Problem Features versus Algorithm Performance on Rugged Multiobjective Combinatorial Fitness Landscapes

Thomas Dirkse   Paolo Servillo
Alejandro Barredo Arrieta
Introduction

The goal of this article is to characterize the impact of problem features on performance of Evolutionary Multi-Objective combinatorial optimization problems. The authors:

- Measured two algorithms for binary optimization with tunable parameters
  - More specifically: runtime of GSEMO and PLS neighborhood search to find a $1+\epsilon$ approximation of the Pareto set
- Defined relevant problem features for fitness landscape and studied their intercorrelation
- Determined individual and joint effects of problem features on performances of both algorithms using multilevel regression
Multi-objective Optimization & Pareto set

- Optimization problem with more than one objective function

Wikipedia:

**Pareto efficiency** or **Pareto optimality** is a state of allocation of resources from which it is impossible to reallocate so as to make any one individual or preference criterion better off without making at least one individual or preference criterion worse off.
Motivation

- These type of problems are often seen in real-life application
- Despite the increasing number of heuristics for EMO, they are mainly based on intuition
- Little knowledge compared to the single objective case
- Identify a number of general-purpose features characterizing problem hardness
Context and definitions

Fitness Landscape Analysis

- A fitness landscape is a **triplet** \((X,N,\varphi)\), where \(X\) is the solution space, \(N\) is a neighborhood relation and \(\varphi\) is a scalar black-box fitness function (here to maximize)
- An **adaptive walk** is a walk such that each solution is better than the previous
  - Its length is an estimator of the diameter of local optima basin of attraction
- The **autocorrelation function** and the **correlation length** allow to characterize the ruggedness of a landscape
  - Computed over a random walk (longer is more accurate), higher correlation length means smoother landscape
Context and definitions

Multi-Objective Optimization

- Maximizing an **objective function** $f : X \rightarrow Z$, which maps any solution to a vector of scalar numbers
- **Solution space** is a discrete set $X = \{0,1\}^N$, where $N$ is the problem size
- An objective vector is **non-dominated** if there does not exists another vector with at least one higher value.
- A solution $x$ is non-dominated, or **Pareto Optima**, if $f(x)$ is non-dominated
Context and definitions

$pMNK$-Landscapes

- Synthetic problem model used for constructing tunable multiobjective multimodal landscapes
- Extend $NK$-Landscapes
  - The objective vector of size $M$, contains each objective function to maximise
  - For each function in the vector and for each variable in a solution there is a uniformly distributed real value associated to the variable and its $K$ interactions
  - The individual contribution of the variable depends on its value and the value of $K$ other (and the fitness of the solution is the average of these values)
  - By increasing the number of interaction $K$ from 0 to $N - 1$, instances are gradually tuned from smooth to rugged
Context and definitions

\( \rho \text{MNK-Landscapes} \)

\[
\begin{align*}
\max & \quad f_i(x) = \frac{1}{N} \sum_{j=1}^{N} f_{ij}(x_j, x_{j1}, \ldots, x_{jk}) \quad i \in \{1, \ldots, M\} \\
\text{s.t.} & \quad x_j \in \{0, 1\} \quad j \in \{1, \ldots, N\}.
\end{align*}
\]

- In \( \rho \text{MNK-Landscapes} \), the real values associated with the interactions additionally follow a multivariate distribution of dimension \( M \), defined by a covariance matrix, where \( \rho \) defines the correlation among the objectives (positive decreases conflict between different objectives)
Context and definitions

PLS & GSEMO

Algorithm 2: PLS

1. Choose an initial solution \( x_0 \) uniformly from \( X \);
2. \( A \leftarrow \{x_0\} \);
3. repeat
   4. Select a non-visited element \( x \) out of \( A \) uniformly;
   5. Create \( \mathcal{N}(x) \) by flipping each bit of \( x \) in turns;
   6. Flag \( x \) as visited;
   7. \( A \leftarrow \) non-dominated solutions from \( A \cup \mathcal{N}(x) \);
4. until all-visited ∨ success ∨ maxeval;

- Local optimizer
- Has termination condition
Context and definitions

PLS & GSEMO

**Algorithm 1: GSEMO**

1. Choose an initial solution $x_0$ uniformly from $X$;
2. $A \leftarrow \{x_0\}$;
3. repeat
4. | Select an element $x$ out of $A$ uniformly;
5. | Create $x'$ by flipping each bit of $x$ with probability $1/N$;
6. | $A \leftarrow$ non-dominated solutions from $A \cup \{x'\}$;
7. until $\text{success} \lor \text{maxeval}$;

- Global optimizer (non-zero probability of reaching any other solution)
- No termination condition
Context and definitions

Estimated Runtime ($\text{ert}$)

- Recall that we are looking for a $(1+\varepsilon)$–approximation of the Pareto set
  - There is always one (Papadimitriou and Yannakakis, 2000).
- Algorithm performances are measured in terms of function evaluations
- In case of success, they record the number of evaluations until a $(1+\varepsilon)$–approximation was found. In case of failure, they restart at random.
- Thus, they obtain a “simulated runtime” from a set of independent Bernoulli trials

\[
\mathbb{E}[T] = \left(\frac{1 - p_s}{p_s}\right) \mathbb{E}[T_f] + \mathbb{E}[T_s] \\
\text{ert} = \left(\frac{1 - \hat{p}_s}{\hat{p}_s}\right) T_{\text{max}} + \frac{1}{t_s} \sum_{i=1}^{t_s} T_i
\]
Experimental Analysis

Performance depending on:

- Problem size \((N)\)
- Epistasis \((K)\)
- Objective space dimension \((M)\)
- Objective correlation \((\rho)\)

These are the parameters of the so called “\(\rho MNK\)-landscapes”
Experimental Analysis

Performance depending on:

- Problem size \((N)\) length of (binary) strings: \(x_1, \ldots, x_n\)
- Epistasis \((K)\)
- Objective space dimension \((M)\)
- Objective correlation \((\rho)\)

These are the parameters of the so called “\(\rho MNK\)-landscapes”
Experimental Analysis

Performance depending on:

- Problem size \((N)\) length of (binary) strings: \(x_1, \ldots, x_n\)
- Epistasis \((K)\) “ruggedness” (interaction between \(x_i\)’s)
- Objective space dimension \((M)\)
- Objective correlation \((\rho)\)

These are the parameters of the so called “\(\rho MNK\)-landscapes”
Experimental Analysis

Performance depending on:

- Problem size $(N)$: length of (binary) strings: $x_1, \ldots, x_n$
- Epistasis $(K)$: “ruggedness” (interaction between $x_i$’s)
- Objective space dimension $(M)$: dimension of the objective vector
- Objective correlation $(\rho)$

These are the parameters of the so called “$\rho$MNK-landscapes”
Experimental Analysis

Performance depending on:

- Problem size \((N)\) length of (binary) strings: \(x_1, ..., x_n\)
- Epistasis \((K)\) “ruggedness” (interaction between \(x_i\)’s)
- Objective space dimension \((M)\) dimension of the objective vector
- Objective correlation \((\rho)\) correlation among the objectives
  \[ \rho(M-1) > -1 \]

These are the parameters of the so called “\(\rho MNK\)-landscapes” (Tunable topology)
Experimental Analysis

Problem instances:

- \( K \in \{2, 4, 6, 8, 10\} \) (epistatic degree)
- \( M \in \{2, 3, 5\} \) (objective space dimension)
- \( \rho \in \{-0.9, -0.7, -0.4, -0.2, 0.0, 0.2, 0.4, 0.7, 0.9\} \) (objective correlation)
- \( N = 18 \) (problem size) \( \rightarrow \) possibility to enumerate all solutions in solution space (\(|X| = 2^{18}\))
Experimental Analysis

Experimental setup:

- 30 different landscapes randomly generated for each combination of parameters → 3300 problem instances

- Target tolerance: $\varepsilon = 0.1$

- Time limit: $T_{max} = 2^N \cdot 10^{-1} < 26251$ function evaluations (if target tolerance is not reached)

- Each algorithm is executed 100 times per instance → estimated runtime is calculated from success rate and expected evaluations
Distribution of $e_{rt}$ w.r.t. objective correlation $\rho$:
Interaction plots between $\text{avg. ert}$ and the benchmark parameters:
What features characterizes multi-objective combinatorial landscapes?

We have already seen:

- Benchmark parameters: $\rho, M, N, K$ (problem size $N$ is kept constant)

Problem features from literature:

- **The Pareto Set:** npo, hv, avgd, maxd, supp
- **The Pareto Graph:** ncomp, lcomp, dconn
- **Ruggedness and Modality:** nplo, ladapt, corhv, corlhv
Table 1: Summary of $\rho$MNK-landscape benchmark parameters and problem instance features investigated in the article.

<table>
<thead>
<tr>
<th>Benchmark parameters (3)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho$</td>
<td>Correlation between the objective function values</td>
</tr>
<tr>
<td>$M$</td>
<td>Number of objective functions</td>
</tr>
<tr>
<td>$K$</td>
<td>Number of variable interactions (epistasis)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Problem features (12)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>npo</td>
<td>Number of Pareto optimal solutions</td>
</tr>
<tr>
<td>hv</td>
<td>Hypervolume (Zitzler et al., 2003) of the Pareto set</td>
</tr>
<tr>
<td>avgd</td>
<td>Average distance between Pareto optimal solutions</td>
</tr>
<tr>
<td>maxd</td>
<td>Maximum distance between Pareto optimal solutions</td>
</tr>
<tr>
<td>supp</td>
<td>Proportion of supported solutions in the Pareto set</td>
</tr>
<tr>
<td>nplo</td>
<td>Number of Pareto local optima</td>
</tr>
<tr>
<td>ladapt</td>
<td>Length of a Pareto-based adaptive walk</td>
</tr>
<tr>
<td>ncomp</td>
<td>Relative number of connected components</td>
</tr>
<tr>
<td>lcomp</td>
<td>Proportional size of the largest connected component</td>
</tr>
<tr>
<td>dconn</td>
<td>Minimal distance to connect the Pareto graph</td>
</tr>
<tr>
<td>corhv</td>
<td>First autocorrelation coefficient of solution hypervolume</td>
</tr>
<tr>
<td>cor1hv</td>
<td>First autocorrelation coefficient of local hypervolume</td>
</tr>
</tbody>
</table>

(References: Knowles and Corne, 2003; Aguirre and Tanaka, 2007; Liegooghe, Verel, Aguirre, and Tanaka, 2013; Paquete et al., 2007; Verel et al., 2013; Paquete and Stützle, 2009; Verel et al., 2011; Paquete and Stützle, 2009; Liegooghe, Verel, Aguirre, and Tanaka, 2013)
Correlation matrix (Pearson) between all feature pairs:

<table>
<thead>
<tr>
<th></th>
<th>corhtv</th>
<th>K</th>
<th>0.58</th>
<th>-0.5</th>
<th>0.98</th>
<th>corhtv</th>
</tr>
</thead>
<tbody>
<tr>
<td>corhtv</td>
<td>0.13</td>
<td>0.14</td>
<td>-0.97</td>
<td>0.14</td>
<td>-0.29</td>
<td>-0.58</td>
</tr>
<tr>
<td>K</td>
<td>0.03</td>
<td>-0.99</td>
<td>0.15</td>
<td>-0.23</td>
<td>-0.59</td>
<td>-0.6</td>
</tr>
<tr>
<td>0.58</td>
<td>-0.36</td>
<td>-0.38</td>
<td>-0.73</td>
<td>0.38</td>
<td>-0.01</td>
<td>-0.73</td>
</tr>
<tr>
<td>-0.5</td>
<td>0.58</td>
<td>-0.73</td>
<td>0.58</td>
<td>-0.5</td>
<td>0.98</td>
<td>-0.51</td>
</tr>
</tbody>
</table>

- **0.21**: Correlation coefficient between feature 1 and feature 2.
- **0.26**: Correlation coefficient between feature 3 and feature 4.
- **0.65**: Correlation coefficient between feature 5 and feature 6.

The table shows the correlation matrix for all features, with values ranging from -1 to 1. The significance of each correlation is indicated by stars: *** indicates p < 0.001, ** indicates p < 0.01, * indicates p < 0.05, and no star indicates p > 0.05.
Some observations:

- The number of objectives $M$ and the objective correlation $\rho$ are moderately correlated with cardinality of the Pareto set $\log(n_{po})$

- $M$ and $\rho$ alone do not explain the number of nondominated solutions. Multilinear regression based on $M$ and $\rho$ explains 70% of the $\log(n_{po})$ variance (high correlation coefficient)

The researchers concluded from this that:

*The impact of many-objective fitness landscapes on the search process cannot be analyzed properly without taking the objective correlation into account.*
Some observations:

Characterizations of the topology of Pareto optimal solutions:

- The relative number of connected components $n_{\text{comp}}$ is positively correlated with the minimal distance to connect all the components $d_{\text{conn}}$, whereas the relative size of the largest component $l_{\text{comp}}$ is negatively correlated with the number of components.

- The minimal distance $d_{\text{conn}}$ is also highly correlated with the maximal Hamming distance between Pareto optimal solutions $\text{maxd}$.
Some observations:

Characterizations of the ruggedness and multimodality of the fitness landscape:

- The correlation between the number of variable interactions (epistasis) $K$ and the number of Pareto global or local optima is low.

- The log-transformed value of $n po$ is highly linearly correlated to the length of a Pareto-based adaptive walk $l_{adap t}$ This potentially allows one to estimate the number of Pareto local optima for large-size problem instances.
Feature-Based Analysis

In this section, the authors try to relate the problem features to the performance (estimated runtime) of EMO algorithms.

- Correlation analysis to measure feature-performance correlation
- Multilinear regression model: To find the relation between each of the features and performance.
- Multilevel regression model: To find the relation between landscape ruggedness, multimodality and the performance of the algorithm.
Feature-performance correlation

- **Plo**: Pareto local optima
- **Hv**: Hypervolume of Pareto set
- **Avgd**: Average distance between Pareto optimal solutions
- **Maxd**: Maximum distance between Pareto optimal solutions
- **Supp**: Proportion of supported solutions
- **Nplo**: Number of Pareto optima
- **Ladapt**: Length of a Pareto-based adaptive walk
- **Ncomp**: Relative number of connected components
- **Lcomp**: Proportional size of the largest connected component
- **Dconn**: Minimum distance to connect the Pareto graph
- **Corhv**: First autocorrelation coefficient of solution hypervolume
- **Corlhv**: First autocorrelation coefficient of local hypervolume
Performance-feature association

- Kendall’s Tau is a statistic used to measure the ordinal association between two measures.
Conditional impact of each feature on performance
Mixed model fitting for each feature
Relative importance of features as performance predictors
Correlation between ruggedness and performance
Conclusions

- This paper proposed a general-purpose methodology to understand the impact of the problem characteristics and fitness landscape features on the performance of EMO algorithms.
- Has shown the importance of ruggedness and multimodality. Pointing to PLS as a better choice when the landscape is rugged but there are few Pareto local optimas.
- Hypervolume of the Pareto set is a key feature for performance.