A Multipopulation Differential Evolution Algorithm for Multimodal Optimization

Daniela Zaharie
West University
Timisoara, Romania
dzaharie@info.uvt.ro
Outline

- Multimodal optimization
- EAs for multimodal optimization
- Multipopulation Differential Evolution
- A multiresolution variant
- Numerical results
- Conclusions
Multimodal optimization (1)

- **Aim:** find all optima (global and/or local) of the objective function

- **Motivation:**
  - give to the decision maker not a single optimal solution but a set of good solutions
  - find all solutions with local optimal behavior

- **Similar with:** multiobjective optimization

- **Applications:**
  - Systems design
  - DNA sequence analysis
  - Detecting peaks in DTA (differential thermal analysis) curves

MENDEL’04
Multimodal optimization (2)

Global optimization

Aim: find a global optimum

Evolutionary approach: population concentrates on the global optima (single powerful species)

Premature convergence: bad

Multimodal optimization

Aim: find all (global/local) optima

Evolutionary approach: different species are formed each one identifying an optimum

Premature convergence: not so bad

MENDEL’04
EAs for multimodal optimization (1)

- Multimodal evolutionary approaches:
  - Sequential niching models
    - Iterative application of an EA
    - At each iteration is identified an optimum
    - The fitness function is derated based on already found optima
      [Beasley et al., 1993]
  - Parallel subpopulation models
    - Divide the population into communicating subpopulations which evolves in parallel
    - Each subpopulation corresponds to a species whose aim is to populate a niche in the fitness landscape and to identify an optimum
    - Speciation is usually assured by a clustering process
      [Bessaou et al., 2000], [Li et al., 2002]
EAs for multimodal optimization (2)

- **Difficulties:**
  - Finding an adequate *niche radius*
  - *Computational cost* of the clustering process (usually $O(m^2)$)

- **Aim of this work:**
  - Analyze the applicability of Differential Evolution to multimodal optimization
  - Develop an algorithm which:
    - Uses the *fast convergence* and *robustness* of DE
    - Uses *few control parameters*
    - *Avoids* a global processing of the entire population (clustering step)
    - Easy to be implemented in *parallel*
Multipopulation DE (1)

- Population structure
  - $s$ subpopulations of fixed size $m$
- Controlled initialization
  - assure landscape exploration
- DE2-type recombination
  - fast convergence
- Random migration
- Convergence?
  - subpopulations variance becomes small
- Archiving
  - collects the best elements of the subpopulations

Subpopulations initialization ($P_1, \ldots, P_s$)

Apply DE to ($P_1, \ldots, P_s$)

Migration

Convergence?

Archiving the best elements
Multipopulation DE (2)

- **Problem:**
  
  find the maxima of \( f : \mathbb{D} \subseteq \mathbb{R}^n \rightarrow \mathbb{R}, \ \mathbb{D} = [a, b]^n \)

- **Resolution factor:**
  
  \[ r = \frac{b-a}{s^{1/n}} \]

- **Subpopulation \( P_i \) is initialized in**
  
  \[ D_i = [a_{1i}, b_{1i}] \times \ldots \times [a_{ni}, b_{ni}] \]

  \[ a_{ji} = a + r k^i_j, \ \ b_{ji} = a^i_j + r \]

  \( k^i_j \in \{0, 1, \ldots, \lfloor s^{1/n} \rfloor - 1\} \) randomly selected

- **Initially the elements of a subpopulation are relatively close to each other**

- **During the evolution the subpopulations can overlap**
Multipopulation DE (3)

Motivation:
Fast convergence to an optimum (the subpopulations find an optimum in their neighbourhood)

Subpopulations initialization \((P_1, \ldots, P_s)\)

Apply DE to \((P_1, \ldots, P_s)\)

Migration

Convergence ?

Archiving the best elements

Parents

\[
\begin{array}{cccccccc}
  x_1 & \ldots & x_i & \ldots & x_j & \ldots & x_k & \ldots & x^* & \ldots & x_m \\
\end{array}
\]

Offspring

\[
\begin{array}{cccccccc}
  z & z & z & z & z & z & Z & \ldots & Z \\
\end{array}
\]

\[
\begin{array}{l}
  z_i = \begin{cases} 
  x^* + F \times (x_j - x_k) & \text{with probability } p \\
  x_i & \text{with probability } 1-p 
  \end{cases} 
\end{array}
\]

Selection: the best one between the parent and the offspring is selected

Motivation:
- Fast convergence to an optimum (the subpopulations find an optimum in their neighbourhood)
Multipopulation DE (4)

- **Random migration:**
  - After a given number of generations the subpopulations exchange information
  - Each element of a subpopulation can be swapped with a migration probability with a randomly selected element from a random subpopulation

- **Migration effects:**
  - Ensures an increase of subpopulations diversity
  - Avoid premature convergence
  - Can guide different subpopulations toward the same optimum (the subpopulations centroids migrate toward the population centroid)

- **Practical remark:**
  - If multiple optima have to be located the migration probability should be small
Multipopulation DE (5)

- Dependence of the number of found optima on the migration probability

Test function: multigaussian – 10 optima
Multipopulation DE (6)

- Archiving:
  - The best element from each subpopulation is stored in an archive.

- Redundancy avoiding: only the elements which:
  - are sufficiently dissimilar
  - belong to different peaks
  - than those already stored are retained.

- To decide if two elements belong to different peaks the hill-valley function is used.
  
  [Ursem, Multinational Evolutionary Algorithms, 1999]
Hill-valley function:

- If there exists $c \in (0,1)$ such that
  
  $$z = cx + (1-c)y \text{ implies } f(z) < f(x) \text{ and } f(z) < f(y)$$

  then there exists a valley between $x$ and $y$

- The decision is based on computing $z$ for some values of $c$
A multiresolution variant (1)

Motivation:
- When many optima have to be found, MDE needs many subpopulations
- If the optima are *unequally spaced* some of them could be missed

Basic idea:
- Apply repeatedly the MDE for different resolution factors
- Hybridization between the sequential and parallel niching methods

(Re)initialization:
- Based on the resolution factor and on the archive content

Communication between different epochs:
- Through the archive

<table>
<thead>
<tr>
<th>Archive initialization</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e := 0$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Controlled (re)initialization</th>
</tr>
</thead>
<tbody>
<tr>
<td>$r_e := (b-a)/(se)^{1/n}$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MDE epoch</th>
</tr>
</thead>
<tbody>
<tr>
<td>DE+migration</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Archiving</th>
</tr>
</thead>
<tbody>
<tr>
<td>$e := e+1$</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Convergence ?</th>
</tr>
</thead>
</table>

| $e < e_{max}$ |

DE+migration

Archiving

Archive initialization

$e := 0$

$e := e+1$

$e < e_{max}$

MENDEL’04
A multiresolution variant (2)

- Idea of controlled (re)initialization
  - At each new epoch \( e \), the elements of subpopulation \( P_i \) are selected from a subdomain

\[
D_i = [a_i^j, b_i^j] \times \ldots \times [a_n^j, b_n^j]
\]

\[
a^j = a + r_e k^j_i, \quad b^j = a^j + r_e, \quad r_e = (b - a) / (se)^{1/n}
\]

\( k^j_i \in \{0, 1, \ldots, [s^{1/n}] - 1\} \) randomly selected

- A random element from \( D_i \) is accepted with the probability

\[
P_a(x) = \frac{1}{1 + \sum_{i=1}^{k} \sigma(x, a_i)}, \quad \sigma(x, a) = \begin{cases} 
1 - \frac{d(x, a)}{r_e / 2} & \text{if } d(x, a) < r_e / 2 \\
0 & \text{otherwise}
\end{cases}
\]

\( A = \{a_1, \ldots, a_k\} \) – the archive

- The selection of elements from \( D_i \) is based on a non-uniform distribution obtained by modifying the uniform distribution by using a sharing function

MENDEL’04
A multiresolution variant (3)

- Illustration of controlled (re)initialization

(Shubert function (2D) on $[-10,10]^2$ – 18 global optima)

$$f(x, y) = \sum_{j=1}^{5} j(\cos(j + 1)x + j) \sum_{j=1}^{5} j(\cos(j + 1)y + j)$$

Epoch 2

Epoch 3

Epoch 4

Epoch 5

Population elements

Current archive

Restricted area around found optima

MDE parameters:
- $m=10$
- $s=20$
- $p_m=0.1$
- 100 generations
- $t_m=10$

MENDEL’04
A multiresolution variant (4)

Illustration of controlled (re)initilization (Shubert function (2D) – 18 global optima)

Population elements

Current archive

Restricted area around found optima

MDE parameters:
\[ m = 10 \]
\[ s = 20 \]
\[ p_m = 0.1 \]
100 generations
\[ t_m = 10 \]

Epochs 6 to 9

MENDEL’04
Aim of experiments:
- Analyze the ability of MMDE to locate multiple optima
- Compare MMDE with other multimodal evolutionary techniques

Experimental setup:
- The population is divided into $s$ subpopulations of fixed size $m$
- DE convergence for a subpopulation: $\text{Var}(X(g)) < 10^{-5}$
- Migration: random
- DE parameters:
  - $p=1$
  - $F$ adaptive:

\[
F(g) = \sqrt{\frac{c(g) - (m-1)/m}{2}}, \quad c(g) = \frac{\text{Var}(X(g-1))}{\text{Var}(X(g))}
\]
Comparison between MDE and MMDE:

Test function: multigaussian

Numerical results (2)

Succes rate (%)  Function evaluations

MDE  MMDE

m=5, g=50

MENDEL’04
Numerical results (3)

Test function: Schaffer 2D

\[ f(x, y) = 0.5 + \frac{\sin^2(\sqrt{x^2 + y^2}) - 0.5}{(1 + 0.001(x^2 + y^2))} , \quad x, y \in [-10, 10] \]

- \[ m=10, s=20, g=50, e=10, p_m=0 \]
- \[ m=10, s=20, g=10 \times 10, e=10, p_m=0.5 \]
Numerical results (4)

Test function: multi-peaks

\[ f(x, y) = x \sin(4\pi x) - y \sin(4\pi y + \pi), \quad x, y \in [-2, 2] \]

- \( m=5, s=50, g=50, e=20, p_m=0 \) (91 elements in the archive)
- \( m=5, s=50, g=10 \times 10, e=20, p_m=0.5 \) (4 elements in the archive)
Comparative results

Test function: Himmelblau

<table>
<thead>
<tr>
<th>MMDE ($m=5, s=10, e_{\text{max}}=2, p_m=0$)</th>
<th>Sequential niching [Beasley, 1993] ($m=26$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fct.eval.</td>
<td>Succes rate</td>
</tr>
<tr>
<td>2665</td>
<td>96%</td>
</tr>
</tbody>
</table>

Test function: multi-peaks [de Castro, 2002]

<table>
<thead>
<tr>
<th>MMDE ($m=5, s=50, e_{\text{max}}=20, p_m=0$)</th>
<th>Opt AI-net (20 cells, 10 clones, 451 gen.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fct. eval.</td>
<td>No. of optima</td>
</tr>
<tr>
<td>76630</td>
<td>88.16</td>
</tr>
</tbody>
</table>
Comparative results

Test function: Shubert 2D

<table>
<thead>
<tr>
<th>Method</th>
<th>Fct. eval.</th>
<th>No. of optima</th>
<th>Method</th>
<th>Fct. eval.</th>
<th>No. of optima</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMDE (m=10, s=50, e_{max}=10, p_{m}=0)</td>
<td>39463</td>
<td>17.26</td>
<td>SCGA [Li et al., 2002] (m=300)</td>
<td>35747</td>
<td>18</td>
</tr>
</tbody>
</table>

Test function: Schaffer 2D

<table>
<thead>
<tr>
<th>Method</th>
<th>Fct. eval.</th>
<th>Succes rate</th>
<th>Method</th>
<th>Fct. eval.</th>
<th>Success rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>MMDE (m=40, s=10, e_{max}=1, p_{m}=0.5)</td>
<td>26253</td>
<td>90%</td>
<td>Island model+speciation [Bessaou et al., 2000] (m=10, s=50)</td>
<td>18000</td>
<td>100%</td>
</tr>
</tbody>
</table>
Conclusions

- Characteristics of MMDE:
  - Exploration ensured by a multi-resolution approach and a controlled (re)initialization of subpopulations
  - Exploitation ensured by a adaptive DE2 variant
  - Preservation of good solutions by a controlled archiving
  - Small subpopulations
  - Migration introduce flexibility:
    - high migration probability: locate one global optima
    - small migration probability: locate all global optima
    - no migration: identify all global/local optima
  - No niche radius
  - No global clustering
  - Easy to parallelize
  - Sensitivity to the number of subpopulations and to the number of epochs