Fitness landscape analysis: an overview for single- and multi-objective optimization problems

Sébastien Verel

LISIC - Université du Littoral Côte d'Opale, Calais, France
http://www-lisic.univ-littoral.fr/~verel/

University Exeter

18th, May, 2017
Personal situation map
Situation map: Université du Littoral Côte d’Opale

LISIC: laboratory of computer science, signal, image

Boulogne (fishing port), Calais (transportation port), Dunkerque (industrial port), Saint Omer (glass industry).

⇝ Lille (0h30), London (1h), Bruxelles (1h15), Paris (1h30), Amsterdam (2h)
Outline

- Short introduction on fitness landscape
  multimodality, ruggedness
- Single-objective optimization :
  Local Optima Network
- Multi-objective optimization :
  Features of fitness landscape
Mono-objective Optimization

- **Search space**: set of candidate solutions
  \[ X \]
- **Objective function**: quality criterion (or non-quality)
  \[ f : X \rightarrow \mathbb{R} \]

- \( X \) discrete: combinatorial optimization
  \( X \subset \mathbb{R}^n \): numerical optimization

Solve an optimization problem (maximization)

\[ X^* = \arg\max_X f \]

or find an approximation of \( X^* \).
Introduction

Multimodal fitness landscape

Local Optima Network

Multiobjective Optimization

Context: black-box optimization

\[ x \rightarrow f(x) \]

No information on the objective definition function \( f \)

Objective function:

- can be irregular, non continuous, non differentiable, etc.
- given by a computation or a simulation
Real-world black-box optimization: first example
PhD of Mathieu Muniglia, Saclay Nuclear Research Centre (CEA), Paris

\( x \rightarrow f(x) \)

\((73, \ldots, 8) \rightarrow \Delta_z P\)

Multi-physic simulator

source: Encyclopædia Britannica Online.
Fitness landscapes in (evolutionary) biology

- Metaphorical uphill struggle across a "fitness landscape"
  - mountain **peaks** represent high "fitness" (ability to survive),
  - **valleys** represent low fitness.

- Evolution proceeds:
  population of organisms
  performs an "**adaptive walk**"

becareful: "2 dimensions instead of many thousands"
Definition of fitness landscape for optimization [Sta02]

Fitness landscape \((X, \mathcal{N}, f)\):

- **search space**: \(X\)
- **neighborhood relation**: \(\mathcal{N} : X \rightarrow 2^X\)
- **objective function**: \(f : X \rightarrow \mathbb{R}\)
Fitness landscape analysis

Algebraic approach, grey-box:

$$\Delta f = \lambda (f - \bar{f})$$

Statistical approach, black-box:

Problems $\mapsto$ Features

$\mapsto$ Algorithm $\mapsto$ Performances
Position of fitness landscape analysis

Selection of local search algorithm: Rice framework

Figure 1.1: A framework for describing the general problems of algorithm selection and performance prediction based on problem features (based Rice’s model [132]).

Position of fitness landscape analysis

Selection of local search algorithm: Rice framework

**Figure 1.1:** A framework for describing the general problems of algorithm selection and performance prediction based on problem features (based Rice’s model [132]).

Multimodal Fitness landscapes

Local optima $x^*$

no neighbor solution with strictly higher fitness value (maximization)

$$\forall x \in \mathcal{N}(x^*), \quad f(x) \leq f(x^*)$$
Sampling local optima

Basic estimator (Alyahya, K., & Rowe, J. E. 2016 [AR16])

Expected proportion of local optima :

Proportion of local optima in a sample of random solutions

- Complexity : \( n \times |\mathcal{N}| \)
- Pros :
  - unbiased estimator, strong and easy statistical tools
- Cons :
  - ”resolution rate” is \( 1/n \)
Sampling local optima by adaptive walks

**Adaptive walk**

\((x_1, x_2, \ldots, x_\ell)\) such that \(x_{i+1} \in \mathcal{N}(x_i)\) and \(f(x_i) < f(x_{i+1})\)
Sampling local optima by adaptive walks

Adaptive walk

\[(x_1, x_2, \ldots, x_\ell)\] such that \(x_{i+1} \in \mathcal{N}(x_i)\) and \(f(x_i) < f(x_{i+1})\)

Hill-Climbing algorithm (first-improvement)

Choose initial solution \(x \in X\)

repeat

choose \(x' \in \{y \in \mathcal{N}(x) \mid f(y) > f(x)\}\)

if \(f(x) < f(x')\) then

\(x \leftarrow x'\)

end if

until \(x\) is a Local Optimum
Sampling local optima by adaptive walks

**Adaptive walk**

\[(x_1, x_2, \ldots, x_\ell) \text{ such that } x_{i+1} \in \mathcal{N}(x_i) \text{ and } f(x_i) < f(x_{i+1})\]

**Hill-Climbing algorithm (first-improvement)**

Choose initial solution \(x \in X\)

repeat

choose \(x' \in \{y \in \mathcal{N}(x) \mid f(y) > f(x)\}\)

if \(f(x) < f(x')\) then

\(x \leftarrow x'\)

end if

until \(x\) is a Local Optimum

**Basin of attraction of \(x^*\)**

\[\{x \in X \mid \text{HillClimbing}(x) = x^*\}\]
Multimodal Fitness landscapes and difficulty

The idea:

- if the size of attractive basin of global optimum is "small",
- then, the "time" to find the global optimum is "long"

Optimisation difficulty:
Number and size of attractive basins (Garnier et al. [GK02])

Feature to estimate basin size:

- **Length of adaptive walks**

  *complexity*: sample size $\times \ell \times |\mathcal{N}|$
Multimodal Fitness landscapes and difficulty

The idea:
- if the size of attractive basin of global optimum is "small",
- then, the "time" to find the global optimum is "long"

Optimisation difficulty:
Number and size of attractive basins (Garnier et al. [GK02])

Feature to estimate basin size:
- **Length of adaptive walks**

*complexity*: sample size $\times \ell \times |\mathcal{N}|$

ex. nk-landscapes with $n = 512$
Random Walk to measure the ruggedness

Random walk:

- \((x_1, x_2, \ldots)\) where \(x_{i+1} \in \mathcal{N}(x_i)\) and equiprobability on \(\mathcal{N}(x_i)\)

The idea:

- if the profile of fitness is irregular,
- then, the "information" between neighbors is low.

Feature:

- Study the fitness profile like a signal
Autocorrelation function of time series of fitnesses along a random walk (Weinberger 90 [Wei90]):

\[ \rho(n) = \frac{\mathbb{E}[(f(x_i) - \bar{f})(f(x_{i+n}) - \bar{f})]}{\text{var}(f(x_i))} \]

Autocorrelation length \( \tau = \frac{1}{\rho(1)} \)

"How many random steps such that correlation becomes insignificant"

- small \( \tau \): rugged landscape
- long \( \tau \): smooth landscape

Complexity: sample size \( \approx 10^3 \)
Ruggedness decreases with the size of those problems

<table>
<thead>
<tr>
<th>Problem</th>
<th>parameter</th>
<th>$\rho(1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>symmetric TSP</td>
<td>$n$ number of towns</td>
<td>$1 - \frac{4}{n}$</td>
</tr>
<tr>
<td>anti-symmetric TSP</td>
<td>$n$ number of towns</td>
<td>$1 - \frac{4}{n-1}$</td>
</tr>
<tr>
<td>Graph Coloring Problem</td>
<td>$n$ number of nodes $\alpha$ number of colors</td>
<td>$1 - \frac{2\alpha}{(\alpha-1)n}$</td>
</tr>
<tr>
<td>NK landscapes</td>
<td>$N$ number of proteins $K$ number of epistasis links</td>
<td>$1 - \frac{K+1}{N}$</td>
</tr>
<tr>
<td>random max-k-SAT</td>
<td>$n$ number of variables $k$ variables per clause</td>
<td>$1 - \frac{k}{n(1-2^{-k})}$</td>
</tr>
</tbody>
</table>
Multimodal fitness landscape

Local Optima Network

Multiobjective Optimization

Introduction

Tuning using ruggedness features: real-world application
PhD of Mathieu Muniglia, Saclay Nuclear Research Centre (CEA), Paris

New control of nuclear power plant to introduce renewable energies.

multi-physic simulator
≈ 30min of computation

Parallel M/S EA:
mutation parameters?

performances is correlated to ruggedness

ruggedness depends on parameters
published at evoCOP 2017:

Towards Landscape-Aware Automatic Algorithm Configuration: Preliminary Experiments on Neutral and Rugged Landscapes.

Arnaud Liefooghe, Bilel Derbel, Sébastien Verel, Hernan Aguirre, and Kiyoshi Tanaka.

See slides presented by Bilel Derbel (thanks to him!)
Joint work

> Marco Tomassini, Lausanes University, Switzerland
> Gabriela Ochoa, University of Stirling, Scotland
> Fabio Daolio, University of Stirling, Scotland

Many thanks!
Key idea: Complex system tools

Principle of variables aggregation

A model for dynamical systems with two scales (time/space)
- Split the state space according to the different scales
- Study the system at the large scale
Key idea: Complex system tools

Principle of variables aggregation

A model for dynamical systems with two scales (time/space)

- Split the state space according to the different scales
- Study the system at the large scale

Variables aggregation for fitness landscape

- At solutions level (small scale):
  - Stochastic local search operator,
  - Exponential number of solutions,
  - Exponential size of the stochastic matrix of the process (Markov chain)

- Projection on a relevant space:
  - Reduce the size of state space
  - Potentially lose some information
  - Relevant information remains when:
    \[ p(op(x)) \approx op'(p(x)) \]
Key idea: Complex system tools

Principle of variables aggregation

A model for dynamical systems with two scales (time/space)
- Split the state space according to the different scales
- Study the system at the large scale

Variables aggregation for fitness landscape

- At solutions level (small scale):
  - Stochastic local search operator,
  - Exponential number of solutions,
  - Exponential size of the stochastic matrix of the process (Markov chain)

- Projection on a relevant space:
  - Reduce the size of state space
  - Potentially lose some information
  - Relevant information remains when:
    \[ p(\text{op}(x)) \approx \text{op'}(p(x)) \]
Key idea: Complex system tools

Complex network

Bring the tools of complex networks analysis to the study the structure of combinatorial fitness landscapes

Methodology

- **Design a network** that represents the landscape
  - Vertices: local optima
  - Edges: a notion of adjacency between local optima

- **Extract features**:
  - "complex" network analysis

- **Use the network features**:
  - search algorithm design, difficulty, etc.

Complex networks

Scale free network
(Watts and Strogatz, 1998 [WS98])

Small world network
(Barabasi and Albert, 1999 [BA99])
Energy surface and inherent networks

**Inherent network**

- **Nodes**: energy minima
- **Edges**: two nodes are connected if the energy barrier separating them is sufficiently low (transition state)

(a) Energy surface
(b) Contours plot: partition of states space into basins of attraction
(c) Landscape as a network


Local Optima Network

Definition: Local Optima Network (LON)

Orienter weighted graph \((V, E, w)\)

- **Notes** \(V\) : set of local optima \(\{LO_1, \ldots, LO_n\}\)
- **Edges** \(E\) : notion of connectivity between local optima

Escape edges

Edge \(e_{ij}\) between \(LO_i\) and \(LO_j\)

if \(\exists x : \text{distance}(LO_i, x) \leq D\) and \(x \in b_j\).

Weights

\(w_{ij} = \#\{x \in X \mid d(LO_i, x) \leq D, x \in b_j\}\)

can be normalized by the number of solutions at distance \(D\)
Local Optima Network

Definition: Local Optima Network (LON)

Orienter weighted graph $(V, E, w)$
- Notes $V$: set of local optima $\{LO_1, \ldots, LO_n\}$
- Edges $E$: notion of connectivity between local optima

Escape edges

Edge $e_{ij}$ between $LO_i$ and $LO_j$
- if $\exists x: \text{distance}(LO_i, x) \leq D$ and $x \in b_j$.

Weights

$w_{ij} = \#\{x \in X \mid d(LO_i, x) \leq D, x \in b_j\}$
- can be normalized by the number of solutions at distance $D$
Features of local optima network

- $nv$: number of vertices
- $lv$: average path length
  \[ d_{ij} = 1/w_{ij} \]
- $lo$: path length to best
- $fnn$: fitness corr.
  \( (f(x), f(y)) \) with \((x, y) \in E\)
- $wii$: self loops
- $wcc$: weighted clust. coef.
- $zout$: out degree
- $y2$: disparity
- $knn$: degree corr.
  \( (\deg(x), \deg(y)) \) with \((x, y) \in E\)
Benchmark : NK-landscapes
[Kauffman 1993] [Kau93]

\[
x \in \{0, 1\}^n \quad f(x) = \frac{1}{n} \sum_{i=1}^{n} f_i(x_j, x_{i_1}, \ldots, x_{i_k})
\]

Two parameters :
- Problem size \( n \)
- Non-linearity \( k < n \)
  (multi-modality, epistatic interactions)
  - \( k = 0 \) : linear problem, one single maxima
  - \( k = n - 1 \) : random problem, number of local optima \( \frac{2^N}{N+1} \)

remarks : ”same” results with QAP, flow shop.
**Structure of Local Optima Network**

- NK-landscapes (small instances): Most of the features are correlated with $K$ relevance of LON definition.

- LON is **not a random** network (NK, QAP, FSSP): Highly clustered network, Distribution of weights and degrees have long tail, etc.
LON to compare of problem difficulty
Local Optima Network for Quadratic Assignment Problem (QAP) [DTVO11]

→ Community detection in LON for
Random instance

Real-like instance

Structure of the LON related to problem difficulty
LON to compare algorithm components

Comparaison of operators for Flow Shop Scheduling Problem

Comparaison of pivot rule in hill-climbing for NK-landscapes

<table>
<thead>
<tr>
<th>$K$</th>
<th>$\bar{n}_e/\bar{n}_v$</th>
<th>$Y$</th>
<th>$d$</th>
<th>$d_{best}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b-LON</td>
<td>f-LON</td>
<td>b-LON</td>
<td>f-LON</td>
</tr>
<tr>
<td>2</td>
<td>0.81</td>
<td>0.96</td>
<td>0.326</td>
<td>0.110</td>
</tr>
<tr>
<td>4</td>
<td>0.60</td>
<td>0.92</td>
<td>0.137</td>
<td>0.033</td>
</tr>
<tr>
<td>6</td>
<td>0.32</td>
<td>0.79</td>
<td>0.084</td>
<td>0.016</td>
</tr>
<tr>
<td>8</td>
<td>0.17</td>
<td>0.65</td>
<td>0.062</td>
<td>0.011</td>
</tr>
<tr>
<td>10</td>
<td>0.09</td>
<td>0.53</td>
<td>0.050</td>
<td>0.009</td>
</tr>
</tbody>
</table>
LON features vs. performance : simple correlation

Algorithm : Iterated Local Search on NK-landscapes with $N = 18$
Performance : $ert = \mathbb{E}(T_s) + \left(\frac{1-p_s}{p_s}\right) T_{max}$

<table>
<thead>
<tr>
<th>$n_v$</th>
<th>$\ddot{d}_{best}$</th>
<th>$\ddot{d}$</th>
<th>fnn</th>
<th>$w_{ii}$</th>
<th>$\ddot{C}^w$</th>
<th>zout</th>
<th>$\ddot{Y}$</th>
<th>knn</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.885</td>
<td>0.915</td>
<td>0.006</td>
<td>-0.830</td>
<td>-0.883</td>
<td>-0.875</td>
<td>0.885</td>
<td>-0.883</td>
<td>-0.850</td>
</tr>
</tbody>
</table>

![Scatter plots showing the correlation between LON features and performance metrics.](scatter_plots.png)
LON features vs. performance: multi-linear regression

1. Multiple **linear** regression on all possible predictors:

\[
\log(ert) = \beta_0 + \beta_1 k + \beta_2 \log(nv) + \beta_2 lo + \cdots + \beta_{10} knn + \varepsilon
\]

2. Step-wise **backward elimination** of each predictor in turn.

<table>
<thead>
<tr>
<th>Predictor</th>
<th>( \hat{\beta}_i )</th>
<th>Std. Error</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Intercept)</td>
<td>10.3838</td>
<td>0.58512</td>
<td>9.24 \cdot 10^{-47}</td>
</tr>
<tr>
<td>lo</td>
<td>0.0439</td>
<td>0.00434</td>
<td>1.67 \cdot 10^{-20}</td>
</tr>
<tr>
<td>zout</td>
<td>-0.0306</td>
<td>0.00831</td>
<td>2.81 \cdot 10^{-04}</td>
</tr>
<tr>
<td>y2</td>
<td>-7.2831</td>
<td>1.63038</td>
<td>1.18 \cdot 10^{-05}</td>
</tr>
<tr>
<td>knn</td>
<td>-0.7457</td>
<td>0.40501</td>
<td>6.67 \cdot 10^{-02}</td>
</tr>
</tbody>
</table>

Multiple R-squared: 0.8494, Adjusted R-squared: 0.8471.
Sampling methodology for large size instances

From the sampling of large-size complex network:

- Random walk on the network
- Breadth-First-Search

**Procedure** $\text{LONSampling}(d, m, l)$

- $x_0 \leftarrow hc(x)$ with $x$ random solution
- for $t \leftarrow 0, \ldots, l - 1$ do
  - $\text{Snowball}(d, m, x_t)$
  - $x_{t+1} \leftarrow \text{RandomWalkStep}(x_t)$
- end for
Performance prediction based on estimated features

- Optimization scenario using off-the-shelf metaheuristics: TS, SA, EA, ILS on 450 instances for NK and QAP.
- Performance measures:
  - average fitness / average rank
- Model of regression:
  - linear model / random forest
- Set of features:
  - basic: 1st autocorr. coeff. of fitness (rw of length $10^3$),
    - Avg. fitness of local optima ($10^3$ hc),
    - Avg. length to reach local optima ($10^3$ hc).
  - lon: see previous,
  - all: basic and lon features
- Quality measure of regression:
  - $R^2$ on cross-validation (repeated random sub-sampling)
Scatter plots of the observed-estimated performance

\[ \text{basic, } R^2 = 0.9327 \]  \hspace{1cm}  \[ \text{lon, } R^2 = 0.9601 \]  \hspace{1cm}  \[ \text{all, } R^2 = 0.9643 \]

Very good prediction using LON features
Algorithm selection: portfolio scenario

- Portfolio of 4 metaheuristics: TS, SA, EA, ILS
- Classification task: selection of one of the best metaheuristic
- Models: logit, random forest, svm
- Quality of classification: error rate (algo. is not one of the best) on cross-validation.

<table>
<thead>
<tr>
<th>Probl.</th>
<th>Feat.</th>
<th>logit</th>
<th>rf</th>
<th>svm</th>
<th>cst</th>
<th>rnd</th>
</tr>
</thead>
<tbody>
<tr>
<td>NK</td>
<td>basic</td>
<td>0.0379</td>
<td>0.0278</td>
<td>0.0158</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>lon</td>
<td>0.0203</td>
<td>0.0249</td>
<td>0.0168</td>
<td>0.4711</td>
<td>0.6749</td>
</tr>
<tr>
<td></td>
<td>all</td>
<td>0.0244</td>
<td>0.0269</td>
<td>0.0165</td>
<td></td>
<td></td>
</tr>
<tr>
<td>QAP</td>
<td>basic</td>
<td>0.0142</td>
<td>0.0107</td>
<td>0.0771</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>lon</td>
<td>0.0156</td>
<td>0.0086</td>
<td>0.0456</td>
<td>0.4222</td>
<td>0.6706</td>
</tr>
<tr>
<td></td>
<td>all</td>
<td>0.0161</td>
<td>0.0106</td>
<td>0.0431</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Multiobjective optimization

Multiobjective optimization problem

- Search space: decision space \( X \)
- Objective function: objective space \( f : X \rightarrow \mathbb{R}^m \)

Component \( f_i \): objective, criterion.
Definition

**Pareto dominance relation**

A solution $x \in X$ **dominates** a solution $x' \in X$ ($x' \prec x$) iff

1. $\forall i \in \{1, 2, \ldots, M\}, f_i(x') \leq f_i(x)$
2. $\exists j \in \{1, 2, \ldots, M\}$ such that $f_j(x') < f_j(x)$
Pareto set, Pareto front

Goal of multi-objective optimization

Find the whole **Pareto Optimal Set**, or a good approximation of the Pareto Optimal Set.
Motivation

- Multi-objective optimization problems are hard:
  underlying single-objective functions, joined functions, etc.
- Learning the problem structure:
  to understand problem difficulty
  to improve algorithm design
  to explain and predict algorithm performance
- Questions:
  What are the relevant problem features?
Joint work

- Arnaud Liefooghe, University of Lille/inria, France
- Fabio Daolio, University of Stirling, UK
- Hernán Aguirre, Shinshu University, Japan
- Kiyoshi Tanaka, Shinshu University, Japan

Slides are from the team, many thanks!
What is a good Pareto set approximation?

**Rule of thumb**
- closeness to the (exact) Pareto front
- well-distributed solutions in the objective space
- the more, the better?

**Quality indicators**
- scalar value that reflects an approximation set quality
- IGD, EPS, R-metrics, HV . . . (all have limitations and biases)
\( \rho \)MNK-landscapes [Kauffman 1993; Aguirre & Tanaka 2007; Verel et al. 2010] general-purpose family of multi-modal pseudo-boolean optimization functions

superposition of \( n \) Walsh functions of order \( k+1 \)

\[
\text{max } f_i(x) = \frac{1}{n} \sum_{j=1}^{n} c_j^i(x_j, x_{j1}, \ldots, x_{jk}), \quad i \in \{1, \ldots, m\} \\
\text{s.t. } x_j \in \{0, 1\}, \quad j \in \{1, \ldots, n\}
\]

Benchmark parameters:

- problem size \( n \) (decision space dimension)
- problem non-linearity \( k < n \) (multi-modality, epistatic interactions)
- number of objective functions \( m \) (objective space dimension)
- objective correlation \( \rho > -\frac{1}{m-1} \)

http://mocobench.sf.net
Some intuitions on objective correlation $\rho$

conflicting objectives
$\rho = -0.9$

independent objectives
$\rho = 0.0$

correlated objectives
$\rho = 0.9$

$m = 2 \quad n = 18 \quad k = 4$
EMO algorithm classes

**Scalarizing approaches**
- multiple aggregations of the objectives (e.g. weighted-sum)
- beware of **unsupported** solutions
- MOSA, MOTS, TPLS, MOEA/D . . .

**Dominance-based approaches**
- search process guided by a dominance relation
- NSGA-II, SPEA2, PAES, PLS, SEMO, AεSεH . . .

**Indicator-based approaches**
- search process guided by a quality indicator
- IBEA, IBMOLS, SMS-EMOA, HypE . . .
Two prototypical dominance-based EMO algorithms

**local search**

multi-objective hill-climber

PLS

[Paquete et al. 2004]

\[ A \leftarrow \{x_0\} \]
\[ \text{repeat} \]
\[ \quad \text{select } x \in A \text{ at random} \]
\[ \quad \text{for all } x' \text{ s.t. } \|x - x'\|_1 = 1 \text{ do} \]
\[ \quad \quad A \leftarrow \text{non-dominated} \]
\[ \quad \quad \text{solutions from } A \cup \{x'\} \]
\[ \quad \text{end for} \]
\[ \text{until stop} \]

**global search**

multi-objective \((1 + 1)\)-EA

G-SEMO

[Laumanns et al. 2004]

\[ A \leftarrow \{x_0\} \]
\[ \text{repeat} \]
\[ \quad \text{select } x \in A \text{ at random} \]
\[ \quad x' \leftarrow x \]
\[ \quad \text{flip each bit } x'_i \text{ with a rate } \frac{1}{n} \]
\[ \quad A \leftarrow \text{non-dominated} \]
\[ \quad \text{solutions from } A \cup \{x'\} \]
\[ \text{until stop} \]
**Benchmark parameters**

Parameters from $\rho$MNK-landscapes

- $n$ problem size  
  (solution space dimension)
- $k$ problem non-linearity  
  (number of epistatic interactions)
- $m$ number of objective functions  
  (objective space dimension)
- $\rho$ objective correlation  
  (correlation between the objective function values)

\[
\begin{align*}
\max \quad & f_i(x) = \frac{1}{n} \sum_{j=1}^{n} c_j^i(x_j, x_{j_1}, \ldots, x_{j_k}) , \quad i \in \{1, \ldots, m\} \\
\text{s.t.} \quad & x_j \in \{0, 1\} , \quad j \in \{1, \ldots, n\}
\end{align*}
\]
Global features from full enumeration (1)

Features from the Pareto set/solution space

- **#po** Pareto optimal (PO) sol.
- **#supp** supported PO solutions
- **hv** PF’s hypervolume
- **#fronts** non-dominated fronts
- **front_ent** entropy of front’s size distribution
Global features from full enumeration (2)
Features from the Pareto set/graph

- $\texttt{podist} \_\texttt{avg}$: avg Hamming distance
- $\texttt{podist} \_\texttt{max}$: max distance (diameter)
- $\texttt{fdc}$: fitness-distance correlation
- $\texttt{#cc}$: connected components
- $\texttt{#sing}$: singletons
- $\texttt{#lcc}$: largest connected comp.
- $\texttt{lcc} \_\texttt{dist}$: avg distance in LCC
- $\texttt{lcc} \_\texttt{hv}$: LCC’s hypervolume
Global features from full enumeration (3)

Local optimality

- #plo: Pareto local optimal (PLO) solutions
- #slo_avg: single-objective local optima (SLO) per objective (avg)
Local features from sampling (1)

Multi-objective random/adaptive walk

random walk sampling (rws)  adaptive walk sampling (aws)
Local features from sampling (2)

Dominance-based metrics

> locally non-dominated solutions in the neighborhood
> supported locally non-dominated solutions in the neighborhood
> neighbors dominated by the current solution
> neighbors dominating the current solution
> neighbors incomparable to the current solution
> average length of aws
Local features from sampling (3)

Hypervolume-based metrics

> (single) solution’s hypervolume
> (single) solution’s hypervolume difference
> neighborhood’s hypervolume
### Summary of problem features (1)

**BENCHMARK parameters (4)**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>(n)</td>
<td>number of (binary) variables</td>
</tr>
<tr>
<td>(k_n)</td>
<td>proportional number of variable interactions (epistatic links) : (k/n)</td>
</tr>
<tr>
<td>(m)</td>
<td>number of objectives</td>
</tr>
<tr>
<td>(\rho)</td>
<td>correlation between the objective values</td>
</tr>
</tbody>
</table>

**GLOBAL FEATURES FROM full enumeration (16)**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>#po</td>
<td>proportion of Pareto optimal (PO) solutions</td>
</tr>
<tr>
<td>#supp</td>
<td>proportion of supported solutions in the Pareto set</td>
</tr>
<tr>
<td>(hv)</td>
<td>hypervolume-value of the (exact) Pareto front</td>
</tr>
<tr>
<td>#plo</td>
<td>proportion of Pareto local optimal (PLO) solutions</td>
</tr>
<tr>
<td>#slo_avg</td>
<td>average proportion of single-objective local optimal solutions per objective</td>
</tr>
<tr>
<td>(podist_avg)</td>
<td>average Hamming distance between Pareto optimal solutions</td>
</tr>
<tr>
<td>(podist_max)</td>
<td>maximal Hamming distance between Pareto optimal solutions (diameter of the Pareto set)</td>
</tr>
<tr>
<td>(po_ent)</td>
<td>entropy of binary variables from Pareto optimal solutions</td>
</tr>
<tr>
<td>(fdc)</td>
<td>fitness-distance correlation in the Pareto set (Hamming dist. in solution space vs. Manhattan dist. in objective space)</td>
</tr>
<tr>
<td>#cc</td>
<td>proportion of connected components in the Pareto graph</td>
</tr>
<tr>
<td>#sing</td>
<td>proportion of isolated Pareto optimal solutions (singletons) in the Pareto graph</td>
</tr>
<tr>
<td>#lcc</td>
<td>proportional size of the largest connected component in the Pareto graph</td>
</tr>
<tr>
<td>(lcc_dist)</td>
<td>average Hamming distance between solutions from the largest connected component</td>
</tr>
<tr>
<td>(lcc_hv)</td>
<td>proportion of hypervolume covered by the largest connected component</td>
</tr>
<tr>
<td>#fronts</td>
<td>proportion of non-dominated fronts</td>
</tr>
<tr>
<td>(front_ent)</td>
<td>entropy of the non-dominated front’s size distribution</td>
</tr>
</tbody>
</table>

**Notes:**
- knowles2003
- aguirre2007
- paquete2007
- liefooghe2013
- paquete2009
- verel2011
- aguirre2007
## Summary of problem features (2)

### Local features from random walk sampling (rws) (17)

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>hv_avg_rws</td>
<td>Average (single) solution’s hypervolume-value</td>
<td></td>
</tr>
<tr>
<td>hv_r1_rws</td>
<td>First autocorrelation coefficient of (single) solution’s hypervolume-values</td>
<td>liefooghe2013</td>
</tr>
<tr>
<td>hud_avg_rws</td>
<td>Average (single) solution’s hypervolume difference-value</td>
<td></td>
</tr>
<tr>
<td>hud_r1_rws</td>
<td>First autocorrelation coefficient of (single) solution’s hypervolume difference-values</td>
<td>liefooghe2013</td>
</tr>
<tr>
<td>nhv_avg_rws</td>
<td>Average neighborhood’s hypervolume-value</td>
<td></td>
</tr>
<tr>
<td>nhv_r1_rws</td>
<td>First autocorrelation coefficient of neighborhood’s hypervolume-value</td>
<td></td>
</tr>
<tr>
<td>#lnd_avg_rws</td>
<td>Average proportion of locally non-dominated solutions in the neighborhood</td>
<td></td>
</tr>
<tr>
<td>#lnd_r1_rws</td>
<td>First autocorrelation coefficient of the proportion of locally non-dominated solutions in the neighborhood</td>
<td></td>
</tr>
<tr>
<td>#lsupp_avg_rws</td>
<td>Average proportion of supported locally non-dominated solutions in the neighborhood</td>
<td></td>
</tr>
<tr>
<td>#lsupp_r1_rws</td>
<td>First autocorrelation coefficient of the proportion of supported locally non-dominated solutions in the neighborhood</td>
<td></td>
</tr>
<tr>
<td>#inf_avg_rws</td>
<td>Average proportion of neighbors dominated by the current solution</td>
<td></td>
</tr>
<tr>
<td>#inf_r1_rws</td>
<td>First autocorrelation coefficient of the proportion of neighbors dominated by the current solution</td>
<td></td>
</tr>
<tr>
<td>#sup_avg_rws</td>
<td>Average proportion of neighbors dominating the current solution</td>
<td></td>
</tr>
<tr>
<td>#sup_r1_rws</td>
<td>First autocorrelation coefficient of the proportion of neighbors dominating the current solution</td>
<td></td>
</tr>
<tr>
<td>#inc_avg_rws</td>
<td>Average proportion of neighbors incomparable to the current solution</td>
<td></td>
</tr>
<tr>
<td>#inc_r1_rws</td>
<td>First autocorrelation coefficient of the proportion of neighbors incomparable to the current solution</td>
<td></td>
</tr>
<tr>
<td>f_cor_rws</td>
<td>Estimated correlation between the objective values</td>
<td></td>
</tr>
</tbody>
</table>

### Local features from adaptive walk sampling (aws) (9)

<table>
<thead>
<tr>
<th>Feature</th>
<th>Description</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>hv_avg_aws</td>
<td>Average (single) solution’s hypervolume-value</td>
<td></td>
</tr>
<tr>
<td>hud_avg_aws</td>
<td>Average (single) solution’s hypervolume difference-value</td>
<td></td>
</tr>
<tr>
<td>nhv_avg_aws</td>
<td>Average neighborhood’s hypervolume-value</td>
<td></td>
</tr>
<tr>
<td>#lnd_avg_aws</td>
<td>Average proportion of locally non-dominated solutions in the neighborhood</td>
<td></td>
</tr>
<tr>
<td>#lsupp_avg_aws</td>
<td>Average proportion of supported locally non-dominated solutions in the neighborhood</td>
<td></td>
</tr>
<tr>
<td>#inf_avg_aws</td>
<td>Average proportion of neighbors dominated by the current solution</td>
<td></td>
</tr>
<tr>
<td>#sup_avg_aws</td>
<td>Average proportion of neighbors dominating the current solution</td>
<td></td>
</tr>
<tr>
<td>#inc_avg_aws</td>
<td>Average proportion of neighbors incomparable to the current solution</td>
<td></td>
</tr>
<tr>
<td>length_aws</td>
<td>Average length of Pareto-based adaptive walks</td>
<td>verel2011</td>
</tr>
</tbody>
</table>
Pairwise feature association (enumerable instances)
Pairwise feature association (large-size instances)

Same association between features from small to large size instances
Experimental setup for large-size instances

Large-size $\rho MNK$-landscapes, constrained random LHS DOE

- problem size $n \in [64, 256]$
- problem non-linearity $k \in [0, 8]$
- number of objectives $m \in [2, 5]$
- objective correlation $\rho \in [-1, 1]$, $\rho > \frac{-1}{m-1}$

1000 problem instances overall

GSEMO and IPLS algorithms

- 30 independent runs per instance
- Fixed budget of 100,000 evaluation calls
- epsilon approximation ratio to best-found non-dominated set
### Prediction accuracy

Cross validation with repeated subsampling, 50 iterations, 90/10 split

<table>
<thead>
<tr>
<th>feature set</th>
<th>MAE</th>
<th>MSE</th>
<th>R²</th>
<th>rank</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>GSEMO</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>all</td>
<td>0.003049</td>
<td>0.000017</td>
<td>0.891227</td>
<td>1</td>
</tr>
<tr>
<td>sampling all</td>
<td>0.003152</td>
<td>0.000018</td>
<td>0.883909</td>
<td>1.3</td>
</tr>
<tr>
<td>sampling rws</td>
<td>0.003220</td>
<td>0.000019</td>
<td>0.878212</td>
<td>2</td>
</tr>
<tr>
<td>sampling aws</td>
<td>0.003525</td>
<td>0.000023</td>
<td>0.854199</td>
<td>3</td>
</tr>
<tr>
<td>$\rho+m+n+k/n$</td>
<td><strong>0.003084</strong></td>
<td><strong>0.000017</strong></td>
<td><strong>0.892947</strong></td>
<td>1</td>
</tr>
<tr>
<td>$\rho+m+n$</td>
<td>0.009062</td>
<td>0.000148</td>
<td>0.065258</td>
<td>4</td>
</tr>
<tr>
<td>$m+n$</td>
<td>0.010813</td>
<td>0.000206</td>
<td>-0.303336</td>
<td>5</td>
</tr>
</tbody>
</table>

| **IPLS**     |       |       |       |      |
| all         | 0.004290 | 0.000034 | 0.886568 | 1    |
| sampling all| 0.004359 | 0.000035 | 0.883323 | 1    |
| sampling rws| 0.004449 | 0.000036 | 0.879936 | 1.3  |
| sampling aws| 0.004663 | 0.000039 | 0.871011 | 2    |
| $\rho+m+n+k/n$ | **0.004353** | **0.000033** | **0.889872** | 1    |
| $\rho+m+n$  | 0.008415 | 0.000119 | 0.600965 | 3    |
| $m+n$       | 0.016959 | 0.000472 | -0.568495 | 4    |
Predicted vs observed values (out-of-folds)
Portfolio accuracy
cross validation with repeated subsampling, 50 iterations, 90/10 split

Portfolio : \{ GSEMO, IPLS \}

<table>
<thead>
<tr>
<th>feature set</th>
<th>error rate</th>
<th>rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>all</td>
<td>0.0128</td>
<td>1</td>
</tr>
<tr>
<td>sampling all</td>
<td>0.0138</td>
<td>1</td>
</tr>
<tr>
<td>sampling rws</td>
<td>0.0150</td>
<td>1</td>
</tr>
<tr>
<td>sampling aws</td>
<td>0.0144</td>
<td>1</td>
</tr>
<tr>
<td>$\rho+m+n+k/n$</td>
<td>0.0134</td>
<td>1</td>
</tr>
<tr>
<td>$\rho+m+n$</td>
<td>0.0824</td>
<td>2</td>
</tr>
<tr>
<td>$m+n$</td>
<td>0.1328</td>
<td>3</td>
</tr>
<tr>
<td>const=GSEMO</td>
<td>0.0880</td>
<td></td>
</tr>
<tr>
<td>const=IPLS</td>
<td>0.7250</td>
<td></td>
</tr>
</tbody>
</table>
Conclusions

Fitness landscape to:

- Understand problem structure from the point of view of local search
- Gain knowledge about problem difficulty
- Explain and predict algorithm performance
- Select algorithm

Local optima network:

- variable aggregation principle + complex network analysis

Multi-objective fitness landscape:

- Relevant features
Perspectives / open issues

- Improving the configuration of algorithms using fitness landscape
- Design cheap features
- Large portfolio of multi-objective algorithms
- Theoretical analysis of LON:
  compute the number of local optima, bassin, etc.
- Fitness landscape for population-based algorithm
- Fitness landscape for continuous and multi-objective optimization
- Extend elementary landscape for another differential equation
- Which "aggregation of variables" shows relevant properties of the optimization problem according to the local search heuristic?

\[
\begin{align*}
X & \xrightarrow{op} X \\
\downarrow^p & \downarrow^p \\
Z & \xrightarrow{op_z} Z
\end{align*}
\]
Khulood Alyahya and Jonathan E Rowe.  
Simple random sampling estimation of the number of local optima.  

A-L. Barabási and R. Albert.  
Emergence of scaling in random networks.  

J. P. K. Doye.  
The network topology of a potential energy landscape: a static scale-free network.  
References II

Fabio Daolio, Marco Tomassini, Sébastien Verel, and Gabriela Ochoa.
Communities of Minima in Local Optima Networks of Combinatorial Spaces.

Josselin Garnier and Leila Kallel.
Efficiency of local search with multiple local optima.

S. A. Kauffman.
*The Origins of Order*.
References III

Katherine M. Malan and Andries P. Engelbrecht.
Fitness landscape analysis for metaheuristic performance prediction.

Peter F. Stadler.
Landscapes and their correlation functions.
P. F. Stadler.
Fitness landscapes.

Frank H Stillinger and Thomas A Weber.

E. D. Weinberger.
Correlated and uncorrelated fitness landscapes and how to tell the difference.