problems; the latter was slightly more effective (fast), while the former was slightly more efficient
- The role of the population size was proved particularly crucial, affecting drastically the performance of the two algorithms




An interesting note
In the model calibration problem, for both algorithms, there exists an "optimal" population size (determined experimentally), for which effectiveness is maximised

## Conclusive remarks

1. The current trend in global optimisation is the combination of ideas obtained from diverse methodological approaches, even the old ones
2. Heuristic schemes manage to handle the typical shortcomings of global optimisation applications, by effectively "balancing" determinism and randomness within the searching procedure
3. A specific characteristic of heuristic methods is the existence of a variety of input arguments that strongly affect their performance (e.g., the population size), and have to be specified by the user
4. The proposed evolutionary annealing-simplex algorithm uses as basis a Nelder-Mead scheme and improves its flexibility by introducing new types of simplex movements; moreover, the adaptive annealing schedule regulates randomness and provides more chances to escape from local optima and handle hard search spaces
5. To increase the effectiveness and efficiency of the proposed method, several improvements could be made towards the parallelisation of the algorithm, the development of criteria for automatic regulation of its input arguments and the incorporation of specific rules to handle the multi-objective nature of most hydroinformatics applications