



HAL
open science

Set-based Multiobjective Fitness Landscapes: definition, properties

Sébastien Verel, Clarisse Dhaenens, Laetitia Jourdan, Arnaud Liefooghe

► To cite this version:

Sébastien Verel, Clarisse Dhaenens, Laetitia Jourdan, Arnaud Liefooghe. Set-based Multiobjective Fitness Landscapes: definition, properties. 4th Workshop on Theory of Randomized Search Heuristics, Mar 2010, Paris, France. hal-00460442

HAL Id: hal-00460442

<https://hal.science/hal-00460442v1>

Submitted on 19 Apr 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Set-based Multiobjective Fitness Landscapes: definition, properties.

ThRaSh 2010

Sébastien Verel^{1,2}

Clarisse Dhaenens^{1,3}

Laetitia Jourdan^{1,3}

Arnaud Liefoghe⁴

<http://www.i3s.unice.fr/~verel>

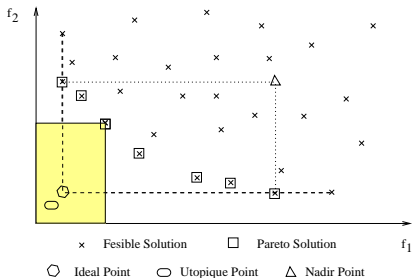
¹DOLPHIN team - INRIA Lille-Nord Europe (France)

²I3S Laboratory - University of Nice-Sophia Antipolis / CNRS (France)

³LIFL - University of Lille 1 / CNRS (France)

⁴University of Coimbra (Portugal)

Multiobjective problem



Multiobjective problem

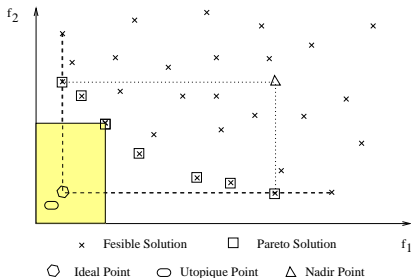
- \mathcal{S} : set of admissible solutions,
- $f : \mathcal{S} \rightarrow \mathbb{R}^p$: multiobjective function to minimize

Dominance (minimization)

x dominate y : $x \prec y$ iff :

- $\forall k = 1..p, f_k(x) \leq f_k(y)$
- $\exists k \in 1..p, f_k(x) < f_k(y)$

Solving a multiobjective problem



Pareto Set (minimization)

$$PS = \{s \in \mathcal{S} \mid \forall z \in \mathcal{S}, z \not\prec s\}$$

Pareto Front

$$PF = \{f(s) \mid s \in PS\}$$

Solving multiobjective problem

Find PS or at least $X \subset PS$
 which is a "good" approximation
 of PS

Difficulty in multiobjective optimization

What does "difficulty" mean ?

- For a given computational time, X far from a "good" approximation of the Pareto Set (distance in objective space to PS, lack of diversity of X , etc.)
- Time, or number of evaluations, for solving the problem is long : class of complexity

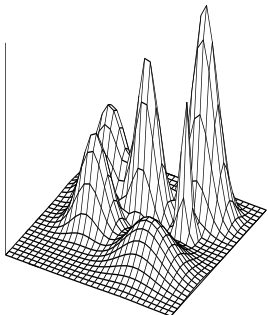
Difficulty in multiobjective optimization

What's make "difficult" a multiobjective problem ?

- The number of objective (criteria)
- The size of the feasible solutions space
- The "difficulty" of the single objective problem (time complexity, ...)
- The correlation between the objectives
- The convexity of the Pareto Front
- The connectedness of the Pareto Set
- ...

Fitness landscape point of view can help...

Fitness landscapes in single-objective optimization



Fitness landscape $(\mathcal{S}, \mathcal{N}, f)$

- \mathcal{S} : set of admissible solutions,
- $\mathcal{N} : \mathcal{S} \rightarrow 2^{\mathcal{S}}$: neighborhood function,
- $f : \mathcal{S} \rightarrow \mathbb{R}$: fitness function.

$$\forall x \in \mathcal{S},$$

$$\mathcal{N}(x) = \{y \in \mathcal{S} \mid P(y = op(x)) > 0\}$$

or

$$\mathcal{N}(x) = \{y \in \mathcal{S} \mid d(y, x) \leq k\}$$

Goal of the fitness landscapes study

- "Geometry" (features) of fitness landscape
⇒ dynamics of a local search algorithm
- Geometry is linked to the problem difficulty :
 - If there are a lot of local optima, the probability to find the global optimum is lower.
 - If the fitness landscape is flat, discovering better solutions is rare.
 - What is the best search direction in the landscape?

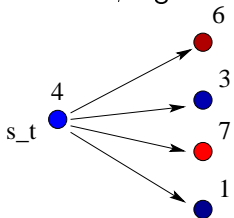
Study of the fitness landscape features
allows to study
the performance of search algorithms

Goal of the fitness landscapes study

- 1 To compare the difficulty of two search spaces :
 - One problem with 2 (or more) possible codings : $(S_1, \mathcal{N}_1, f_1)$ and $(S_2, \mathcal{N}_2, f_2)$
different coding, mutation operator, fitness function, etc.
Which one is easier to solve?
- 2 To choose the algorithm :
 - analysis of global geometry of the landscape
Which algorithm can I use?
- 3 To tune the parameters :
 - *off-line* analysis of structure of fitness landscape
Which is the best mutation operator ? the size of the population ? etc.
- 4 To control the parameters during the run :
 - *on-line* analysis of structure of fitness landscape
Which is the optimal mutation operator according to the estimation of structure ?

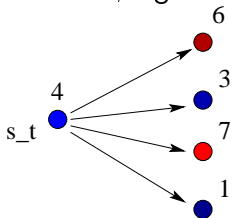
Links with local search

Solutions, eighbors, fitness :

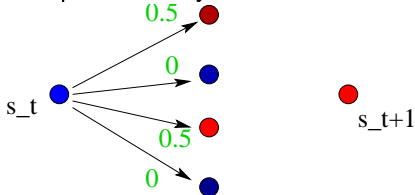


Links with local search

Solutions, eighbors, fitness :



Put prob. from your heuristic :



- Fitness landscape point of view :
before putting a particular heuristic
- Sample the neighborhood to have information on local features of the search space
- From local information :
deduce some global features like general shape of search space, "difficulty", etc.

Multiobjective fitness landscapes (moFiL)

3 ideas to define multiobjective Fitness Landscapes (moFiL)

- Change **nothing** from the mono-objective definition (the most studied from 2003) :
Fitness function is the multiobjective function
- Change the **fitness function** by an indicator such as hypervolume :
Ruggedness can be studied
- Change the **search space** by the set of sets of solutions :
"Real" search space viewed by efficient local searches

Fitness function is the multiobjective function

Multidimensional multiobjective Fitness Landscape $(\mathcal{S}, \mathcal{N}, f)$

- \mathcal{S} : set of admissible solutions,
- $\mathcal{N} : \mathcal{S} \rightarrow 2^{\mathcal{S}}$: neighborhood defined by local operator or distance on \mathcal{S} ,
- $f : \mathcal{S} \rightarrow \mathbb{R}^p$: multiobjective function.

Main idea to analysis of multiobjective fitness landscapes :

- Dominance relation

Overview on previous tools (1)

- **Pareto Set** [AT04] [AT07] :

Number, distribution over landscape

+ :Information on the final goal

– :Information only on the final goal

- Partition the search space into **dominance levels** [PCS04]
[GD09] :

connection between level of dominance

+ :Information on structure of search space

– :Ways to reach PS? small search space

Overview on previous tools (2)

- **Connectedness** [EK97] [GKR06] [PS09] :

number of clusters, size of cluster by the graph induced by Pareto set and neighborhood relation (edges if the distance $\leq k$)

+ :If connected, easy way to find the PS

– :Huge PS are intractable, final goal, often not connected

- **Fitness distance correlation, random walk from Pareto loc. opt.** [GD06a] [GD08] [GD06b] :

correlation between distances in obj. and sol. space.

+ :Show the correlation of fitness

*– :The reference solution change, no correlation after few steps,
no scale*

Remarks on multidimensional moFiL

- Majority of works use the concept of dominance :
connectedness, size, estimation of distribution Pareto set,
partition by dominance levels

Limited results :

- No enough information for the design of metaheuristics
- Few understanding of the complexity of multiobjective problems from the fitness landscapes analysis

Change the fitness function by an indicator

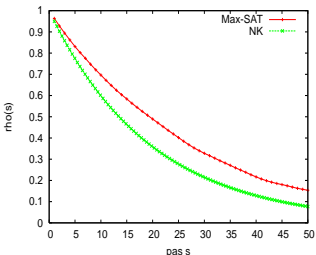
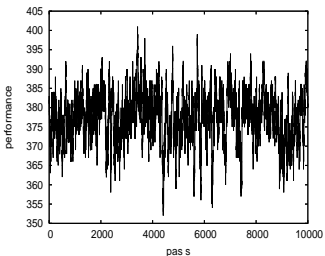
Indicator multiobjective Fitness Landscape $(\mathcal{S}, \mathcal{N}, i)$

- \mathcal{S} : set of admissible solutions,
- $\mathcal{N} : \mathcal{S} \rightarrow 2^{\mathcal{S}}$: neighborhood defined by local operator or distance on \mathcal{S} ,
- $i : 2^{\mathcal{S}} \rightarrow \mathbb{R}$: indicator (such as hypervolume, ϵ dominance) on set of solutions

Main idea to analysis of multiobjective fitness landscapes :

- Value of the indicator (hypervolume) of the neighborhood :
$$i(\mathcal{N}(s) \cup \{s\}) = hv(\mathcal{N}(s) \cup \{s\})$$

Rugged/smooth fitness landscapes



Autocorrelation of time series of fitnesses $(f(s_1), f(s_2), \dots)$ along a random walk (s_1, s_2, \dots) [Wei90] :

$$\rho(n) = \frac{E[(f(s_i) - \bar{f})(f(s_{i+n}) - \bar{f})]}{\text{var}(f(s_i))}$$

autocorrelation length $\tau = \frac{1}{\rho(1)}$

- small τ : rugged landscape
- long τ : smooth landscape

A model of multiobjective problem

Multiobjective *MNK*-landscapes with objective correlation

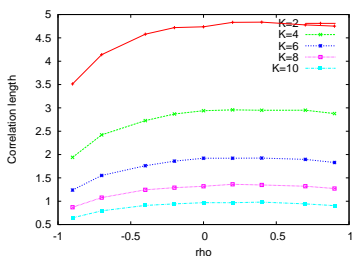
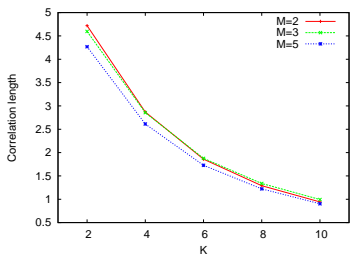
MNK landscapes [AT04]

$$\forall m \in [1, M], f_{NK,m}(s) = \frac{1}{N} \sum_{i=1}^N f_{i,m}(s_i, s_{i_1}, \dots, s_{i_K})$$

- M objectives which are NK – landscapes
- Same interaction : same K , and same s_i
- Each fitness component $f_{i,m}$ by extension : $y_{s_i, s_{i_1}, \dots, s_{i_K}}^{i,m}$.

- MNK : y_{\dots}^{im} are randomly and independently drawn from $[0, 1)$.
- in ρMNK : $(y_{s_i, s_{i_1}, \dots, s_{i_K}}^{i1}, \dots, y_{s_i, s_{i_1}, \dots, s_{i_K}}^{iM}) \sim$ multivariate uniform law of dimension M :
objective correlation between objective can be tuned precisely

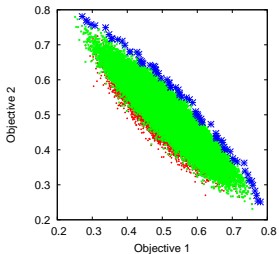
Results on *MNK*-landscapes with objective correlation



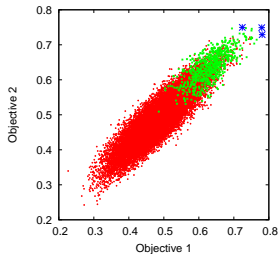
- Dimension has low influence on correlation length
- Objective correlation change slowly the correlation length
- The non-linearity (epistasis) has the main influence

Analysis

- Definition of the ruggedness in indicator moFiL.
- Ruggedness is not correlated to the number of Pareto Local Optima in moFiL.
- Argument in favour to optimize the hypervolume rather than the dominance



$$K = 4, \rho = -0.9$$



$$K = 4, \rho = 0.9$$

Remarks on indicator moFiL

Pros :

- Estimation of non-linearity / ruggedness of the landscapes
- Explain "local" difficulty of multiobjective problem

Cons :

- Metaheuristics manipulate set of solutions, not single solution

Change the search space by the set of sets of solutions

Set based multiobjective Fitness Landscape (Σ_q, N_q, i)

- Σ_q : set of sets of feasible solutions of maximum size q

$$\Sigma_q = \{\sigma \subset \mathcal{S} : |\sigma| \leq q\}, \text{ so } |\Sigma_q| = \binom{|\mathcal{S}|}{q}$$

- N_q : insert, delete, or change one solution in the set
- $i : \Sigma_q \rightarrow \mathbb{R}$: indicator (such as hypervolume, ϵ dominance) on set of solutions

Main idea to analysis of multiobjective fitness landscapes :

- All the standard tools of Fitness Landscapes analysis

Neighborhood N_q : Insert, delete, or change 1 solution

- Neighborhood $N_q : \Sigma_q \rightarrow 2^{\Sigma_q}$:
defined from neighborhood $\mathcal{N} : \mathcal{S} \rightarrow 2^{\mathcal{S}}$
- Neighbors of $\sigma \in \Sigma_q$ are given by 3 operators :
 - Insert one solution which is neighbor of one of solution of σ
 - Delete one solution
 - Change one solution by its neighbor

$$N_q(\sigma) = N_q^i(\sigma) \cup N_q^d(\sigma) \cup N_q^c(\sigma)$$

$N_q^i(\sigma)$: Insert one solution in the set

σ' neighbor of $\sigma = \{s_1, \dots, s_k\}$

- if $k = q$: $\sigma' = \sigma$
- if $0 < k < q$: $\sigma' = \{s_1, \dots, s_k, s_{k+1}\}$ with $s_{k+1} \in \mathcal{N}(s_i)$
- if $k = 0$: $\sigma' = \{s\}$ with $s \in \mathcal{S}$

Sizes

when $0 < k < q$,

- if s_{k+1} is not in σ , $|\sigma'| = |\sigma| + 1$
- $|N_q^i(\sigma)| \leq \sum_{i=1}^k |\mathcal{N}(s_i)|$
 $|N_q^i(\sigma)| \leq k\lambda$ if neighborhood size is cst.

$N_q^d(\sigma)$: Delete one solution from the set

σ' neighbor of $\sigma = \{s_1, \dots, s_k\}$

- if $k = 0$: $\sigma' = \sigma$
- if $0 < k$: $\sigma' = \{s_1, \dots, s_{i-1}, s_{i+1}, \dots, s_k\}$ with $i \in 1..k$

Sizes

when $0 < k$,

- $|\sigma'| = |\sigma| - 1$
- $|N_q^d(\sigma)| = k$

$N_q^c(\sigma)$: Change one solution of the set σ' neighbor of $\sigma = \{s_1, \dots, s_k\}$

- if $k = 0$: $\sigma' = \sigma$
- if $0 < k$: $\sigma' = \{s_1, \dots, s_{i-1}, s'_i, s_{i+1}, \dots, s_k\}$
with $i \in 1..k$ and $s'_i \in \mathcal{N}(s_i)$

Sizes

when $0 < k$,

- if s'_i is not in σ , $|\sigma'| = |\sigma|$
- $|N_q^c(\sigma)| \leq \sum_{i=1}^k |\mathcal{N}(s_i)|$
 $|N_q^i(\sigma)| \leq k\lambda$ if neighborhood size is cst.
- Total size of $N_q(\sigma)$: $|N_q(\sigma)| \leq k + 2 \sum_{i=1}^k |\mathcal{N}(s_i)|$
 $|N_q(\sigma)| \leq k(2\lambda + 1)$ if neighborhood size is cst.

Links with local search : example of efficient LS

"On set based-based multiobjective optimization" Zitzler et al, 2009

Set based Algo. for MO

Generate initial set P of size q with random solutions

repeat

$P' \leftarrow \text{randomSetMutation}(P)$

$P'' \leftarrow \text{heuristicSetMutation}(P)$

if $P'' \preceq P$ then

$P \leftarrow P''$

else

if $P' \preceq P$ then

$P \leftarrow P'$

end if

end if

until termination criterion is not true

RandomSetMutation

randomly choose $s_1, \dots, s_k \in S$ with $s_i \neq s_j$

randomly select $p_1, \dots, p_k \in P$ with $p_i \neq p_j$

$P' \leftarrow P \setminus \{s_1, \dots, s_k\} \cup \{p_1, \dots, p_k\}$

return P'

heuristicSetMutation

Generate $s_1, \dots, s_k \in S$ based on P (selection and mutation of solutions from P)

$P' \leftarrow P \cup \{s_1, \dots, s_k\}$

while $|P''| > q$ do

for all $a \in P''$ do

$\delta_a \leftarrow i(P'' \setminus \{a\}) - i(P'')$

end for

choose $p \in P''$ with $\delta_p = \min_{a \in P''} \delta_a$

$P'' \leftarrow P'' \setminus \{p\}$

end while

return P''

Tools to analysis set-based multiobjective fitness landscapes

- Density of states :
What can we expect from random search ?
- Adaptive walk, local optima :
What is the number of local optima, and size of basin ?
- Random walk, ruggedness :
Is the solution in the neighborhood are random ?
- Fitness cloud, evolvability :
What is the distribution of fitness in the neighborhood ?
Can the solutions go to better fitness values ?
- Neutral walk, neutrality :
Is there a lot of solutions with the same fitness ?
Is there information in and around the neutral networks ?
- Local Optima Network :
What are the links between basins of attraction in the landscapes ?

Conclusion and perspectives

Summary

- Previous works on multidimensional moFiL : Fitness is the multiobjective function based on dominance, information on Pareto Set (goal of optimization)
- Indicator moFiL : Fitness is the value of indicator based on random walk, information on ruggedness
- Set based moFiL : search space is the set of sets
Use the "real" search space and elementary operators on sets

Conclusion and perspectives

Summary

- Previous works on multidimensional moFiL : Fitness is the multiobjective function based on dominance, information on Pareto Set (goal of optimization)
- Indicator moFiL : Fitness is the value of indicator based on random walk, information on ruggedness
- Set based moFiL : search space is the set of sets
Use the "real" search space and elementary operators on sets

Perspectives

- Use it !
- To decide the best encoding and operator for example in multiobjective TVRP (see M-E Marmion)

Non commercial communication

Special Session

Fitness landscapes and metaheuristics

- at international conference META'2010, Djerba Island, Tunisia, October 28-30th, 2010.
- Extended abstract 2 pages
- Post publication for selected papers in COR, ITOR, JMMA
- Deadline may 5, 2010
- web page :
<http://www2.lifl.fr/META10/index.php?n=Main.InfoFiL>

For Further Reading



Hernán E. Aguirre and Kiyoshi Tanaka.

Insights on Properties of Multiobjective MNK-Landscapes.

In *2004 Congress on Evolutionary Computation (CEC'2004)*, volume 1, pages 196–203, Portland, Oregon, USA, June 2004. IEEE Service Center.



Hernán E. Aguirre and Kiyoshi Tanaka.

Working principles, behavior, and performance of moeas on mnk-landscapes.

European Journal of Operational Research, 181(3) :1670 – 1690, 2007.



M. Ehrgott and K. Klamroth.

Connectedness of efficient solutions in multiple criteria combinatorial optimization.

European Journal of Operational Research, 97 :159–66, 1997.

For Further Reading



Deon Garrett and Dipankar Dasgupta.

Analyzing the performance of hybrid evolutionary algorithms for the multiobjective quadratic assignment problem.

In *Congress on Evolutionary Computation (CEC 2006)*, pages 1710–1717. IEEE Service Center, July 2006.



Deon Garrett and Dipankar Dasgupta.

Analyzing the performance of hybrid evolutionary algorithms for the multiobjective quadratic assignment problem.

In *Congress on Evolutionary Computation (CEC'2006)*, pages –. IEEE Service Center, July 2006.






Deon Garrett and Dipankar Dasgupta.

Multiobjective landscape analysis and the generalized assignment problem.

pages 110–124, 2008.

For Further Reading

-  Deon Garrett and Dipankar Dasgupta.
Plateau connection structure and multiobjective metaheuristic performance.
In *Congress on Evolutionary Computation (CEC'2009)*, pages –. IEEE Service Center, July 2009.
-  Jochen Gorski, Kathrin Klamroth, and Stefan Ruzika.
Connectedness of Efficient Solutions in Multiple Objective Combinatorial Optimization.
Technical report, Kaiserslauterer uniweiter, 2006.
-  Luis Paquete, Marco Chiarandini, and Thomas Stützle.
Pareto Local Optimum Sets in the Biobjective Traveling Salesman Problem : An Experimental Study.
In Xavier Gandibleux, Marc Sevaux, Kenneth Sörensen, and Vincent T'kindt, editors, *Metaheuristics for Multiobjective Optimisation*, pages 177–199, Berlin, 2004. Springer. Lecture

For Further Reading



Luís Paquete and Thomas Stützle.

Clusters of Non-dominated Solutions in Multiobjective Combinatorial Optimization : An Experimental Analysis.

In Vincent Barichard, Matthias Ehrgott, Xavier Gandibleux, and Vincent T'Kindt, editors, *Multiobjective Programming and Goal Programming. Theoretical Results and Practical Applications*, pages 69–77. Springer, Lecture Notes in Economics and Mathematical Systems, Vol. 618, 2009. ISBN 978-3-540-85645-0.



E. D. Weinberger.

Correlated and uncorrelated fitness landscapes and how to tell the difference.

In *Biological Cybernetics*, pages 63 :325–336, 1990.