Negatively Correlated Search

Ke Tang

USTC-Birmingham Joint Research Institute in Intelligent Computation and Its Applications (UBRI)
School of Computer Science and Technology
University of Science and Technology of China, China

May 2016 @ ECOLE’16
Outline

• Motivation

• Negatively Correlated Search: The General Idea

• NCS-C: A simple Instantiation

• More Interpretations

• Summary and Future Work
Outline

• Motivation
  • Negatively Correlated Search: The General Idea
  • NCS-C: A simple Instantiation
  • More Interpretations
• Summary and Future Work
Motivation

- Population-based Search methods (e.g., EAs)

<table>
<thead>
<tr>
<th>A Unified/Simplified Framework</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Generate the initial population $P(0)$ of size $N$ at random and set $t \leftarrow 0$;</td>
</tr>
<tr>
<td>• Repeat:</td>
</tr>
<tr>
<td>Evaluate the fitness of each individual in $P(0)$;</td>
</tr>
<tr>
<td>Generate population $P(t+1)$ based on $P(t+1)$;  // recombination, selection, etc.</td>
</tr>
<tr>
<td>$t \leftarrow t+1$;</td>
</tr>
<tr>
<td>• Until halting criteria are satisfied.</td>
</tr>
</tbody>
</table>

- It seems that EAs are quite similar to Simulated Annealing, except for the use of a *population*. But why shall we use a population?
Motivation

• Answers offered in the literature:
  – Parallel/distributed search in the solution space
    → the optimal/near-optimal solution could be found more efficiently?

• We’re not merely talking about parallel search, otherwise
  – The same region might be repeatedly visited
    → waste of time
  – Linear speed-up would be the best case one can expect
    → not “more efficiently” in the sense of AI.

• Search with a population ↔ a population of search processes.

• The search processes shouldn’t be independent, but *correlated*.
Outline

• Motivation

• Negatively Correlated Search: The General Idea
  • NCS-C: A simple Instantiation
  • More Interpretations

• Summary and Future Work
Negatively Correlated Search

- Consider the case of hill climbing (a search process), how to correlate two HC processes?

- At time step $t$, communication is made such that two HCs will search different regions in the next iteration(s).

<table>
<thead>
<tr>
<th>Steps of Hill Climbing</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Generate an initial solution $x_0$ at random and set $i=0$;</td>
</tr>
<tr>
<td>2. Calculate the fitness $f(x_0)$</td>
</tr>
<tr>
<td>3. Repeat:</td>
</tr>
<tr>
<td>4. Apply a search operator to $x_i$ to generate a new solution $x_i'$;</td>
</tr>
<tr>
<td>5. Preserve the solution with higher fitness as $x_{i+1}$;</td>
</tr>
<tr>
<td>6. $i \leftarrow i+1$;</td>
</tr>
<tr>
<td>7. Until halting criteria are satisfied.</td>
</tr>
</tbody>
</table>
negatively correlated
Negatively Correlated Search

- Suppose solutions $x_i$ and $x_j$, corresponding to two HCs.

  $x_i$ \hspace{0.5cm} \text{Search Operator} \hspace{0.5cm} x_i'$

  $x_j$ \hspace{0.5cm} \text{Search Operator} \hspace{0.5cm} x_j'$

- For each HC, the solution with both high fitness and large distance to the other HC(s) is preferred.

- Generate an initial solution $x_0$ at random and set $i=0$;
- Calculate the fitness $f(x_0)$
- Repeat:
  1. Apply a search operator to $x_i$ to generate a new solution $x_i'$;
  2. Preserve the solution with higher fitness and larger distance to $x_j$ as $x_{i+1}$;
  3. $i \leftarrow i+1$;
- Until halting criteria are satisfied.
Negatively Correlated Search

• There are many metrics for measuring the distance between distributions, NCS adopts the Bhattacharyya distance.

  
  **Continuous case**

  \[ D_B(p_i, p_j) = -\ln \left( \int \sqrt{p_i(x)p_j(x)} \, dx \right) \]

  
  **Discrete case**

  \[ D_B(p_i, p_j) = -\ln \left( \sum_{x \in X} \sqrt{p_i(x)p_j(x)} \right) \]

  Probability density function unknown: Sampling + Estimation

• In case of \( N (N>2) \) HCs, the distance of a distribution \( p_i \) to the other distributions is measured as:

  \[ \text{Corr}(p_i) = \min_{j \neq i} \{ D_B(p_i, p_j) \} \]
• Fitness and distance may be conflicting:
  – large distance: explore areas not to be visited by others (exploration).
  – high fitness: strive to find a better solution (exploitation).

• The following heuristic is adopted to control the trade-off between two criteria:

\[
\begin{cases}
\text{discard } x_i, & \text{if } \frac{f(x'_i)}{\text{Corr}(p'_i)} < \lambda \\
\text{discard } x'_i, & \text{otherwise}
\end{cases}
\]

where $\lambda > 0$ is a parameter to balance exploration and exploitation.

• Normalization is needed before using the heuristic since fitness and correlation may of different scales.
The general framework of NCS

- Generate the initial population $P(0)$ of size $N$ at random and set $t \leftarrow 0$;
- Evaluate the fitness of all initial solutions;
- Reserve the solution with the highest fitness in an external archive ($BestFound$);
- Repeat:
  
  For $i = 1$ to $N$
  
  1. Generate a new solution $x_i'$ based on; $x_i$;
  
  2. Compute $f(x_i)$, $Corr(p_i)$ and $Corr(p_i')$
  
  EndFor
  
  For $i = 1$ to $N$
  
  1. Update $BestFound$;
  
  2. Update $x_i$ based on normalized $f$ and $Corr$;
  
  EndFor
  
  $t \leftarrow t+1$;
- Until halting criteria are satisfied.
Negatively Correlated Search

The general framework of NCS

- Generate the initial population $P(0)$ of size $N$ at random and set $t \leftarrow 0$;
- Evaluate the fitness of all initial solutions;
- Reserve the solution with the highest fitness in an external archive ($BestFound$);
- Repeat:
  
  For $i = 1$ to $N$
  
  1. Generate a new solution $x'_i$ based on $x_i$;
  2. Compute $f(x_i)$, $\text{Corr}(p_i)$ and $\text{Corr}(p'_i)$;

  EndFor
  
  For $i = 1$ to $N$
  
  1. Update $BestFound$;
  2. Update $x_i$ based on normalized $f$ and $\text{Corr}$;

  EndFor

  $t \leftarrow t+1$;

- Until halting criteria are satisfied.
Negatively Correlated Search

• NCS can also be understood as a multi-agent system.

• It consists of a population of search processes (agents).

• Agents communicate periodically to behave diversely.

Let’s cover different parts.

How to handle the task?

I’ll take the left part.

Brilliant! Let me cover the right part.
Negatively Correlated Search

• NCS is not restricted to a population of HCs or (1+1) EAs, but only requires a 1-to-1 mapping between parents and offspring, which holds for many EAs, such as Particle Swarm Optimizer (PSO) and Differential Evolution (DE).

• NCS is neither restricted to comparisons of individuals, it can be adapted to other scenarios where a selection is needed, e.g., adaption of search operators/control parameters.
Outline

• Motivation

• Negatively Correlated Search: The General Idea

• NCS-C: A simple Instantiation

• More Interpretations

• Summary and Future Work
NCS-C: A Simple Instantiation

• A simple instantiation of NCS, namely NCS-C is implemented to demonstrate the effectiveness of NCS on continuous multimodal optimization problems.

• The key issues for specifying a NCS instantiation.
  – The individual search processes
  – The calculation of Bhattacharyya distance
  – Setting the control parameter $\lambda$
Specifications of NCS-C

- The individual search process: (1+1) ES with Gaussian mutation operator, i.e.,

\[ x'_{id} = x_{id} + \mathcal{N}(0, \sigma_i) \]

where a solution \( x_i = [x_{i1}, \ldots, x_{id}, \ldots, x_{iD}] \)

- This means a new solution \( x'_i \) is obtained by sampling the following multivariate normal distribution:

\[ \mathcal{N}(x_i, \Sigma_i), \text{ where } \Sigma_i = \sigma_i^2 I \]
Specifications of NCS-C

• The Bhattacharyya distance between two multivariate Gaussian distributions is:

\[ D_B(p_i, p_j) = \frac{1}{8} (x_i - x_j)^T \Sigma^{-1} (x_i - x_j) + \frac{1}{2} \ln \left( \frac{\det \Sigma}{\sqrt{\det \Sigma_i \det \Sigma_j}} \right) \]

where \( \Sigma = \frac{\Sigma_i + \Sigma_j}{2} \)
Specifications of NCS-C

• $\sigma_i$ is initialized to the same value for all search processes.
• Each $\sigma_i$ is adapted according to the 1/5 successful rule in [1].

$$\sigma_i = \begin{cases} 
\frac{\sigma_i}{r} & \text{if } \frac{c}{\text{epoch}} > 0.2 \\
\sigma_i \times r & \text{if } \frac{c}{\text{epoch}} < 0.2 \\
\sigma_i & \text{if } \frac{c}{\text{epoch}} = 0.2 
\end{cases}$$

• Lambda initialized to, and shrinks over time.

$$\lambda_t = \mathcal{N}(1, 0.1 - 0.1 \times \frac{t}{T_{max}})$$

## Results of NCS-C

- 9 algorithms, 25 runs, 3e+06 evaluation on 20 multimodal problems of CEC2005 benchmark set.

<table>
<thead>
<tr>
<th>Algo.</th>
<th>PHC</th>
<th>SA</th>
<th>TS</th>
<th>SS</th>
<th>GL-25</th>
<th>SaDE</th>
<th>CMA-ES</th>
<th>CLPSO</th>
<th>NCS-C</th>
</tr>
</thead>
<tbody>
<tr>
<td>(F_6)</td>
<td>2.61E+01±2.35E+01</td>
<td>3.90E+02±4.09E+01</td>
<td>7.00E+03±1.01E+04</td>
<td>2.17E+05±6.41E+04</td>
<td>2.13E+01±1.02E+01</td>
<td>4.76E+01±3.35E+01</td>
<td>0.00E+00±0.00E+00</td>
<td>4.80E+00±3.55E+00</td>
<td>2.08E+01±3.61E+00</td>
</tr>
<tr>
<td>(F_7)</td>
<td>9.86E-04±2.76E-03</td>
<td>2.21E+00±1.84E+00</td>
<td>1.64E-02±1.59E-02</td>
<td>1.40E+00±7.31E-02</td>
<td>2.78E-02±3.76E-02</td>
<td>1.95E-02±1.37E-02</td>
<td>1.84E-03±4.59E-03</td>
<td>4.63E-01±7.31E-02</td>
<td>1.69E-02±1.38E-02</td>
</tr>
<tr>
<td>(F_8)</td>
<td>2.00E+01±1.29E-02</td>
<td>2.10E+01±7.13E-02</td>
<td>2.01E+01±3.60E-02</td>
<td>2.09E+01±5.12E-02</td>
<td>2.10E+01±4.76E-02</td>
<td>2.09E+01±5.62E-02</td>
<td>2.03E+01±5.60E-02</td>
<td>2.10E+01±1.22E-02</td>
<td>2.00E+01±1.22E-02</td>
</tr>
<tr>
<td>(F_9)</td>
<td>1.07E+02±2.13E+01</td>
<td>2.41E+02±8.62E+01</td>
<td>4.83E+02±9.60E+01</td>
<td>2.57E+02±3.85E+01</td>
<td>2.63E+02±5.64E+00</td>
<td>1.99E-01±4.06E-01</td>
<td>4.12E-02±1.38E+00</td>
<td>0.00E+00±1.38E+00</td>
<td>9.36E+01±1.38E+00</td>
</tr>
<tr>
<td>(F_{10})</td>
<td>9.64E+01±1.84E+01</td>
<td>2.17E+02±8.69E+01</td>
<td>7.92E+02±1.43E+02</td>
<td>3.48E+02±9.51E+01</td>
<td>1.35E+02±6.67E+00</td>
<td>5.08E+01±1.32E+00</td>
<td>4.97E+01±1.31E+00</td>
<td>1.06E+02±1.79E+01</td>
<td>9.03E+01±1.79E+01</td>
</tr>
<tr>
<td>(F_{11})</td>
<td>1.57E+01±1.89E+00</td>
<td>2.70E+01±2.18E+00</td>
<td>1.89E+01±4.44E+00</td>
<td>2.58E+01±4.55E+00</td>
<td>3.15E+01±8.45E+00</td>
<td>1.68E+01±2.82E+00</td>
<td>6.23E+00±1.47E+00</td>
<td>2.53E+01±1.37E+01</td>
<td>1.27E+00±1.27E+00</td>
</tr>
<tr>
<td>(F_{12})</td>
<td>7.53E+03±6.72E+03</td>
<td>6.06E+03±5.30E+03</td>
<td>2.28E+03±3.39E+03</td>
<td>1.18E+04±7.82E+03</td>
<td>6.83E+03±4.34E+03</td>
<td>3.11E+03±1.25E+03</td>
<td>1.28E+04±1.53E+04</td>
<td>1.96E+04±4.44E+03</td>
<td>1.57E+03±1.52E+03</td>
</tr>
<tr>
<td>(F_{13})</td>
<td>4.32E+00±9.03E-01</td>
<td>1.33E+01±1.04E+01</td>
<td>1.19E+01±3.36E+00</td>
<td>2.80E+01±4.44E+00</td>
<td>7.88E+00±5.79E+00</td>
<td>3.72E+00±5.89E-01</td>
<td>3.35E+00±8.52E-01</td>
<td>2.09E+00±8.04E-01</td>
<td>4.54E+00±8.04E-01</td>
</tr>
<tr>
<td>(F_{14})</td>
<td>1.34E+01±2.11E-01</td>
<td>1.47E+01±1.07E-01</td>
<td>1.42E+01±3.11E-01</td>
<td>1.35E+01±3.92E-01</td>
<td>1.29E+01±2.71E-01</td>
<td>1.26E+01±1.95E-01</td>
<td>1.47E+01±2.64E-01</td>
<td>1.24E+01±3.31E-01</td>
<td>1.24E+01±3.31E-01</td>
</tr>
<tr>
<td>(F_{15})</td>
<td>3.79E+02±5.35E+01</td>
<td>5.72E+02±1.18E+02</td>
<td>8.42E+02±3.19E+02</td>
<td>4.33E+02±4.75E+01</td>
<td>3.00E+02±7.62E-02</td>
<td>3.60E+02±6.51E+01</td>
<td>5.13E+02±2.69E+02</td>
<td>6.33E+01±4.87E+01</td>
<td>3.15E+02±5.68E+01</td>
</tr>
</tbody>
</table>
## Results of NCS-C

<table>
<thead>
<tr>
<th>Algo.</th>
<th>PHC</th>
<th>SA</th>
<th>TS</th>
<th>SS</th>
<th>GL-25</th>
<th>SaDE</th>
<th>CMA-ES</th>
<th>CLPSO</th>
<th>NCS-C</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F_{18}$</td>
<td>1.42E+02</td>
<td>3.77E+02</td>
<td>5.96E+02</td>
<td>4.21E+02</td>
<td>1.44E+02</td>
<td>8.16E+01</td>
<td>3.39E+02</td>
<td>1.76E+02</td>
<td>1.21E+02</td>
</tr>
<tr>
<td></td>
<td>±4.36E+01</td>
<td>±1.93E+02</td>
<td>±3.35E+02</td>
<td>±1.89E+00</td>
<td>±7.76E+01</td>
<td>±6.90E+01</td>
<td>±2.99E+02</td>
<td>±3.25E+01</td>
<td>±1.53E+01</td>
</tr>
<tr>
<td>$F_{17}$</td>
<td>1.90E+02</td>
<td>6.46E+02</td>
<td>8.75E+02</td>
<td>3.28E+02</td>
<td>1.58E+02</td>
<td>7.31E+01</td>
<td>4.15E+02</td>
<td>2.36E+02</td>
<td>1.55E+02</td>
</tr>
<tr>
<td></td>
<td>±3.94E+01</td>
<td>±3.12E+02</td>
<td>±3.34E+02</td>
<td>±1.29E+02</td>
<td>±7.17E+01</td>
<td>±2.79E+01</td>
<td>±3.07E+02</td>
<td>±4.37E+01</td>
<td>±2.40E+01</td>
</tr>
<tr>
<td>$F_{18}$</td>
<td>9.10E+02</td>
<td>8.23E+02</td>
<td>9.29E+02</td>
<td>8.32E+02</td>
<td>9.06E+02</td>
<td>8.75E+02</td>
<td>9.04E+02</td>
<td>9.10E+02</td>
<td>8.79E+02</td>
</tr>
<tr>
<td></td>
<td>±1.98E+00</td>
<td>±1.60E+01</td>
<td>±1.60E+02</td>
<td>±4.00E+01</td>
<td>±1.49E+00</td>
<td>±6.32E+01</td>
<td>±1.86E-01</td>
<td>±2.15E+01</td>
<td>±8.68E+01</td>
</tr>
<tr>
<td>$F_{19}$</td>
<td>9.09E+02</td>
<td>8.23E+02</td>
<td>9.54E+02</td>
<td>8.45E+02</td>
<td>9.07E+02</td>
<td>9.07E+02</td>
<td>9.25E+02</td>
<td>9.14E+02</td>
<td>8.93E+02</td>
</tr>
<tr>
<td></td>
<td>±1.74E+00</td>
<td>±1.40E+01</td>
<td>±1.92E+02</td>
<td>±7.77E+01</td>
<td>±1.71E+01</td>
<td>±4.08E+01</td>
<td>±1.07E+02</td>
<td>±1.79E+00</td>
<td>±4.12E+01</td>
</tr>
<tr>
<td>$F_{20}$</td>
<td>9.09E+02</td>
<td>8.29E+02</td>
<td>1.01E+03</td>
<td>8.24E+02</td>
<td>9.07E+02</td>
<td>8.83E+02</td>
<td>9.04E+02</td>
<td>9.14E+02</td>
<td>8.81E+02</td>
</tr>
<tr>
<td></td>
<td>±1.92E+00</td>
<td>±3.46E+01</td>
<td>±1.95E+02</td>
<td>±8.86E-01</td>
<td>±1.54E+00</td>
<td>±5.84E+01</td>
<td>±2.32E-01</td>
<td>±1.19E+00</td>
<td>±1.23E+02</td>
</tr>
<tr>
<td>$F_{21}$</td>
<td>4.96E+02</td>
<td>8.47E+02</td>
<td>9.08E+02</td>
<td>8.22E+02</td>
<td>5.00E+02</td>
<td>5.00E+02</td>
<td>5.12E+02</td>
<td>5.00E+02</td>
<td>5.00E+02</td>
</tr>
<tr>
<td></td>
<td>±1.81E+01</td>
<td>±1.03E+02</td>
<td>±3.43E+02</td>
<td>±2.60E+02</td>
<td>±4.83E-13</td>
<td>±2.09E-13</td>
<td>±6.00E+01</td>
<td>±2.38E-13</td>
<td>±2.32E-13</td>
</tr>
<tr>
<td>$F_{22}$</td>
<td>9.41E+02</td>
<td>7.45E+02</td>
<td>1.34E+03</td>
<td>5.74E+02</td>
<td>9.28E+02</td>
<td>9.33E+02</td>
<td>8.24E+02</td>
<td>9.70E+02</td>
<td>9.06E+02</td>
</tr>
<tr>
<td></td>
<td>±2.11E+01</td>
<td>±2.25E+02</td>
<td>±1.60E+02</td>
<td>±1.27E+02</td>
<td>±1.07E+01</td>
<td>±2.00E+01</td>
<td>±1.59E+01</td>
<td>±1.04E+01</td>
<td>±1.31E+01</td>
</tr>
<tr>
<td>$F_{23}$</td>
<td>5.43E+02</td>
<td>8.36E+02</td>
<td>1.31E+03</td>
<td>9.62E+02</td>
<td>5.34E+02</td>
<td>5.34E+02</td>
<td>5.35E+02</td>
<td>5.34E+02</td>
<td>5.71E+02</td>
</tr>
<tr>
<td></td>
<td>±1.57E+12</td>
<td>±1.13E+02</td>
<td>±1.10E+02</td>
<td>±3.27E+02</td>
<td>±4.21E-04</td>
<td>±2.26E-03</td>
<td>±1.88E+00</td>
<td>±1.57E-04</td>
<td>±2.99E+01</td>
</tr>
<tr>
<td>$F_{24}$</td>
<td>2.00E+02</td>
<td>3.69E+02</td>
<td>1.57E+03</td>
<td>2.35E+02</td>
<td>2.00E+02</td>
<td>2.00E+02</td>
<td>2.00E+02</td>
<td>2.00E+02</td>
<td>2.00E+02</td>
</tr>
<tr>
<td></td>
<td>±3.59E+02</td>
<td>±2.79E+02</td>
<td>±1.04E+02</td>
<td>±8.36E+01</td>
<td>±2.96E-09</td>
<td>±2.00E+01</td>
<td>±2.67E-12</td>
<td>±2.72E-12</td>
<td>±2.72E-12</td>
</tr>
<tr>
<td>$F_{25}$</td>
<td>1.35E+03</td>
<td>1.43E+03</td>
<td>2.00E+03</td>
<td>1.32E+03</td>
<td>2.17E+02</td>
<td>2.13E+02</td>
<td>2.07E+02</td>
<td>2.00E+02</td>
<td>2.22E+02</td>
</tr>
<tr>
<td></td>
<td>±3.59E+02</td>
<td>±6.85E+01</td>
<td>±7.30E+01</td>
<td>±3.66E+01</td>
<td>±1.59E-01</td>
<td>±1.15E+00</td>
<td>±6.30E+00</td>
<td>±1.96E+00</td>
<td>±1.37E+01</td>
</tr>
<tr>
<td>Wilcoxon-Test</td>
<td>16-1-3</td>
<td>16-0-4</td>
<td>18-1-1</td>
<td>16-0-4</td>
<td>12-5-3</td>
<td>10-3-7</td>
<td>12-2-6</td>
<td>11-2-7</td>
<td>—</td>
</tr>
</tbody>
</table>
Fig. 1: The Top-K, K = 1, 2, 9, curves of the algorithms.
Results of NCS-C

Search trajectories of NCS-C, PHC, SaDE, and CLPSO on the 2-d problem ($F_{19}$ of CEC2005 benchmark set)
Outline

• Motivation

• Negatively Correlated Search: The General Idea

• NCS-C: A simple Instantiation

• More Interpretations

• Summary and Future Work
More interpretations

• NCS is relevant but rather different from a number of well-established search mechanisms.
NCS vs. Diversity Maintenance

- Existing diversity maintenance schemes, e.g., Niching (fitness sharing, crowding, etc.) and Scatter Search, emphasize distance (diversity) between *individuals* in the search space.

- NCS encourages diversity between *search behaviors*. 
NCS vs. Multi-population EAs

• In Multi-population EAs, search processes (sub-populations) communicate to share the promising region that has been found. Shared information will attract search processes to move towards each other.

• In NCS, a search process is pushed against the others through communication.
NCS vs. TS

• Tabu search employs Tabu list to keep track of previously visited solutions/regions, i.e., it *looks into the past*.

• NCS considers the *future behaviors* of search processes. (“anticipate” the positions of newly generated solutions).
Outline

• Motivation

• Negatively Correlated Search: The General Idea

• NCS-C: A simple Instantiation

• More Interpretations

• Summary and Future Work
A new population-based search framework, NCS, is described.

NCS maintains a population of search processes that show negatively correlated behaviors by explicitly encouraging the distances between probability distributions corresponding to search processes.

The potential of NCS is demonstrated by a simple instantiation, i.e., NCS-C.

NCS isn’t restricted to selecting offspring in (1+1) EAs, but can also encompass existing EAs as its component search process, as well as be used for adapting search operators.
Directions for Future Research

• Deeper (theoretical) analysis on the pros/cons of NCS.

• Sampling techniques for the estimation of distance/correlation.

• With appropriate definition of search behaviors and correlations, NCS can be generalized to other problem classes.
  – Seeking multiple optima
  – Multi-objective optimization
  – Multi-task optimization
Collaborators

Mr. Peng Yang

Professor Xin Yao
Reference


• Matlab Codes of NCS-C available at: http://staff.ustc.edu.cn/~ketang/codes/NCS.html
Thanks for your time!

Questions/comments?