Exploratory Landscape Analysis
– Advanced Tutorial at GECCO 2017 –

Pascal Kerschke

Information Systems & Statistics,
University of Münster, Germany

Mike Preuss

Berlin, July 15 - 19, 2017

GECCO '17 Companion, July 15-19, 2017, Berlin, Germany

©2017 Copyright is held by the owner/author(s).

ACM ISBN 978-1-4503-4939-0/17/07.
http://dx.doi.org/10.1145/3067695.3067696

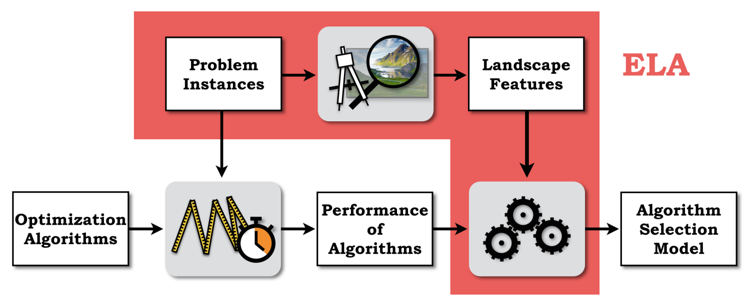
Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.

Agenda

General Idea of Exploratory Landscape Analysis

General Idea of ELA

- algorithm selection problem[1]
  - ⇝ find the individually best suited algorithm for an unseen optimization problem



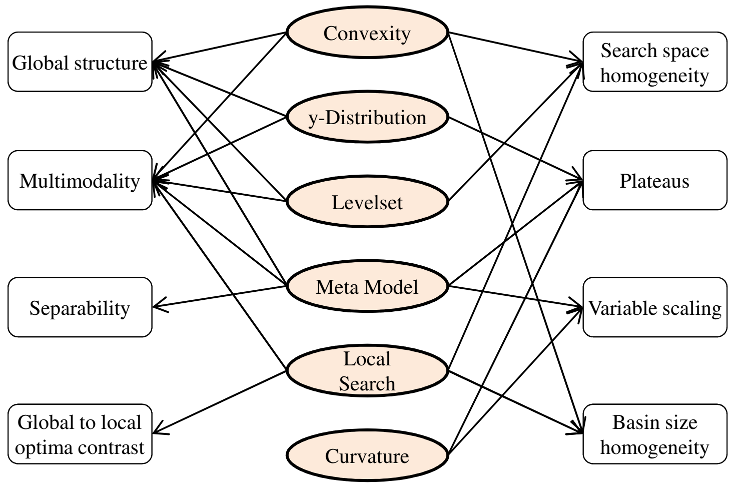
[1] Rice, J. (1976). *The Algorithm Selection Problem.* In Advances in Computers (pp. 65-118).

General Idea of ELA

Exploratory Landscape Analysis (ELA):

- we aim at finding the right algorithm
- but also at improving problem or algorithm/problem dependency understanding
- basic idea (exploratory!): we start with **very simple features without clear purpose**
- match existing high-level features (expert knowledge) with our ELA features
- currently: mostly continuous (black-box) (global) optimization, but also in other domains (e.g. TSP)

General Idea of ELA



Mersmann, O., Preuss, M. & Trautmann, H. (2010). *Benchmarking Evolutionary Algorithms: Towards Exploratory Landscape Analysis.* In Proceedings of PPSN XI (pp. 71 - 80).

General Idea of ELA

- we do not know functional relationships when designing features
- but we can match them to high-level characteristics (multimodality, funnel structure, etc.) of optimization problems
- this enables recognizing important problem properties quickly
- based on initial design of samples $x_{i1}, \dots, x_{iD}$ and their corresponding fitness value y_i, $i = 1, \dots, n$
- given an evaluated initial design (initial population?), most ELA features are for free
- there are already several different feature sets

General Idea of ELA

What is the difference to Fitness Landscape Analysis (FLA)?

- basic idea is similar: extract knowledge on problems in order to select proper algorithm
- however, our viewpoint is always set-based: no single feature has to explain anything on its own, the combination is important
- we heavily rely on Machine Learning for composing good feature sets
- additionally, we strive for (very) small sample sizes, in the range of initial generation samples (100D points and less)

Block 1 - Part I

A Brief Introduction into the FLACCO-GUI

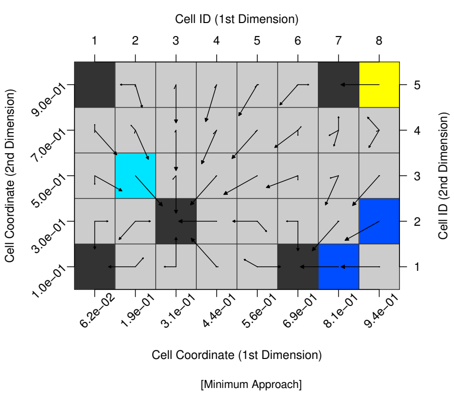
Block 1 - Part II

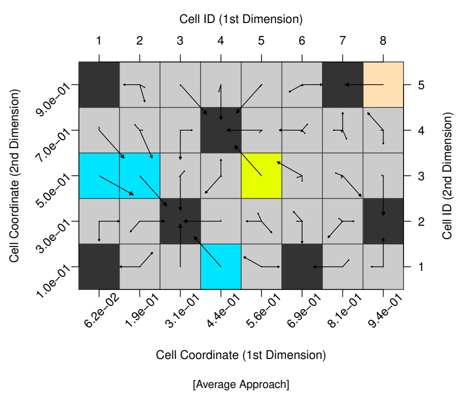
ELA for Single-Objective Global Optimization Problems

Block 1 - Part II: Single-Objective Global Opt.

General Cell Mapping Features

Kerschke, P., Preuss, M., Hernández, C., Schütze, O., Sun, J.-Q., Grimme, C., Rudolph, G., Bischl, B. & Trautmann, H. (2014). *Cell Mapping Techniques for Exploratory Landscape Analysis*. In Proceedings of EVOLVE 2014 (pp. 115 – 131).

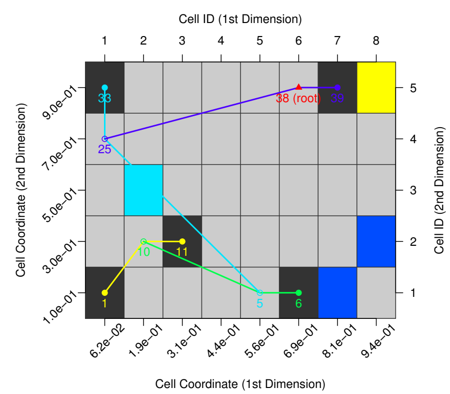


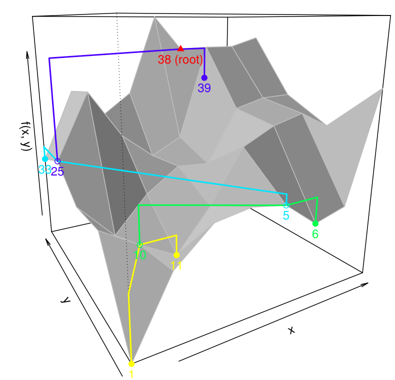


Block 1 - Part II: Single-Objective Global Opt.

Barrier Tree Features

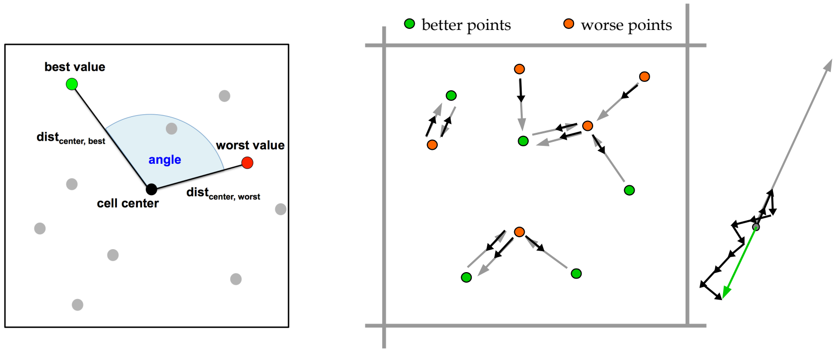
Hernández, C., Schütze, O., Emmerich, M. T. M., & Xiong, F. R. (2014). *Barrier Tree for Continuous Landscapes by Means of Generalized Cell Mapping*. In Proceedings of EVOLVE 2014.





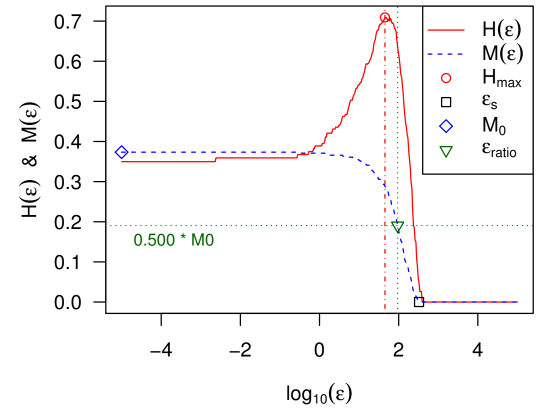
Cell Mapping Features

Kerschke, P., Preuss, M., Hernández, C., Schütze, O., Sun, J.-Q., Grimme, C., Rudolph, G., Bischl, B. & Trautmann, H. (2014). *Cell Mapping Techniques for Exploratory Landscape Analysis*. In Proceedings of EVOLVE 2014 (pp. 115 - 131).



Information Content Features

Muñoz, M. A., Kirley, M., Halgamuge, S. K. (2015). *Exploratory Landscape Analysis of Continuous Space Optimization Problems using Information Content*. In IEEE Transactions on Evolutionary Computation (pp. 74 – 87).



Dispersion Features

Lunacek, M. & Whitley, D. (2006). *The Dispersion Metric and the CMA Evolution Strategy*. In Proceedings of GECCO 2006 (pp. 477 – 484).

Length Scale Features

Morgan, R. & Gallagher M. (2015). *Analyzing and Characterising Optimization Problems Using Length Scale*. In Soft Computing (pp. 1 – 18).

Ruggedness Features

Malan, K. M. & Engelbrecht, A. P. (2013). *Ruggedness, Funnels and Gradients in Fitness Landscapes and the Effect on PSO Performance*. In Proceedings of CEC 2013 (pp. 963 – 970).

Fitness Distance Correlation Features

Jones, T. & Forrest, S. (1995). *Fitness Distance Correlation as a Measure of Problem Difficulty for Genetic Algorithms*. In Proceedings of ICGA 1995 (pp. 184 – 192).

Violation Landscape Features

Malan, K. M., Oberholzer, J. F. & Engelbrecht, A. P. (1995). *Characterising Constrained Continuous Optimisation Problems*. In CEC 2015 (pp. 1351 – 1358).

Hill Climbing Features

Abell, T., Malitsky, Y. & Tierney, K. (2013). *Features for Exploiting Black-Box Optimization Problem Structure*. In Proceedings of LION 2013 (pp. 30 – 36).

Block 1 - Part II: Single-Objective Global Opt.

Nearest Better Clustering Features

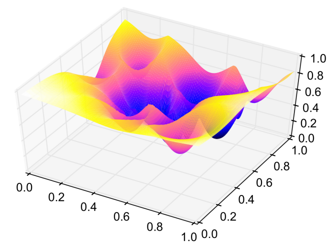
Kerschke, P., Preuss, M., Wessing, S. & Trautmann H. (2015). *Detecting Funnel Structures by Means of Exploratory Landscape Analysis*. In Proceedings of GECCO 2015 (pp. 265 – 272).

Meta-Model & NBC Features ⇝ Funnel Detection

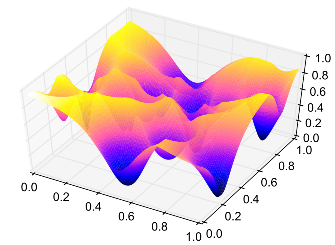
Block 1 - Part II: Single-Objective Global Opt.

Funnel Detection

- funnel: local optima are located near to each other and pile up to an "upside-down mountain"
- knowledge about underlying global structure, i.e., funnels, helps selecting the right algorithm



(a) funnel



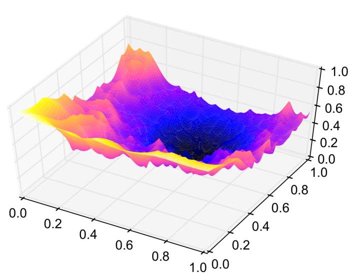
(b) non-funnel ("random")

Kerschke, P., Preuss, M., Wessing, S. & Trautmann H. (2016). Low-Budget Exploratory Landscape Analysis on Multiple Peaks Models. In Proceedings of GECCO 2016 (pp. 229-236)

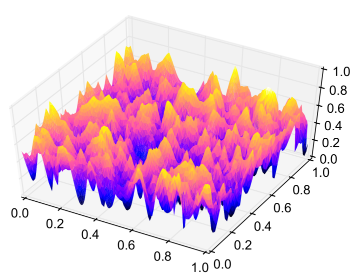
Block 1 - Part II: Single-Objective Global Opt.

Funnel Detection

- different algorithm candidates for either category
- wide variety within the classes "funnel" and "non-funnel"



(a) funnel



(b) non-funnel ("random")

Kerschke, P., Preuss, M., Wessing, S. & Trautmann H. (2016). Low-Budget Exploratory Landscape Analysis on Multiple Peaks Models. In Proceedings of GECCO 2016 (pp. 229-236)

Block 1 - Part II: Single-Objective Global Opt.

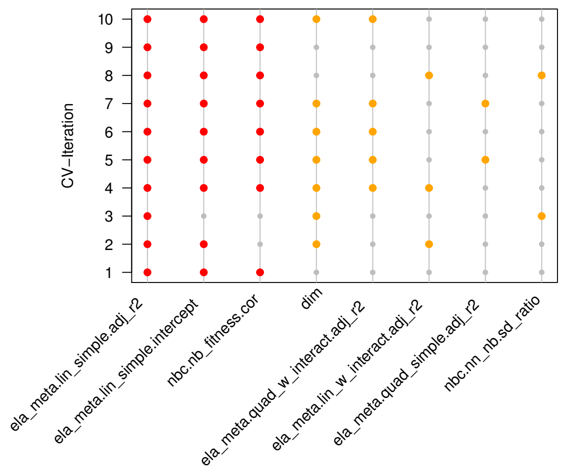
Funnel Detection

- detailed results in our GECCO 2016 paper
- used MPM2[2] to generate a set of 4,000 training instances
- initial designs of size $50 \times D$ observations (small!)
- trained four classifiers (random forest, rpart, kknn and ksvm)
- only used a total of 8 Meta-Model and NBC features
- validated results on BBOB and subset of problems from CEC-2013 niching competition

[2] multiple peaks model 2 generator, available in python (`optproblems0.9`, Wessing, S.) and R (`smoof`, Bossek, J.)

Block 1 - Part II: Single-Objective Global Opt.

Funnel Detection



Kerschke, P., Preuss, M., Wessing, S. & Trautmann H. (2016). Low-Budget Exploratory Landscape Analysis on Multiple Peaks Models. In Proceedings of GECCO 2016 (pp. 229-236)

Block 1 - Part III

Introduction into FLACCO and its GUI

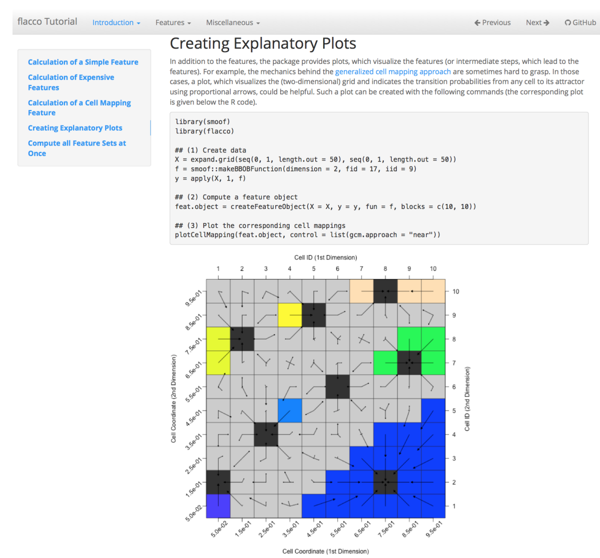
Block 1 - Part III: `FLACCO` and its GUI

- `flacco`: **F**eature-Based **L**andscape **A**nalysis of **C**ontinuous and **C**onstraint **O**ptimization Problems
- unified interface for multiple (single-objective) sets of configurable features
- stable release on CRAN[3] / developers version on GitHub[4]
- multiple vizualisation techniques (partially shown on these slides)
- tracks # of function evaluations <u>and</u> run time - per feature set

[3] Stable Release: `https://cran.r-project.org/package=flacco`

[4] Developers Version: `https://github.com/kerschke/flacco`

Block 1 - Part III: `FLACCO` and its GUI



Tutorial: `http://kerschke.github.io/flacco-tutorial/site/`

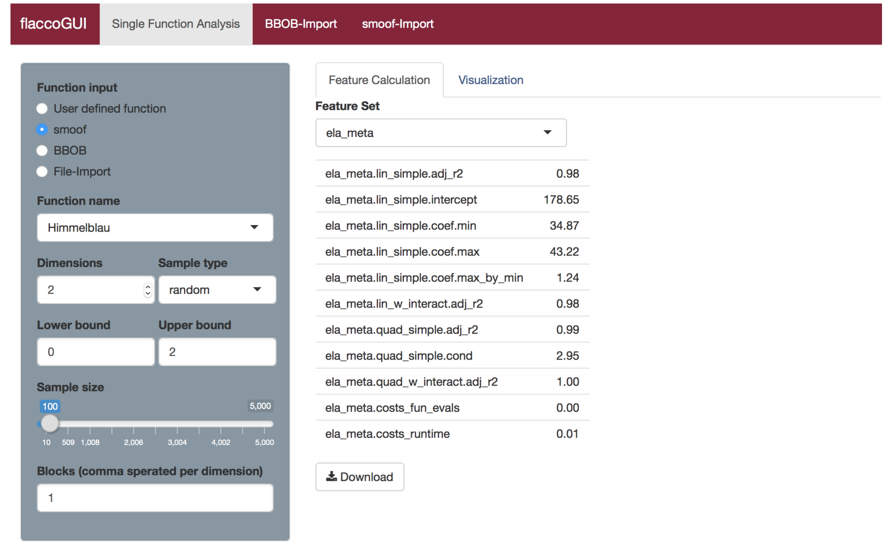
Block 1 - Part III: FLACCO and its GUI

- flacco also comes with a GUI, which provides many functionalities of the package itself
- the GUI can be started (within R) using the commands below:

```
> # first, install "flacco" from CRAN
> install.packages("flacco", dependencies = TRUE)
>
> # then, load the package and start the app
> library(flacco)
> runFlaccoGUI()
```

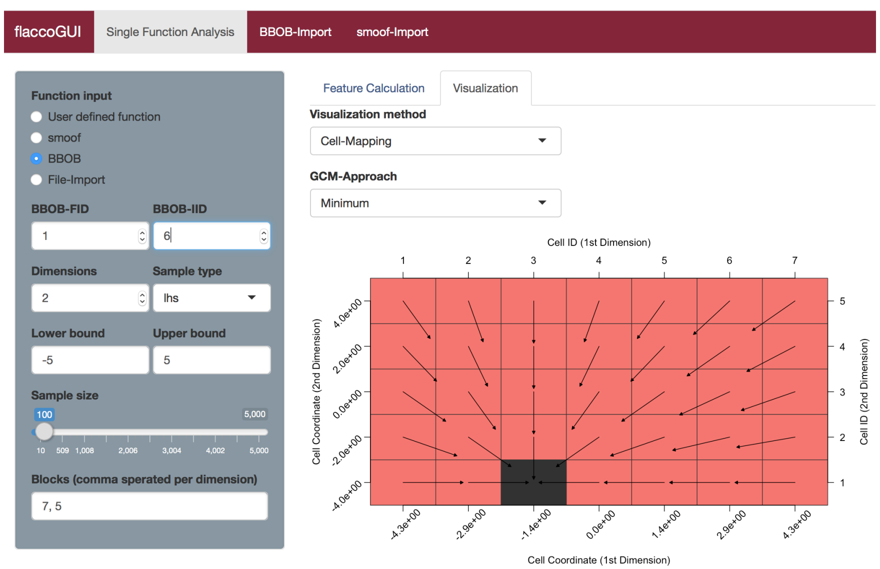
- alternatively, one can completely avoid the usage of R and use the online version of the GUI, which can be found here: https://flaccogui.shinyapps.io/flaccogui

Block 1 - Part III: FLACCO and its GUI



https://flaccogui.shinyapps.io/flaccogui/

Block 1 - Part III: FLACCO and its GUI



https://flaccogui.shinyapps.io/flaccogui/

Block 1 - Part III: FLACCO and its GUI

Further information on flacco and/or its GUI can be found here:

- GECCO 2017 workshop paper:
  Hanster, C. & Kerschke, P. (2017). *flaccogui: Exploratory Landscape Analysis for Everyone*. In Proceedings of GECCO 2017.
- presentation at the EvoSoft-Workshop at GECCO 2017:
  http://dev.heuristiclab.com/trac.fcgi/wiki/EvoSoft
- CEC 2016 paper:
  Kerschke, P. & Trautmann, H. (2016). *The R-Package FLACCO for Exploratory Landscape Analysis with Applications to Multi-Objective Optimization Problems*. In Proceedings of CEC 2016.

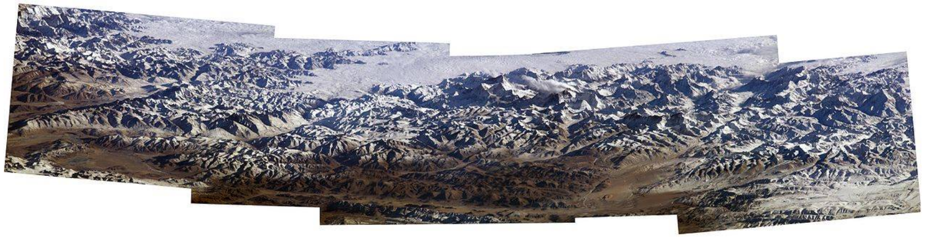
Block 1 - Part IV

Live-Session Using FLACCO and its GUI

Block 2

ELA for Single-Objective Multimodal Optimization Problems

Block 2: Single-Objective Multimodal Optimization



Block 2: Single-Objective Multimodal Optimization

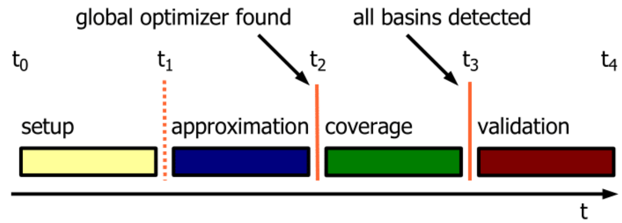
- core difference: we are looking for solution sets, not for one optimal solution
- sample definition:
  "In a multimodal optimization task, the main purpose is to find multiple optimal solutions (global and local), so that the user can have a better knowledge about different optimal solutions in the search space and as and when needed, the current solution may be switched to another suitable optimum solution."
  (from Deb, Saha: Multimodal Optimization Using a Bi-Objective Evolutionary Algorithm, ECJ, 2012)
- many things are fuzzy here

Block 2: Single-Objective Multimodal Optimization

- different aims possible
- currently most important (competitions): multiglobal
  = find all search space points that are globally optimal
- two main algorithmic approaches:
  - parallel, large populations
  - sequential, coordinated restarts
- several components that may be used: archives, clustering methods, methods for obtaining well distributed samples
- ELA could be helpful for selecting components/methods

Block 2: Single-Objective Multimodal Optimization

- funnel detection is important, because many methods need to partition space into basins
- for two-stage methods, we know that **restart organization** does not make much sense for global optimization
- but it does for multimodal optimization (because we have to look "everywhere", Preuss: Multimodal Optimization by Means of Evolutionary Algorithms, Springer 2015)



Block 2: Single-Objective Multimodal Optimization

- how do t_2 and t_3 depend on organizing restarts well?

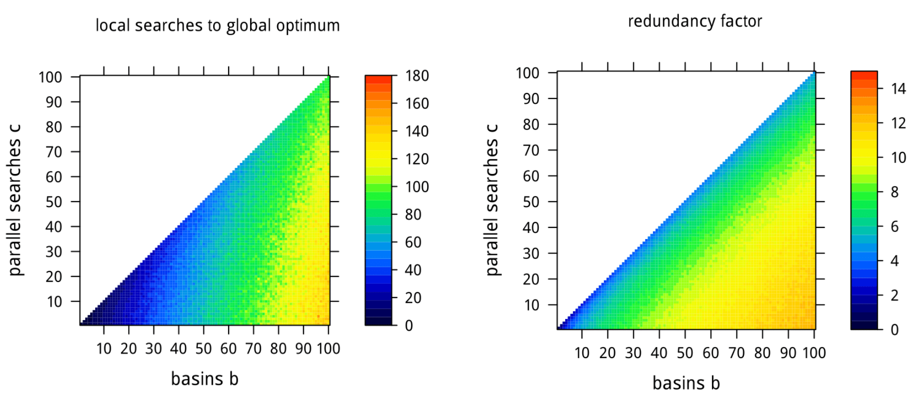


Figure: left: t_2 (time to global optimum), right: t_3 (time until all basins have been visited) for unequal basin sizes (1:10) and moderately well working ($p = 0.5$) basin identification

Block 2: Single-Objective Multimodal Optimization

So what do we need ELA to do for us?

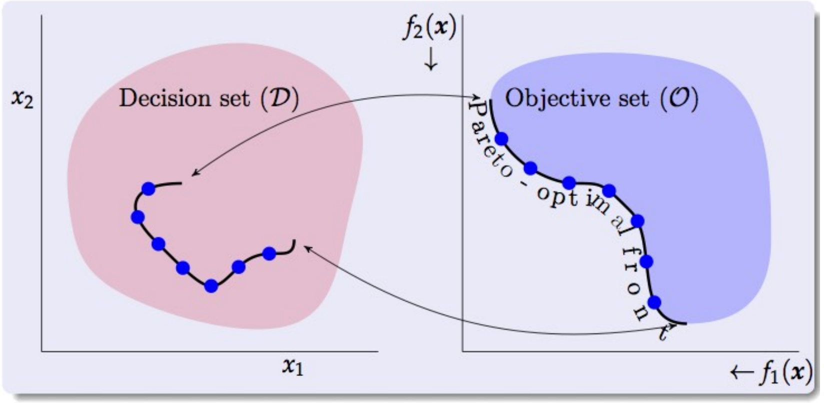
- having an idea of how many basins are there would be great
- knowing how homogenous basin sizes are is important
- but even if we do not understand everything, it would be good to properly match algorithms/components to problems
- there is little activity in this direction: get active!

Block 3 - Part I

ELA for Multi-Objective "Global" Optimization Problems

Block 3 - Part I: Multi-Objective "Global" Opt.

$$\text{minimize } \boldsymbol{F}(\boldsymbol{x}) = \langle f_1(\boldsymbol{x}), f_2(\boldsymbol{x})\rangle ,\quad \text{with } \boldsymbol{x} \in \mathcal{D} \subseteq \mathbb{R}^2 .$$



source: lmarti.github.io

Block 3 - Part I: Multi-Objective "Global" Opt.

- in single-objective optimization, ELA has shown to be useful for describing the problem landscape based on a small initial design
- currently, there exist almost no landscape features for continuous multi-objective optimization problems
- first approaches[5,6] towards ELA in the multi-objective setting

[5] Kerschke, P. & Trautmann, H. (2016). *The R-Package FLACCO for Exploratory Landscape Analysis with Applications to Multi-Objective Optimization Problems*. In Proceedings of CEC 2016.

[6] Kerschke, P., Wang, H., Preuss, M., Grimme, C., Deutz, A., Trautmann, H., & Emmerich, M. (2016). *Towards Analyzing Multimodality of Multiobjective Landscapes*. In Proceedings of PPSN 2016 (pp. 962-972)

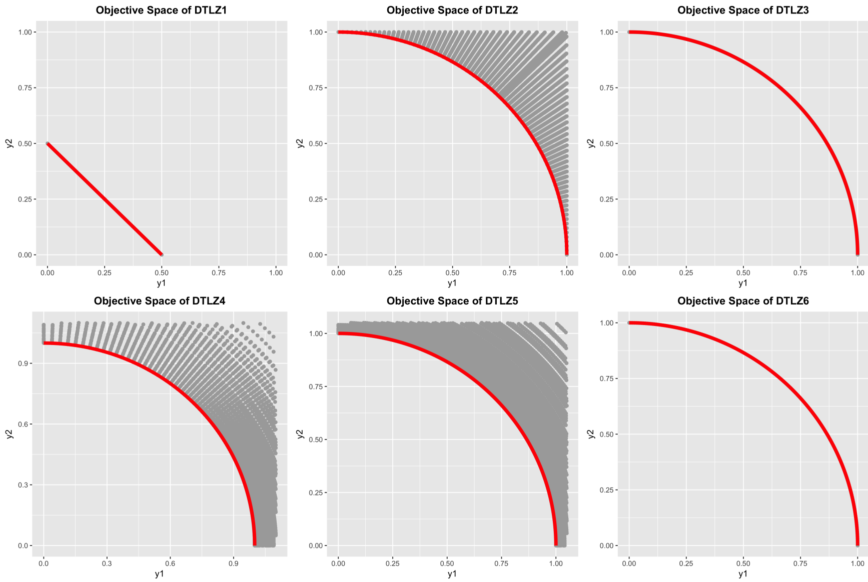
Block 3 - Part I: Multi-Objective "Global" Opt.

- DTLZ1 to DTLZ7[7] and ZDT1 to ZDT6[8] (without ZDT5)
  $\rightsquigarrow$ 120 instances (12 functions with 10 replicates each)
- initial designs: $100 \times D$ samples with $D = 3$
- considered 131 artificially designed "interaction-features":
  - all 15 feature sets except for GCM and Barrier Trees
  - aggregated objectives (objective 1 / objective 2) per feature
  - discarded runtimes, as well as all features that contained infinite or non-defined values

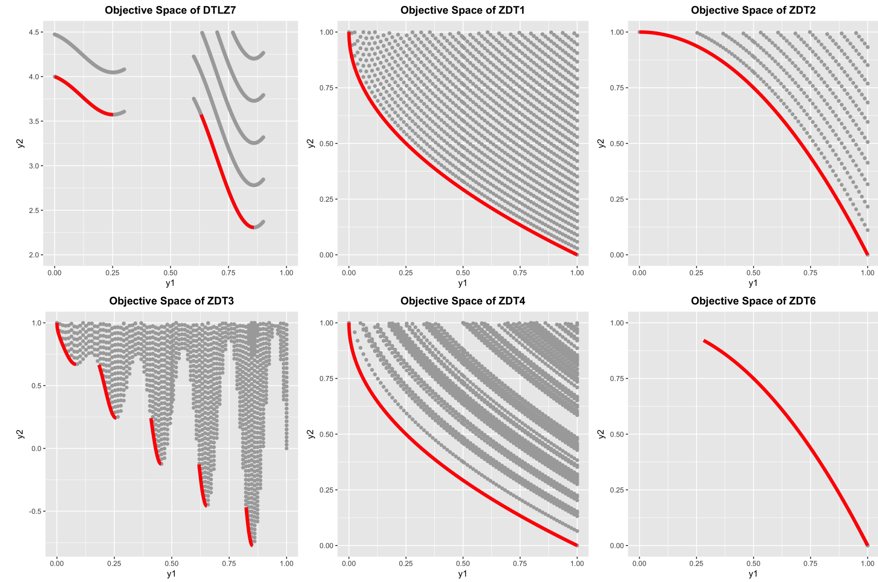
[7] Deb, K., Thiele, L., Laumanns, M. & Zitzler, E. (2001). Scalable Multi-Objective Optimization Test Problems. In Proceedings of CEC 2002 (pp. 825 - 830)

[8] Zitzler, E., Deb, K. & Thiele (2000). Comparison of Multiobjective Evolutionary Algorithms: Empirical Results. In Journal of Evolutionary Computation (pp. 173 - 195)

Block 3 - Part I: Multi-Objective “Global” Opt.



Block 3 - Part I: Multi-Objective “Global” Opt.

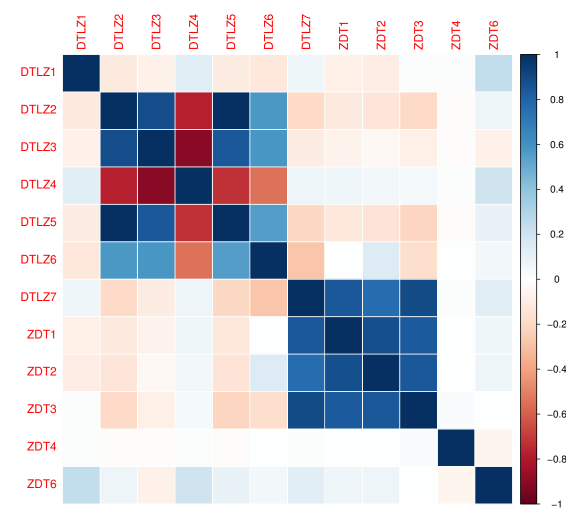


Block 3 - Part I: Multi-Objective “Global” Opt.

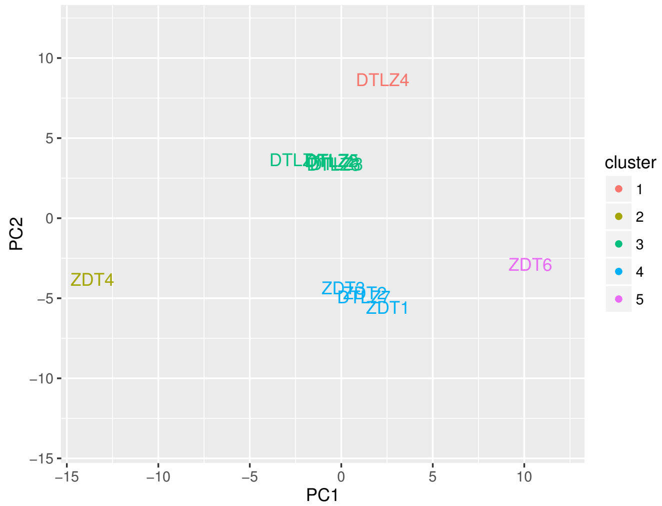
- the objective spaces of the 12 MOPs show some similarities across the problems, e.g.
  - DTLZ2 ≈ DTLZ4
  - ZDT4 ≈ DTLZ5 ≈ ZDT1
- the objective space of DTLZ7 looks very different to all the others

Do the features meet our expectations and group the MOPs accordingly?

Block 3 - Part I: Multi-Objective “Global” Opt.



Kerschke, P. & Trautmann, H. (2016). *The R-Package FLACCO for Exploratory Landscape Analysis with Applications to Multi-Objective Optimization Problems.* In: Proceedings of CEC 2016.



Kerschke, P. & Trautmann, H. (2016). *The R-Package FLACCO for Exploratory Landscape Analysis with Applications to Multi-Objective Optimization Problems*. In: Proceedings of CEC 2016.

Why do the features group the MOPs differently?

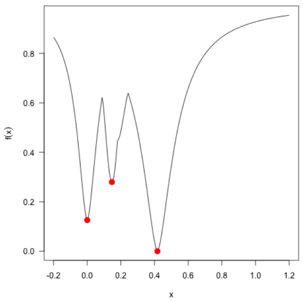
- many features are based on the decision space
- remember: many algorithms also act in the decision space (e.g., mutation / recombination within EAs)
- did not use any sophisticated features (just the feature-wise ratios between the objectives)

We need something to characterize and/or distinguish the multi-objective landscapes!

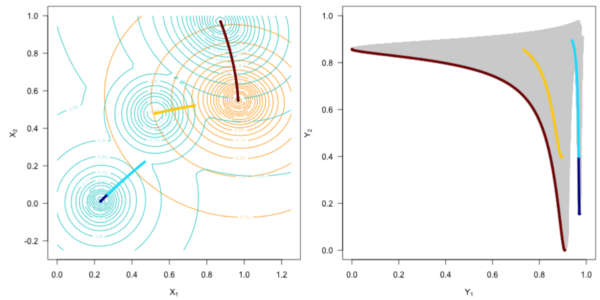
1. start with a "white-box"-approach and "measure" some (rather obvious) characteristics
2. once we know which of the characteristics might be useful, we can (and should!) develop landscape features that measure the information of these characteristics

ELA for Multi-Objective Multimodal Optimization Problems

Single-Objective Optimization



Multi-Objective Optimization



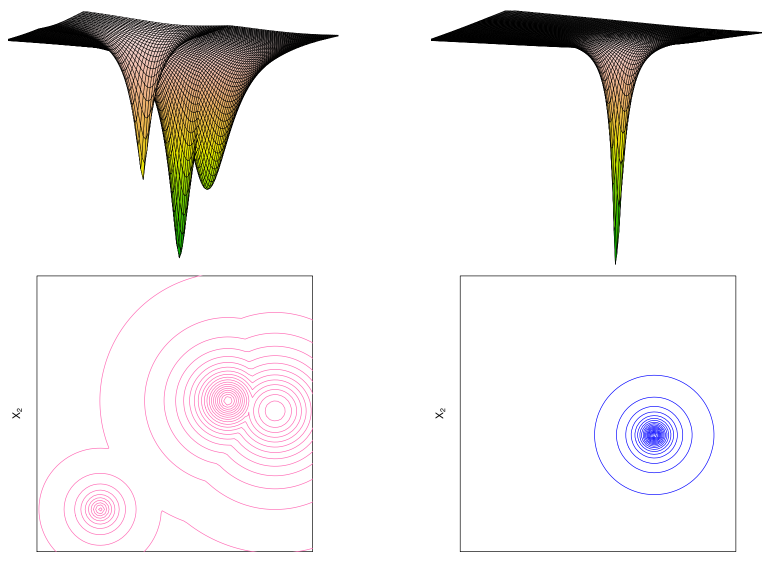
Block 3 - Part II: Multi-Objective Multimodal Opt.

- definition of multimodality for multi-objective problems within our PPSN paper from 2016:
  Kerschke, P., Wang, H., Preuss, M., Grimme, C., Deutz, A., Trautmann, H., & Emmerich, M. (2016). *Towards Analyzing Multimodality of Multiobjective Landscapes*. In: Proceedings of PPSN XIV, Edinburgh, Scotland, pp. 962–972 (Best Paper Award).
- visualized multimodality on a set of simple, but configurable problems ⇝ bi-objective mixed-sphere problems
  (using an adaptation of the MPM2-generator[9])

[9] multiple peaks model 2 generator, available in python (`optproblems0.9`, Wessing, S.) and R (`smoof`, Bossek, J.)

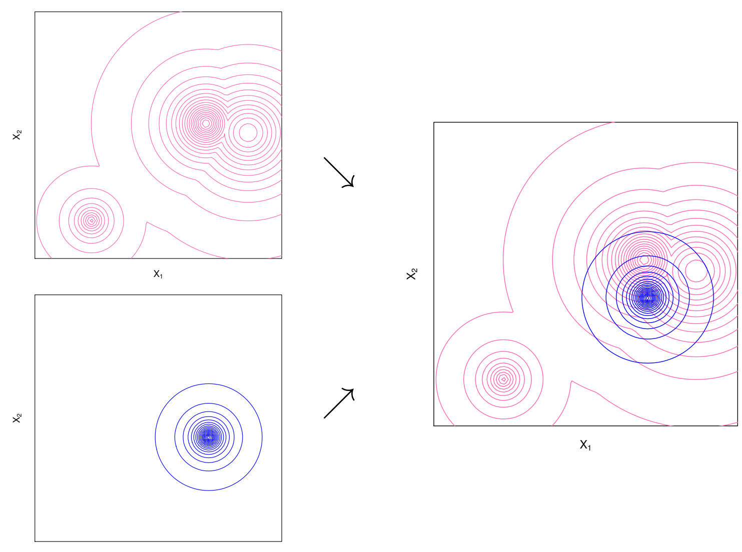
Block 3 - Part II: Multi-Objective Multimodal Opt.

Mixed-Sphere Problems



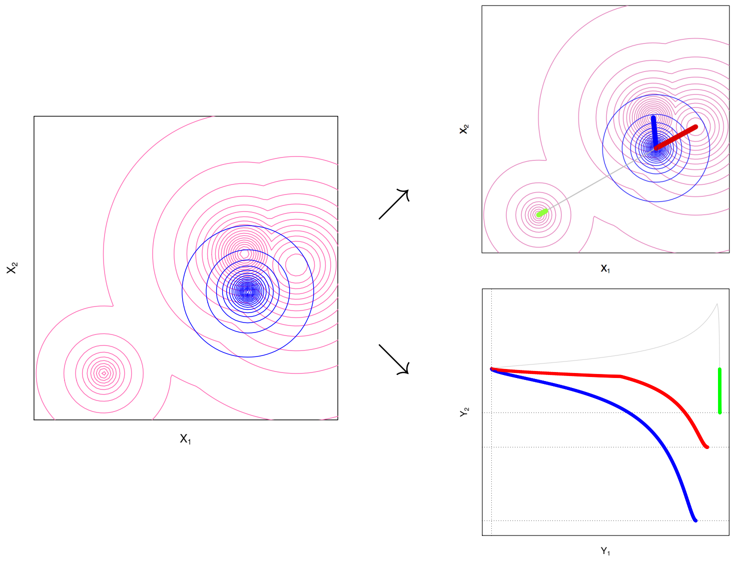
Block 3 - Part II: Multi-Objective Multimodal Opt.

Mixed-Sphere Problems



Block 3 - Part II: Multi-Objective Multimodal Opt.

Mixed-Sphere Problems



Block 3 - Part II: Multi-Objective Multimodal Opt.

Characteristics $\neq$ Features

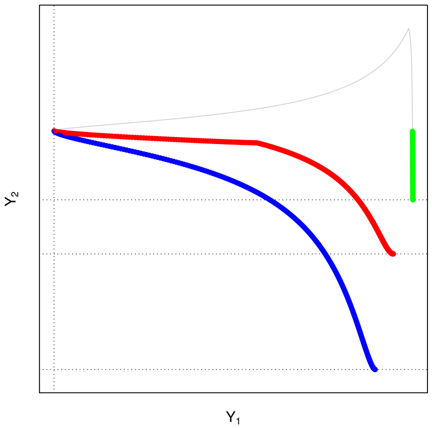
- characteristics use knowledge of the entire landscape (white-box)
- features are based on a small (!) sample of points from the problem

Block 3 - Part II: Multi-Objective Multimodal Opt.

Measuring the Multimodality

possible characteristics:

1. percentage of counts of global to local Pareto fronts
2. percentage of lengths of global to local Pareto fronts
3. (1) for connected fronts
4. (2) for connected fronts
5. ...



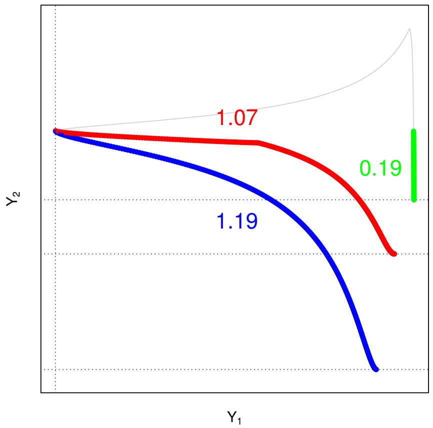
$$\frac{1}{1+1+1} = \frac{1}{3} \approx 0.33$$

Block 3 - Part II: Multi-Objective Multimodal Opt.

Measuring the Multimodality

possible characteristics:

1. percentage of counts of global to local Pareto fronts
2. percentage of lengths of global to local Pareto fronts
3. (1) for connected fronts
4. (2) for connected fronts
5. ...



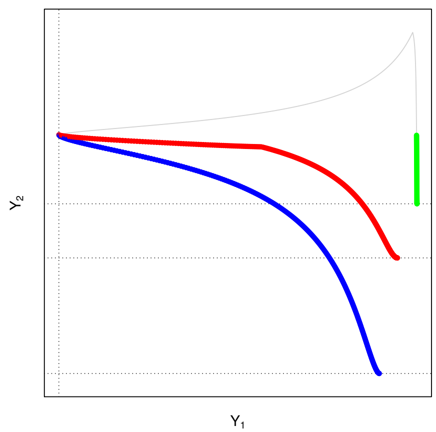
$$\frac{1.19}{1.19+1.07+0.19} = \frac{1.19}{2.45} \approx 0.49$$

Block 3 - Part II: Multi-Objective Multimodal Opt.

Measuring the Multimodality

possible characteristics:

1. percentage of counts of global to local Pareto fronts
2. percentage of lengths of global to local Pareto fronts
3. (1) for connected fronts
4. (2) for connected fronts
5. ...



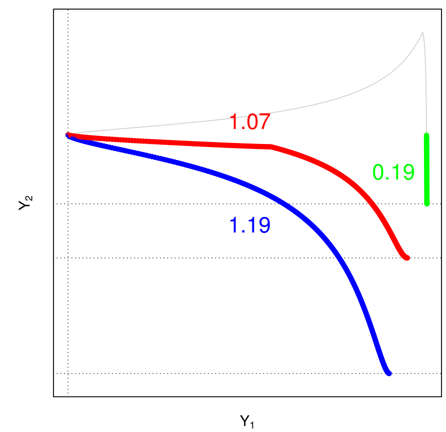
$$\frac{1+1}{1+1+1} = \frac{2}{3} \approx 0.67$$

Block 3 - Part II: Multi-Objective Multimodal Opt.

Measuring the Multimodality

possible characteristics:

1. percentage of counts of global to local Pareto fronts
2. percentage of lengths of global to local Pareto fronts
3. (1) for connected fronts
4. (2) for connected fronts
5. ...

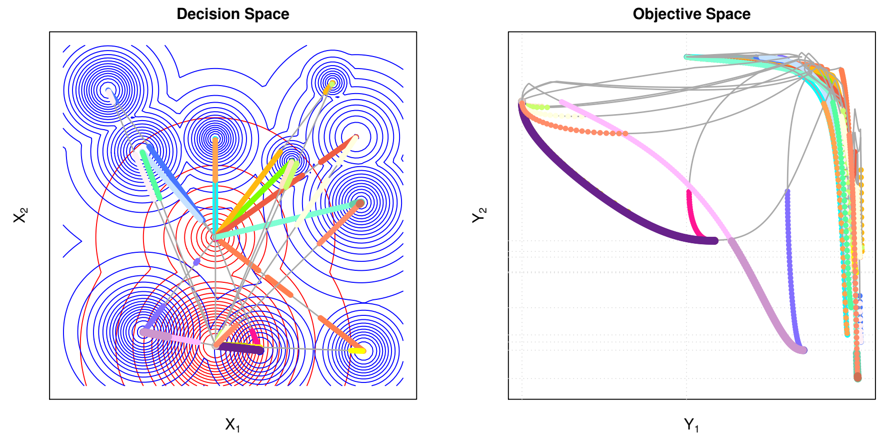


$$\frac{1.19+1.07}{1.19+1.07+0.19} = \frac{2.26}{2.45} \approx 0.92$$

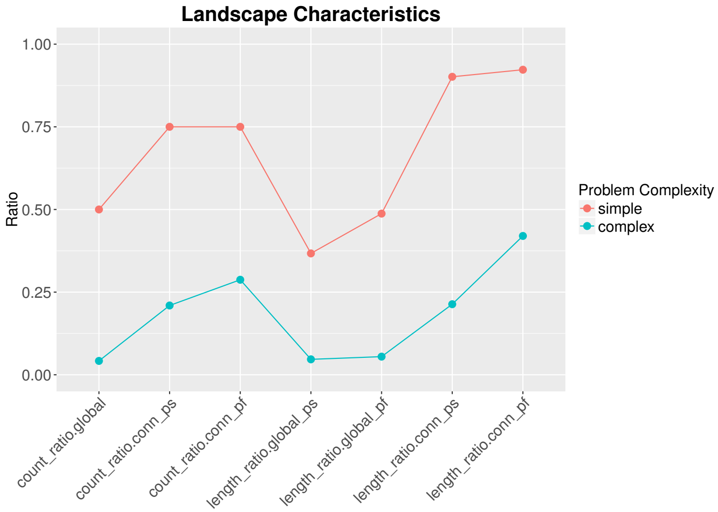
Block 3 - Part II: Multi-Objective Multimodal Opt.

Quite simple for small problems. But what happens if the problems become (just a little bit) more multimodal?

Block 3 - Part II: Multi-Objective Multimodal Opt.



Block 3 - Part II: Multi-Objective Multimodal Opt.



Kerschke, P., Wang, H., Preuss, M., Grimme, C., Deutz, A., Trautmann, H., & Emmerich, M. (2016). *Towards Analyzing Multimodality of Multiobjective Landscapes.* In Proceedings of PPSN 2016 (pp. 962-972)

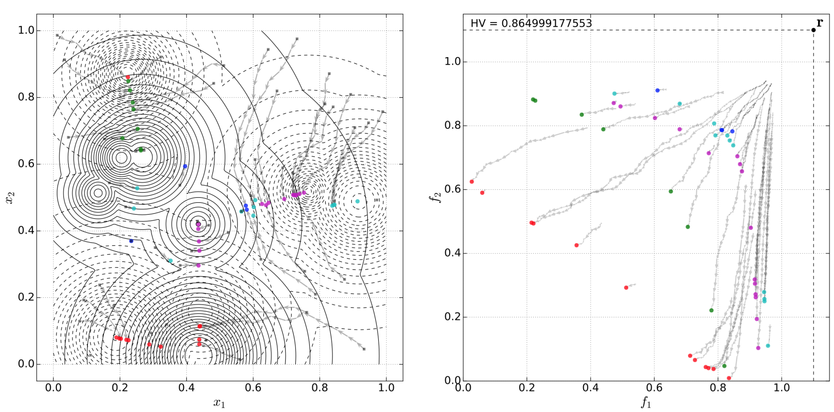
Block 3 - Part II: Multi-Objective Multimodal Opt.

Why is it necessary / useful to know the multimodality?

- optimizers behave differently:
  - a general optimizer rather detects the local fronts
  - a global optimizer tries to find the global Pareto front(s)

Block 3 - Part II: Multi-Objective Multimodal Opt.

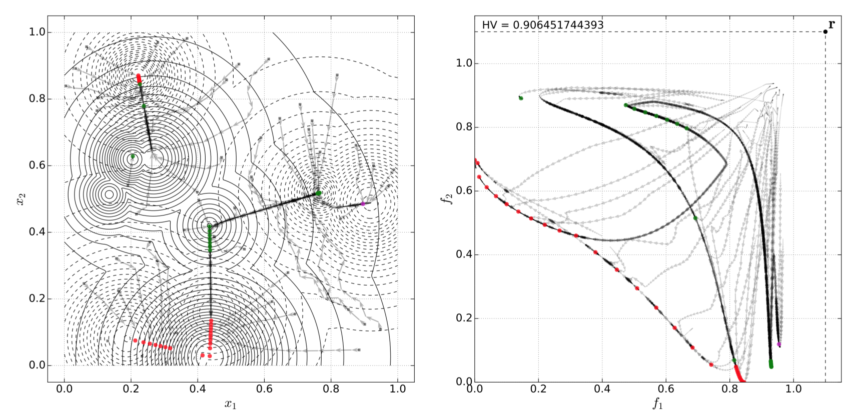
Naive Stochastic Local Search (SLS) Algorithm:



Kerschke, P., Wang, H., Preuss, M., Grimme, C., Deutz, A., Trautmann, H., & Emmerich, M. (2016). *Towards Analyzing Multimodality of Multiobjective Landscapes.* In Proceedings of PPSN 2016 (pp. 962-972)

Block 3 - Part II: Multi-Objective Multimodal Opt.

Hypervolume Indicator Gradient Ascent Multi-Objective Optimization (HIGA-MO) Algorithm:

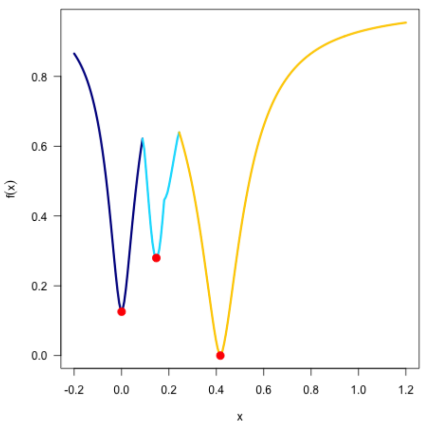


Kerschke, P., Wang, H., Preuss, M., Grimme, C., Deutz, A., Trautmann, H., & Emmerich, M. (2016). *Towards Analyzing Multimodality of Multiobjective Landscapes.* In Proceedings of PPSN 2016 (pp. 962-972)

Block 3 - Part II: Multi-Objective Multimodal Opt.

Basin of Attraction

Single-Objective Optimization

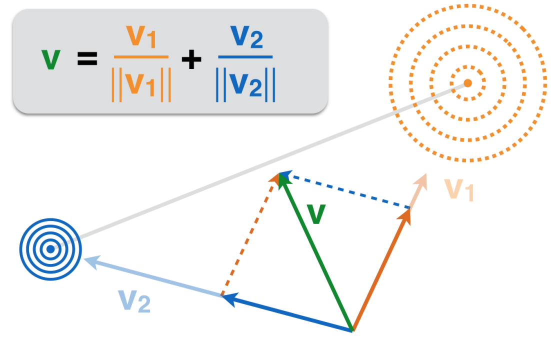


Multi-Objective Optimization

Block 3 - Part II: Multi-Objective Multimodal Opt.

Visualizing the Basins of Attraction

1. Combined Gradient



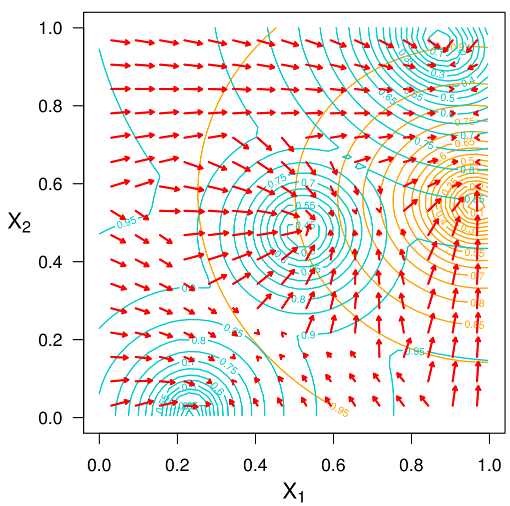
- length of combined gradient $\longleftrightarrow$ cone of dominance
- move to next cell $\Longleftrightarrow$ $\|\mathbf{v}\| > \varepsilon$

Kerschke, P. & Grimme, C. (2017). *An Expedition to Multimodal Multi-Objective Optimization Landscapes*. In Proceedings of EMO 2017 (pp. 329-343)

Block 3 - Part II: Multi-Objective Multimodal Opt.

Visualizing the Basins of Attraction

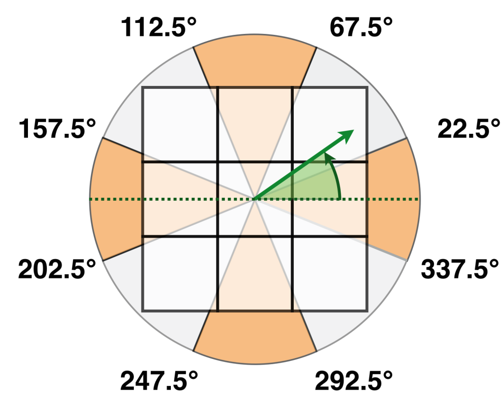
2. Gradient Field



Block 3 - Part II: Multi-Objective Multimodal Opt.

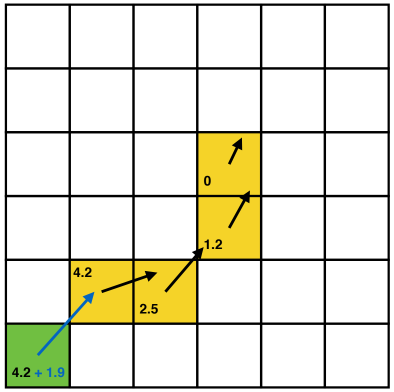
Visualizing the Basins of Attraction

3. Direction for Next Step

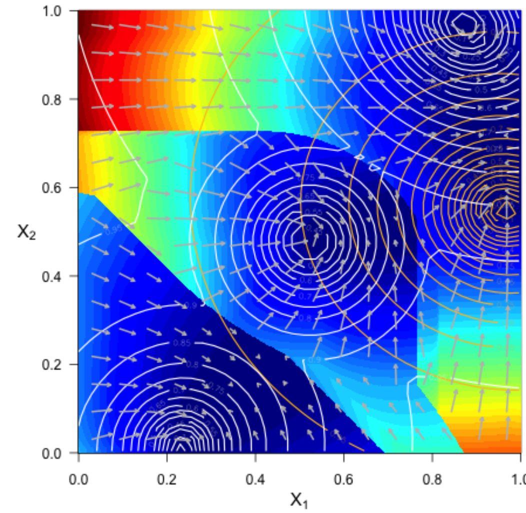


Kerschke, P. & Grimme, C. (2017). *An Expedition to Multimodal Multi-Objective Optimization Landscapes*. In Proceedings of EMO 2017 (pp. 329-343)

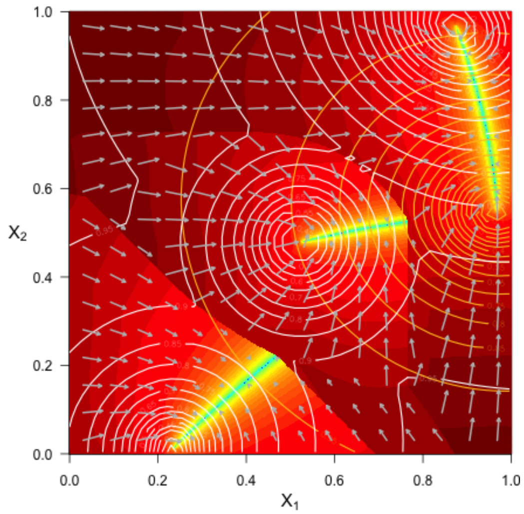
4. Cumulated Gradient Paths

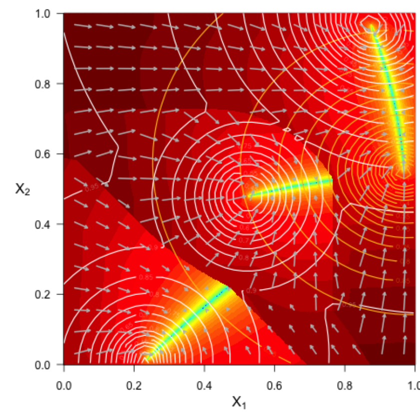


5. Heatmap of Cumulated Gradient Paths



5. Heatmap of Cumulated Gradient Paths

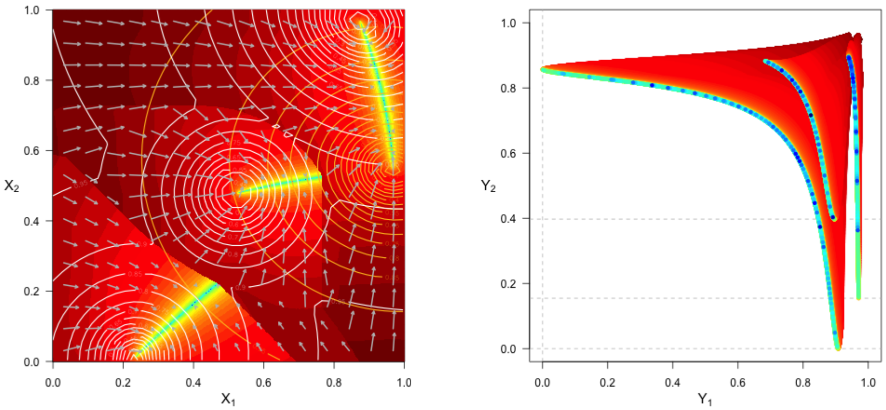




- basin of attraction
- (joint vs. disconnected) local efficient sets
- multi-objective ball
- discontinuities & ridges

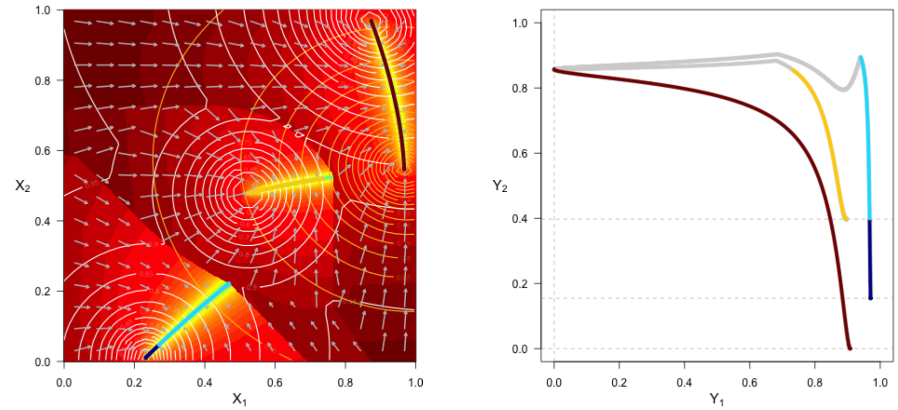
Block 3 - Part II: Multi-Objective Multimodal Opt.

Visualizing the Basins of Attraction



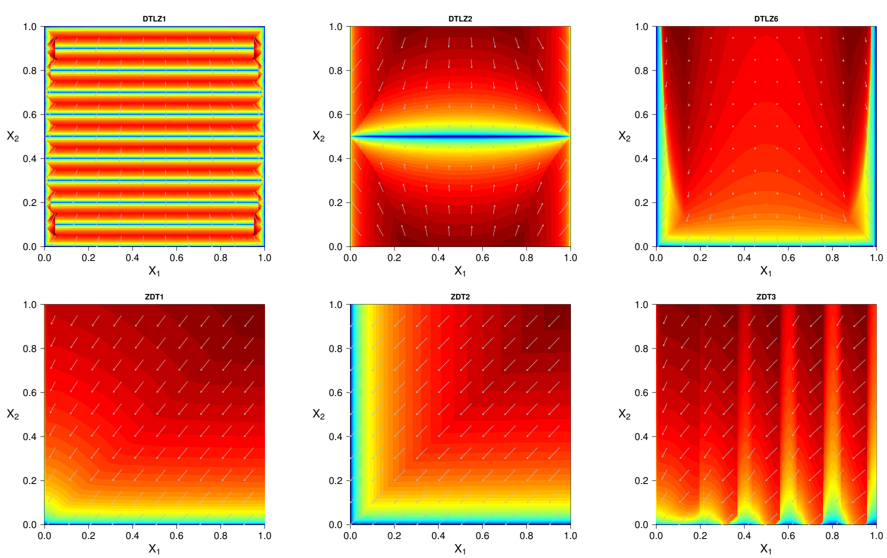
Block 3 - Part II: Multi-Objective Multimodal Opt.

Visualizing the Basins of Attraction



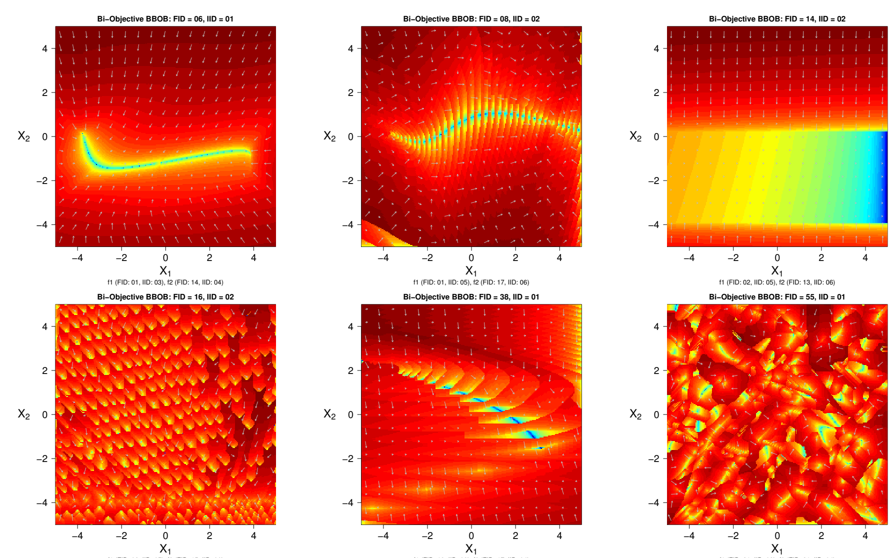
Block 3 - Part II: Multi-Objective Multimodal Opt.

Visualizing the Basins of Attraction – DTLZ & ZDT



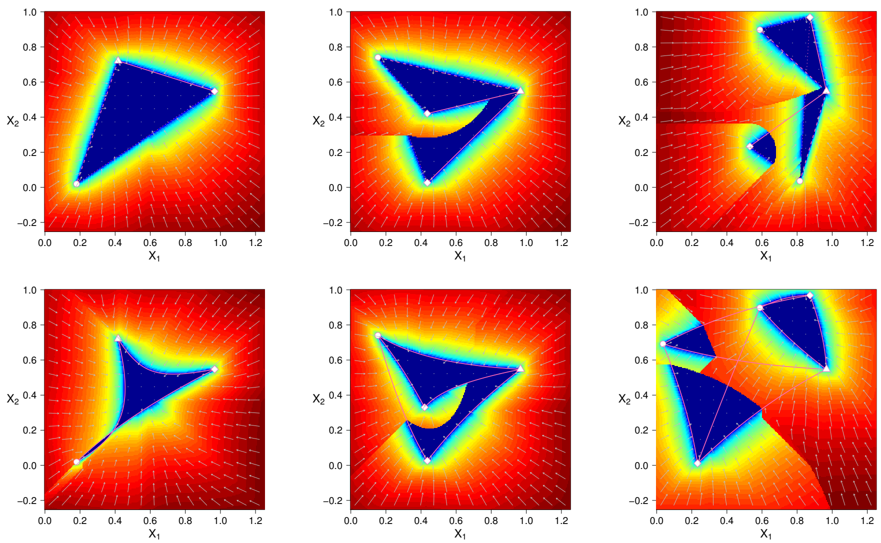
Block 3 - Part II: Multi-Objective Multimodal Opt.

Visualizing the Basins of Attraction – Bi-Objective BBOB



Block 3 - Part II: Multi-Objective Multimodal Opt.

Visualizing the Basins of Attraction – 3 Objectives



Closing

Closing

- enhance `flacco` with more ELA features
- how can we find the smallest most informative feature set?
- by how much can we still reduce the size of the initial designs without losing (too much) information?!
- where can we find representative real-world problems / appropriate benchmarks?
- can we transfer landscape features from / to different domains?
- use ELA features for improved algorithm selection and/or configuration on different benchmarks (e.g., BBOB)