Seeking multiple solutions: multimodal optimization using niching methods

Professor Xiaodong Li
School of Science (Computer Science and Software Engineering)
RMIT University, Melbourne, Australia
Email: xiaodong.li@rmit.edu.au
URL: https://titan.csit.rmit.edu.au/~e46507/
Multi-Modal Optimization?

Curiosity's view of "Mount Sharp" (September 9, 2015)

Mount Sharp rises from the middle of Gale Crater; the green dot marks Curiosity's landing site (north is down).

Source: https://en.wikipedia.org/wiki/Curiosity_(rover)
Outline

• Background on Multi-Modal Optimization (MMO), i.e., finding multiple good solutions, in meta-heuristics
• What are the benefits for studying MMO?
• What are niching methods?
• Some real-world examples
• Classic niching methods: fitness sharing, crowding, clearing, and speciation, etc.
• PSO and DE for niching methods
• Niching in dynamic and multiobjective optimization
• IEEE CIS Taskforce on Multi-Modal Optimization
• Summary
• References
What is multimodal optimization?

- **Multi-Modal Optimization (MMO):** to locate multiple optimal (or close to optimal) solutions in the search space.
- This is different from a conventional optimization method which has a common goal of seeking to locate a single global optimum.
- **MMO** problems represent an important class of optimization problems. Many real-world optimization problems are multimodal by nature, that is, multiple satisfactory solutions exist (several real-world examples of MMO problems are provided in subsequent slides).
- From a decision maker’s point of view, it might be desirable to locate all global optima and/or some local optima that are also considered as being satisfactory.
Methods for MMO

• Optimization methods specifically designed for solving MMO problems, often called multimodal optimization or niching methods, are predominately developed from the field of meta-heuristic algorithms, which covers a family of population-based stochastic optimization algorithms, including evolutionary algorithms, evolutionary strategies, particle swarm optimization, differential evolution, and so on.

• These meta-heuristic algorithms are shown particularly effective in solving multimodal optimization problems, if equipped with specifically designed diversity preserving mechanisms, commonly referred to as niching methods.
What are the benefits?

• A decision maker may be interested to know whether there exist multiple equally good solutions before making a final decision.
• Important for a sensitivity study of a problem, and helps develop more robust solutions to the problem.
• Plays an important role in keeping a diverse population of candidate solutions, hence helps prevent the population from converging prematurely to a sub-optimum.
• May increase the probability of finding the global optimum.
Ecological inspiration

- In natural ecosystems, individual species must compete to survive by taking on different roles. Different species evolve to fill different niches (or subspaces) in the environment that can support different types of life.
What are niching methods?

• According to the Oxford Dictionary, a niche refers to “a role taken by a type of organism within its community”; and a species refers to “a group of living organisms consisting of similar individuals capable of exchanging genes or interbreeding”.

• These concepts of niches, species and speciation can be adopted in a population-based meta-heuristic algorithm (typically an evolutionary algorithm), to encourage the population to evolve different species targeting different optimal solutions in the search space.
MMO Publication trends

- Despite niching methods first appeared more than 30 years ago, currently niching techniques are experiencing a revival, attracting researchers from across a wide range of research fields.
MMO Application areas

- Agricultural and Biological Sciences
- Arts and Humanities
- Biochemistry, Genetics and Molecular Biology
- Business, Management and Accounting
- Chemical Engineering
- Chemistry
- Computer Science
- Decision Sciences
- Earth and Planetary Sciences
- Economics, Econometrics and Finance
- Energy
- Engineering
- Environmental Science
- Immunology and Microbiology
- Materials Science
- Mathematics
- Medicine
- Multidisciplinary
- Neuroscience
- Nursing
- Pharmacology, Toxicology and Pharmaceutics
- Physics and Astronomy
- Psychology
- Social Sciences
Engineering example: truss topology design

• In topology optimization, the connectivity of members in a truss is to be determined. There exist multiple different topologies with almost equal overall weight in truss-structure design problems as the members in the ground structure increase.

• The resulting solution of truss-structure optimization design problems becomes “multi-modal” with large number of truss members.

Some nodes in the ground structure may or may not be removed. The optimal structure is found as a subset of the ground structure.


Truss topology design

- Sharing scheme is used to compute the similarity between different topology design solutions.
- The sharing fitness is a reduced one from the original fitness, in order to discourage solutions in the vicinity.

Binary PSO is run based on the sharing fitness values, and multiple dissimilar truss topologies are derived and saved.
Truss topology design

Multiple optimal truss topologies found by BPSO with niching.
Trust structure design using Bilevel and niching aspects

- Formulate the truss problem as a **bilevel** optimization problem
- A new bilevel PSO niching method locates multiple optimal solutions
- **Stable topologies** can be found in the **upper level**
- The **optimized sizes** of the members of these topologies can be found in the **lower level**
- **Niching** at the **upper level**
- **Standard optimizer** is used at the **lower level** to optimize a bilevel truss problem

Trust structure design examples

Continuum structural topology optimization

Drug Molecule Design (I)

• Search for **molecular structures** with specific pharmacological or biological activity that influence the behavior of certain targeted cells

• **Objectives**: Maximization of potency of drug & Minimization of side-effects

• **Aim**: provide the medicinal chemist a **set of diverse molecular structures** that can be promising candidates for further research
  
  – **Fit solutions may** result in finding structures that **fail in practice**
  
  – The chemist **desires a set of promising structures** rather than only one single solution

Drug Molecule Design (II)

- Dynamic Niche Sharing technique incorporated to MOEA

Scheduling Problems

• Project Management
  – Optimize productivity
• Makespan, Due dates
  – Maximize revenue
  – Minimize delays

• Job shop Scheduling


Artificial examples

Demos

• DE/nrand/1 on Himmelblau 2D:
  https://www.youtube.com/watch?v=M32JdyBmVLc

• DE/nrand/1 for Shubert 2D:
  https://www.youtube.com/watch?v=miy3VK_8KwU

• DE/nrand on Deb's function with 100 optima 500 pop size:
  https://www.youtube.com/watch?v=ymNRKlrSyQU
Classic niching methods
Fitness sharing

- A sharing function can be used to degrade an individual’s fitness based on the presence of other neighbouring individuals.

- During selection, many individuals in the same neighbourhood would degrade each other’s fitness, thereby limiting the number of individuals occupying the same niche.

Crowding methods

• Originally by De Jong (1975), and later modified by Mahfoud (1995).

• **Crowding** usually consists of two phases:
  – Pairing phase: pairing each offspring with a similar individual in the current population; and
  – Replacement phase: which of the two will remain in the population?

• **Deterministic Crowding** selects the fittest individual in each pair in the replacement phase. Probabilistic Crowding selects the surviving individual for each pair based on a probabilistic formula that takes fitness into account.


Deterministic crowding

Algorithm 1: The pseudocode of deterministic crowding.

1: Select two parents, $p_1$ and $p_2$ randomly, without replacement
2: Generate two offspring $c_1$ and $c_2$
3: if $d(p_1, c_1) + d(p_2, c_2) \leq d(p_1, c_2) + d(p_2, c_1)$ then
4:    if $f(c_1) > f(p_1)$ then replace $p_1$ with $c_1$
5:    if $f(c_2) > f(p_2)$ then replace $p_2$ with $c_2$
6: else
7:    if $f(c_2) > f(p_1)$ then replace $p_1$ with $c_2$
8:    if $f(c_1) > f(p_2)$ then replace $p_2$ with $c_1$
9: end if

Each offspring tends to compete for survival with its most similar parent.
Clearing

- Proposed by Petrowski (1996); inspired by the principle of sharing of limited resources within each subpopulation (or species). The clearing procedure only supplies the resources to the best individuals in each subpopulation.

- All individuals fall within $r$ distance from the best $k$ individuals (below shows $k = 2$) from the population are cleared. This process is repeated until the whole population is considered.

Niching with PSO and DE
PSO niching methods

• In particle swarm optimization (PSO), each particle has its own memory remembering its best known position so far, and share this information with other particles.

• At each iteration, each particle is propelled towards the area defined by the stochastic average of its own known best position and the swarm best position.

• The notion of memory associated with each particle is unique to PSO, and this property can be used to induce niching behaviour: a swarm can be divided into two parts, an explorer-swarm consisting of the current particles, and a memory-swarm, comprising of only best known positions of individual particles.

Speciation-based PSO

An example of how to determine the species seeds from the population at each iteration. $s_1$, $s_2$, and $s_3$ are chosen as the species seeds. Note that $p$ follows $s_2$.

Ring topology based niching PSO

- Given a reasonably large population uniformly distributed in the search space, the ring topology based niching PSOs are able to form stable niches across different local neighbourhoods, eventually locating multiple global/local optima.
- This method can operate as a niching algorithm by using individual particles’ local memories to form a stable network retaining the best positions found so far,
Ring topology based niching PSO

• Results on Shubert 2D function (two snapshots during a simulation run).
DE niching methods

- Studies on the dynamics of DE suggest that the DE individuals are inclined to cluster around either local or global optima after some numbers of iterations.
- Inspired by this observation, the DE mutation operator in a classic DE variant DE/nrand/1, was altered to incorporate the nearest neighbour concept, in order to induce the niching effect.
- Instead of using the base vector the usual way, its nearest neighbour is always chosen as the actual base vector.

\[ v_{g+1}^i = x_{g}^{NN_i} + F(x_{g}^{r1} - x_{g}^{r2}), \quad (1) \]

\[ v_{g+1}^i = x_{g}^{NN_i} + F(x_{g}^{r1} - x_{g}^{r2}) + F(x_{g}^{r3} - x_{g}^{r4}), \quad (2) \]

Nearest-better clustering

- **The basic idea**: basins of attraction are indicated by contour lines. Each individual connects to its nearest neighbour which is better; clustering is done via cutting the longest lines (Preuss 2010). However, still need to set a few niching parameters.
- This NBC (nearest-better clustering) idea combined with CMA-ES produces a niching algorithm that won the top place in the CEC’2013 niching competition.

CEC’2013 niching benchmark

• A common platform for evaluating and comparing different niching algorithms.
• 20 benchmark multimodal functions with different characteristics.
• 5 accuracy levels: $\epsilon \in \{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}$
• The benchmark suite and the performance measures have been implemented in: C/C++, Java, MATLAB.

CEC 2013/2015/2016 competitions

- IEEE CEC niching competitions at 2013, 2015 and 2016, with the latest results available at the following URL:
  - http://www.epitropakis.co.uk/cec16-niching/competition/
  - https://github.com/mikeagn/CEC2013

The "Competition on Niching Methods for Multimodal Optimization" will be held as part of the IEEE Congress on Evolutionary Computation (IEEE CEC) 2020. A suite of twenty benchmark multimodal functions with different characteristics and levels of difficulty is provided.
20 test functions

<table>
<thead>
<tr>
<th>Id</th>
<th>Dim.</th>
<th># GO</th>
<th>Name</th>
<th>Characteristics</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_1$</td>
<td>1</td>
<td>2</td>
<td>Five-Uneven-Peak Trap</td>
<td>Simple, deceptive</td>
</tr>
<tr>
<td>$F_2$</td>
<td>1</td>
<td>5</td>
<td>Equal Maxima</td>
<td>Simple</td>
</tr>
<tr>
<td>$F_3$</td>
<td>1</td>
<td>1</td>
<td>Uneven Decreasing Maxima</td>
<td>Simple</td>
</tr>
<tr>
<td>$F_4$</td>
<td>2</td>
<td>4</td>
<td>Himmelblau</td>
<td>Simple, non-scalable, non-symmetric</td>
</tr>
<tr>
<td>$F_5$</td>
<td>2</td>
<td>2</td>
<td>Six-Hump Camel Back</td>
<td>Simple, not-scalable, non-symmetric</td>
</tr>
<tr>
<td>$F_6$</td>
<td>2,3</td>
<td>18,81</td>
<td>Shubert</td>
<td>Scalable, #optima increase with D, unevenly distributed grouped optima</td>
</tr>
<tr>
<td>$F_7$</td>
<td>2,3</td>
<td>36,216</td>
<td>Vincent</td>
<td>Scalable, #optima increase with D, unevenly distributed optima</td>
</tr>
<tr>
<td>$F_8$</td>
<td>2</td>
<td>12</td>
<td>Modified Rastrigin</td>
<td>Scalable, #optima independent from D, symmetric</td>
</tr>
<tr>
<td>$F_9$</td>
<td>2</td>
<td>6</td>
<td>Composition Function 1</td>
<td>Scalable, separable, non-symmetric</td>
</tr>
<tr>
<td>$F_{10}$</td>
<td>2</td>
<td>8</td>
<td>Composition Function 2</td>
<td>Scalable, separable, non-symmetric</td>
</tr>
<tr>
<td>$F_{11}$</td>
<td>2,3,5,10</td>
<td>6</td>
<td>Composition Function 3</td>
<td>Scalable, non-separable, non-symmetric</td>
</tr>
<tr>
<td>$F_{12}$</td>
<td>2,3,5,10</td>
<td>8</td>
<td>Composition Function 4</td>
<td>Scalable, non-separable, non-symmetric</td>
</tr>
</tbody>
</table>
Performance measures

**Peak Ratio (PR)** measures the average percentage of all known global optima found over multiple runs:

$$PR = \frac{\sum_{run=1}^{NR} \# \text{ of Global Optima}_i}{\text{(# of known Global Optima) } \times \text{(# of runs)}}$$

**Who is the winner:**

- The participant with the highest average Peak Ratio performance on all benchmarks wins.
- In all functions the following holds: the higher the PR value, the better.
CEC 2013/2015 niching competition top 3 entries

• **(NMMO) Niching Migratory Multi-Swarm Optimiser:**

• **(NEA2) Niching the CMA-ES via Nearest-Better Clustering:**

• **(LSEAGP) Localised Search Evolutionary Algorithm using a Gaussian Process:**
Algorithm performances

Accuracy level $\varepsilon = 10^{-5}$
Participants’ performance

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Statistics</th>
<th>Friedman’s Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Median</td>
<td>Mean</td>
</tr>
<tr>
<td>NMMSO</td>
<td>0.9885</td>
<td>0.8221</td>
</tr>
<tr>
<td>NEA2</td>
<td>0.8513</td>
<td>0.7940</td>
</tr>
<tr>
<td>LSEAEA</td>
<td>0.9030</td>
<td>0.7477</td>
</tr>
<tr>
<td>dADE/nrand/1</td>
<td>0.7488</td>
<td>0.7383</td>
</tr>
<tr>
<td>LSEAGP</td>
<td>0.7900</td>
<td>0.7302</td>
</tr>
<tr>
<td>CMA-ES</td>
<td>0.7550</td>
<td>0.7137</td>
</tr>
<tr>
<td>N-VMO</td>
<td>0.7140</td>
<td>0.6983</td>
</tr>
<tr>
<td>ALNM</td>
<td>0.7920</td>
<td>0.6594</td>
</tr>
<tr>
<td>PNA-NSGAll</td>
<td>0.6660</td>
<td>0.6141</td>
</tr>
<tr>
<td>NEA1</td>
<td>0.6496</td>
<td>0.6117</td>
</tr>
<tr>
<td>DE/nrand/2</td>
<td>0.6667</td>
<td>0.6082</td>
</tr>
<tr>
<td>dADE/nrand/2</td>
<td>0.7150</td>
<td>0.6931</td>
</tr>
<tr>
<td>DE/nrand/1</td>
<td>0.6396</td>
<td>0.5809</td>
</tr>
<tr>
<td>DELS-aj</td>
<td>0.6667</td>
<td>0.5760</td>
</tr>
<tr>
<td>CrowdingDE</td>
<td>0.6667</td>
<td>0.5731</td>
</tr>
<tr>
<td>DELG</td>
<td>0.6667</td>
<td>0.5706</td>
</tr>
<tr>
<td>DECG</td>
<td>0.6567</td>
<td>0.5516</td>
</tr>
<tr>
<td>IPOP-CMA-ES</td>
<td>0.2600</td>
<td>0.3625</td>
</tr>
<tr>
<td>MEA</td>
<td>0.2075</td>
<td>0.3585</td>
</tr>
<tr>
<td>A-NSGAll</td>
<td>0.0740</td>
<td>0.3275</td>
</tr>
<tr>
<td>MSSPSO</td>
<td>0.0000</td>
<td>0.2188</td>
</tr>
</tbody>
</table>
Niching in dynamic optimization
SPSO for tracking optima

• In contrast to optimization towards a static optimum, in a dynamic environment the goal is to track as closely as possible the dynamically changing optima.

• A useful strategy to ensure good tracking of the global optimum in a dynamic environment, is to maintain multiple species at all the optima found so far, regardless whether they are globally or locally optimal.

• By maintaining individual species at each local optimum, it helps tremendously in case when such a local optimum turns into a global optimum.

Niching in multiobjective optimization
EMO solution diversity

• Although diversity maintenance is a much common issue in any population-based metaheuristics, it is possible to use niching methods for maintaining solution diversity. An early example is the Niched-Pareto GA (NGPA) (Horn, et al., 1994), which is a multiobjective GA using a variant of fitness sharing to maintain Pareto solution diversity in the objective space. Another example is the crowding distance metric used in NSGA-II (Deb, et al., 2002).

• Much attention has been given to maintaining solution diversity in the objective space, however, little attention has been given to how to maintain solution diversity in the decision space. See next slide.


Diversity in both spaces

• A multiobjective evolutionary algorithm (e.g., multiobjective Niching-CMA) can produce a much more diverse set of efficient solutions (i.e., solutions in the decision space), without sacrificing objective space diversity (Shir, et. al. 2009).

![Decision space vs. Objective space](image)

An example where two solutions that are close in the objective space but their corresponding points in the decision space are further apart.

Omni-Optimizer

• Allows degeneration of NSGA-II into a single objective multimodal optimization method (i.e., a niching method).
• A variable space crowding distance metric is used to encourage distant solutions in the decision space to remain in the population.
• Distant solutions with similar or equal objective function values will survive.
• Omni-Optimizer can degenerate to a niching method for multiobjective multimodal optimization, capable of finding multiple Pareto-optimal fronts.

IEEE CIS Taskforce on MMO

- **The key objective** is to promote research on multi-modal optimization, including its development, education and understanding of sub topic areas of multi-modal optimization. Further info: [http://www.epitropakis.co.uk/ieee-mmo/](http://www.epitropakis.co.uk/ieee-mmo/)

- **Current chair**: Michael G. Epitropakis (Lancaster University, UK).

- **Vice-Chairs**: Andries Engelbrecht (University of Pretoria, South Africa), and Xiaodong Li (RMIT University, Australia).

- **Members**: Carlos A. Coello Coello, Kalyanmoy Deb, Andries Engelbrecht, Michael G. Epitropakis, Jonathan Fieldsend, Jian-Ping Li, Xiaodong Li, Jonathan Mwaura, Konstantinos Parsopoulos, Vassilis Plagianakos, Mike Preuss, Bruno Sareni, Ofer M. Shir, Patrick Siarry, P. N. Suganthan, Michael N. Vrahatis, Bo-Yang Qu, Simon Wessing, Xin Yao.

- **Past and planned activities**:
  - IEEE CEC 2010, 2013, and 2015 special session and/or competitions on “Niching Methods for Multimodal Optimization”.
  - International Workshop on "Advances in Multimodal Optimization", PPSN 2014.
  - Tutorial at WCCI 2016.
  - A repository for publications and source codes.
Summary

• Niching methods have been studied for the past few decades, and now experience a revival, as more people from diverse backgrounds find its relevance in their own disciplinary areas.
• Niching methods can be developed using other meta-heuristics, apart from evolutionary algorithms.
• Niching has its application in many problem solving domains, e.g., dynamic optimization and multiobjective optimization.
• A good starting point for new comers: several survey papers are available, plus recently a new book by Mike Preuss.
• Many open research questions and challenges to be addressed.
• Many possible real-world applications of niching methods.

References


References