Continuous (meta-)optimization

Vladimir Bicik

Computational Intelligence Group
Faculty of Information Technology
Czech Technical University in Prague

Advisor: Ing. Pavel Kordik, Ph.D.

17.12.2011

Introduction

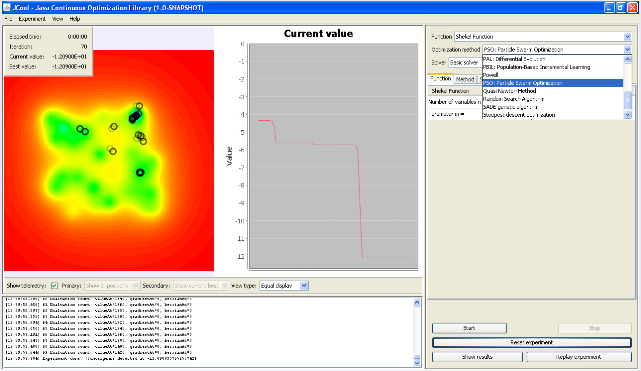
Continuous Optimization

- Seeks a (global) minimum of an arbitrary continuous function
- The function is usually complex, multimodal and multidimensional
- Usually an analytical gradient is available, but not always
- Even less usual is analytical Hessian
- The function is considered a black box $\Longrightarrow$ black-box optimization
- Many different approaches, exhaustively mapped

Application Environment

JCool

- Project resulting from M. Hvizdos' Master's Thesis
- Testing and benchmarking of optimization methods
- Currently contains 18 methods and 33 testing functions



Gradient Methods

- Different use of the Hessian matrix:
  1. Conjugate Gradient: does not use at all
  2. Levenberg-Marquardt: uses and adjusts
  3. *quasi*Newton: not directly, approximates

4. Orthogonal search – optimization dimension by dimension
5. Powell's method – improved OS by folding the already taken steps

Covariance Matrix Adaptation Evolution Strategy

- Sampling of a normal distribution of a multidimensional vector
- Covariance matrix used to describe dependence between parameters

Ant Colony Algorithms

- Directly simulate ant behaviour (CACO, API)
- Extension of the original algorithm by discretization (AACA)
- Extension of the original algorithm by probabilistic sampling (ACO*, DACO)

Genetic Algorithms

- Differential Evolution (DE, SADE)
- Vector of probabilities used to sample the population (PBIL)

- Simulation of a bevy in search of food (PSO)
- Combined algorithms (HGAPSO)

Suite of test functions

- Unimodal and multimodal functions
- Multidimensional functions, many of which configurable by parameters
- For the most of these analytical gradient and Hessian is available
- Values of the global minima are known, including their positions

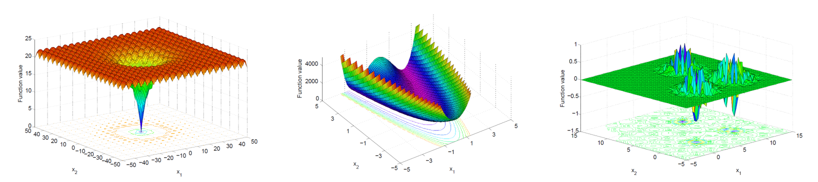


Figure: Examples of implemented test functions.

Benchmarking

- 100 runs, limit to 2000 iterations
- Each parameter tested in it's full range
- Success rate and # of iterations recorded

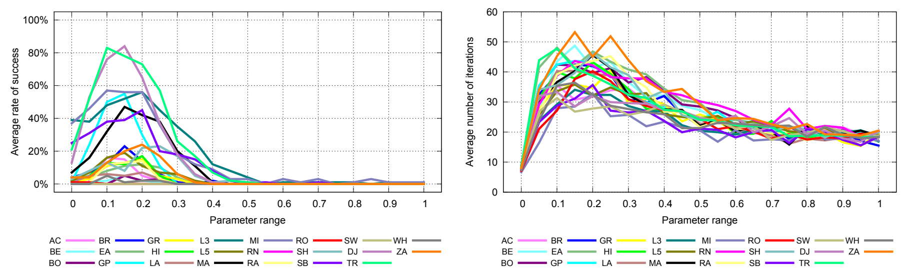


Figure: PBIL, likelihood of a mutation, step-size 0.05

Benchmarking

- Recommended values of optimization method parameters
- Different sets of parameter values for different function types

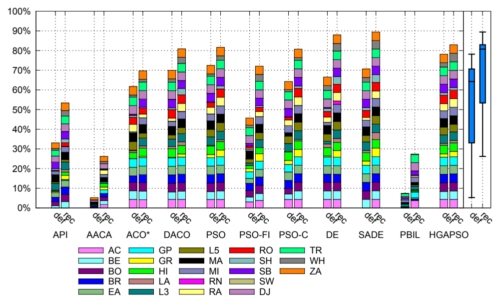


Figure: Comparison of the original and recommended parameter values

Numeric Methods

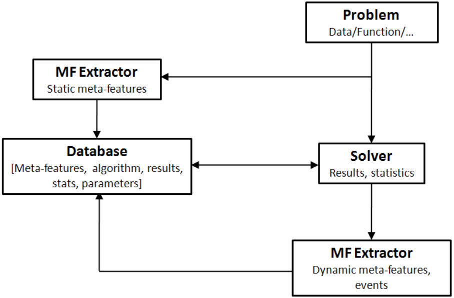
- Very precise
- More effective
- Poor global convergence
- Useful for landmarking

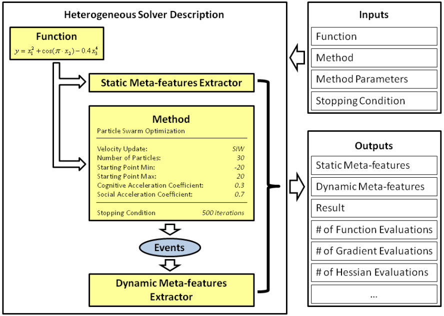
Nature Inspired Methods

- Although less precise, these can handle hard functions
- Time demanding computation, more iterations needed

Optimization of Optimization

- Technique of identifying the best algorithm for the given task
- *No free lunch theorem* (Wolpert and Macready, 1997)
- Once identified, optimal parameter values should be supplied as well
- Ultimately a repository will be created, storing
  - meta-features
  - algorithm + parameter values
  - achieved results
- Optimizing GAME models

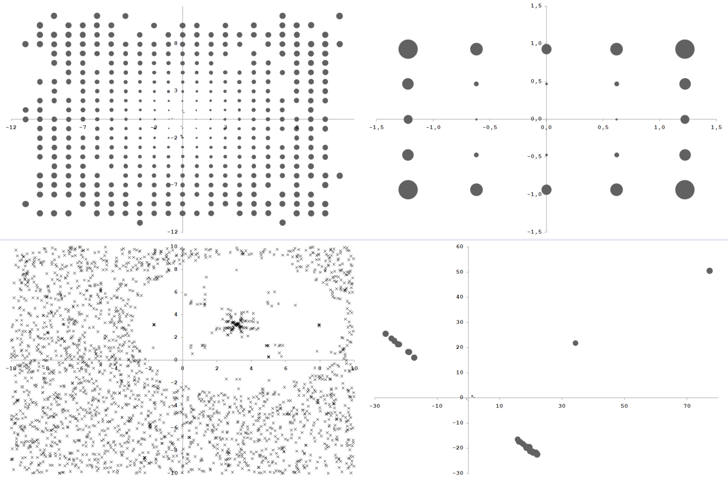

Problem
Data/Function/...
MF Extractor
Static meta-features
Database
[Meta-features, algorithm, results, stats, parameters]
Solver
Results, statistics
MF Extractor
Dynamic meta-features, events


Heterogeneous Solver Description
Function
Static Meta-features Extractor
Method
Particle Swarm Optimization
Velocity Update: SIW
Number of Particles: 30
Starting Point Min: -20
Starting Point Max: 20
Cognitive Acceleration Coefficient: 0.3
Social Acceleration Coefficient: 0.7
Stopping Condition 500 iterations
Events
Dynamic Meta-features Extractor
Inputs
Function
Method
Method Parameters
Stopping Condition
Outputs
Static Meta-features
Dynamic Meta-features
Result
# of Function Evaluations
# of Gradient Evaluations
# of Hessian Evaluations
...

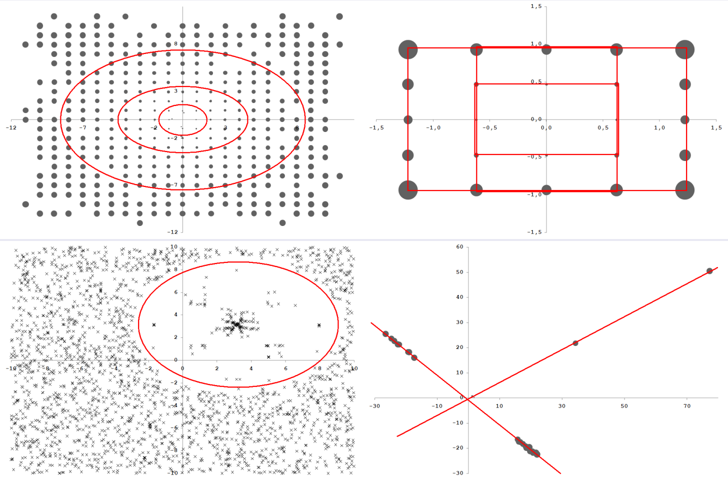
Landmarking

- Sampling the function value surface either by a grid or by a simple and fast algorithm
  - *quasi*-Newton method selected
- Since the function is a black box, no other information can be collected
- Aiming to answer which method should be used, with what parameter values and where is a good starting point
  - Average Delta Value
  - Average Step Length
  - Number of Different Minima
  - Value Difference to Trip Length
  - ...

Early Data



Early Data



-12 -7 -2 3 8
8 3 -2 -7 -12
1,5 1,0 0,5 0,0 -0,5 -1,0 -1,5
-1,5 -1,0 -0,5 0,0 0,5 1,0 1,5
10 8 6 4 2 0 -2 -4 -6 -8 -10
-10 -8 -6 -4 -2 0 2 4 6 8 10
60 50 40 30 20 10 0 -10 -20 -30
-30 -10 10 30 50 70

Current Stage

Feasibility?

Landmarking

- More complex functions are needed
  - or at least multidimensional functions must be tested
- Computational cost of finding a symmetry
- Broader set of static meta-features is yet to be identified