UNIVERSITY
OF APPLIED SCIENCES
UPPER AUSTRIA

HEURISTIC AND
EVOLUTIONARY
ALGORITHMS
LABORATORY

Vectorial Genetic Programming – Optimizing Segments for Feature Extraction

Philipp Fleck, Stephan Winkler, Michael Kommenda, Michael Affenzeller

Eurocast 2022, 2022-02-23

HAGENBERG | LINZ | STEYR | WELS

Melanie Bracewell
@meladoodle

For the second year in a row, despite many people agreeing, I have been the only one at work to dress up for Halloween

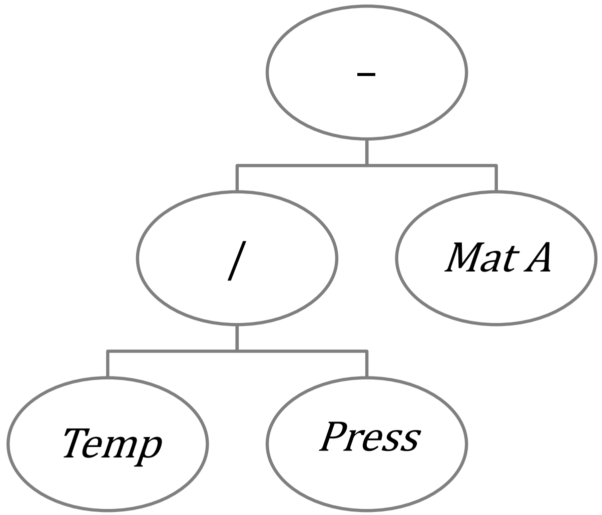
9:19 PM · Oct 30, 2018

2.2K

Symbolic Regression

ID	Mat A	Mat B	Temp.	Press.	…	Quality
1	10.3	22.0	25.0	1.40		**5.0**
2	12.0	23.0	22.8	1.45		**5.2**
3	11.5	22.4	24.8	1.55		**5.3**
n	…	…	…	…		**…**

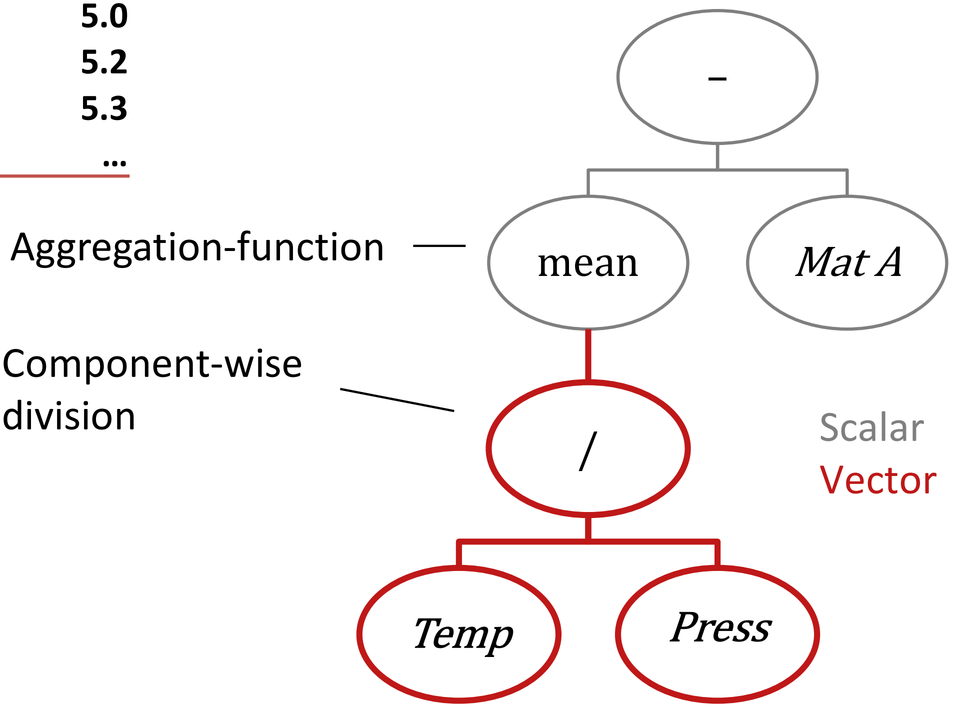
$$Quality = \frac{Temp}{Press} - Mat\ A$$



Symbolic Regression with Vectors

ID	Mat A	Mat B	Temp.	Press.	Quality
1	10.3	22.0	[20, 25, 30, ..., 25]	[1.5, 1.4, 1.8, ..., 1.2]	**5.0**
2	12.0	23.0	[18, 22, 23, ..., 28]	[1.6, 1.3, 1.7, ..., 1.2]	**5.2**
3	11.5	22.4	[20, 21, 28, ..., 30]	[1.6, 1.6, 1.7, ..., 1.3]	**5.3**
n	...	...	...	...	**...**

$$Quality = \mathrm{mean}\left(\frac{Temp}{Press}\right) - Mat\ A$$



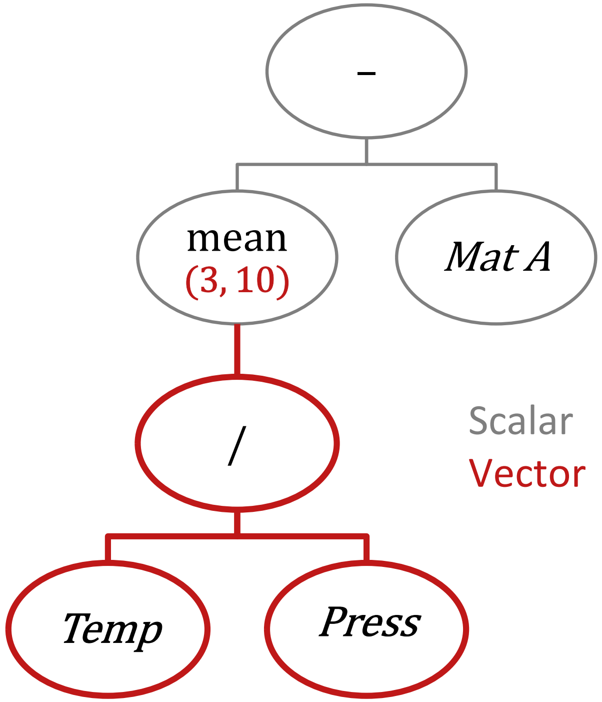
I. Azzali, L. Vanneschi, S. Silva, I. Bakurov, and M. Giacobini, "**A Vectorial Approach to Genetic Programming**," EuroGP, 2019.
P. Fleck, S. Winkler, M. Kommenda, M. Affenzeller, "**Grammar-based Vectorial Genetic Programming for Symbolic Regression**," GPTP, 2021.

Symbolic Regression with Vectors and Aggregation Windows

ID	Mat A	Mat B	Temp.	Press.	Quality
1	10.3	22.0	[20, 25, 30, ..., 25]	[1.5, 1.4, 1.8, ..., 1.2]	**5.0**
2	12.0	23.0	[18, 22, 23, ..., 28]	[1.6, 1.3, 1.7, ..., 1.2]	**5.2**
3	11.5	22.4	[20, 21, 28, ..., 30]	[1.6, 1.6, 1.7, ..., 1.3]	**5.3**
n	...	...	...	...	**...**

$$Quality = \mathrm{mean}_{3 \le t \le 10}\left(\frac{Temp}{Press}\right) - Mat\ A$$

Aggregation-window



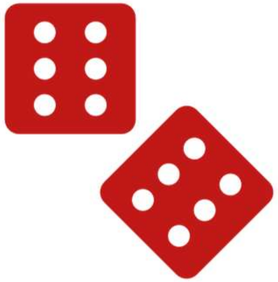
Optimization Sub-Problems in Symbolic Regression

	Genetic Programming
Model structure	✓
Local parameters	~
Aggregation-window	~

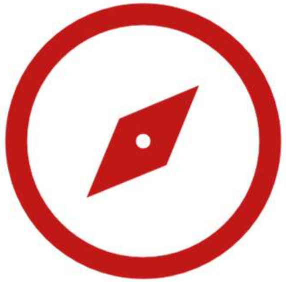
Kommenda, Michael, et al. "**Effects of constant optimization by nonlinear least squares minimization in symbolic regression**." *Proceedings of the 15th annual conference companion on Genetic and evolutionary computation*. 2013.

Optimization Strategies

- **Iterative random sampling**
  - (1, λ)-ES

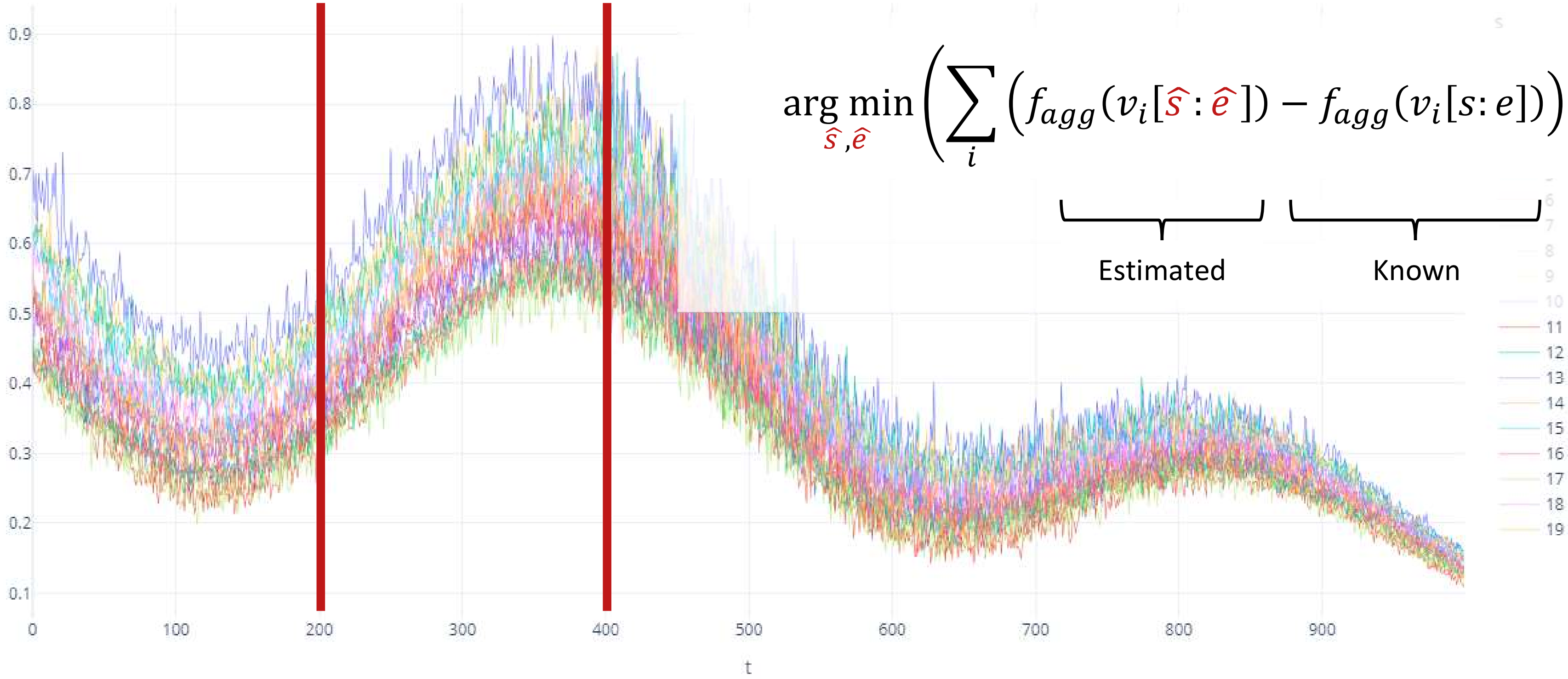


- **Iterative guided sampling**
  - Using an approximate gradient



Segment Optimization Problem

Known bounds + aggregation function → targets
(e.g. mean, std dev, …)

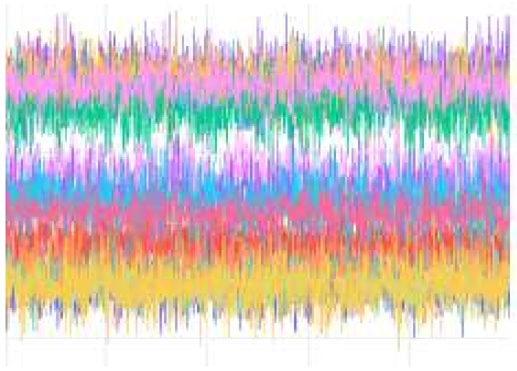


$$\underset{\hat{s},\hat{e}}{\arg\min}\left(\sum_i \left(\underbrace{f_{agg}(v_i[\hat{s}:\hat{e}])}_{\text{Estimated}} - \underbrace{f_{agg}(v_i[s:e])}_{\text{Known}}\right)^2\right)$$

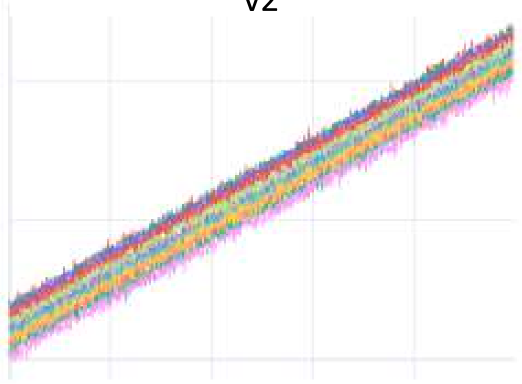
UNIVERSITY OF APPLIED SCIENCES UPPER AUSTRIA

Random Vector Problem Instances

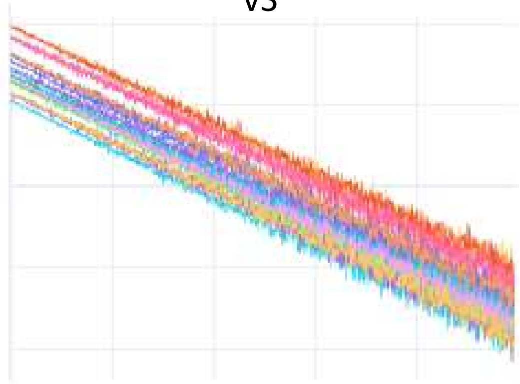
v1



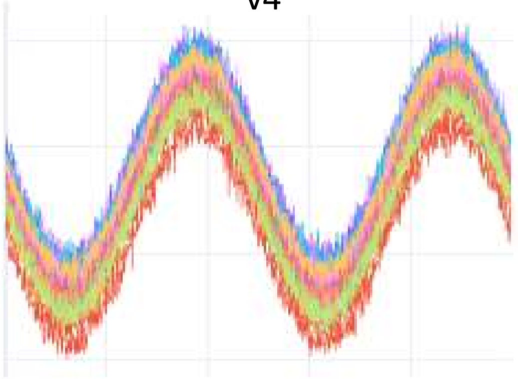
v2



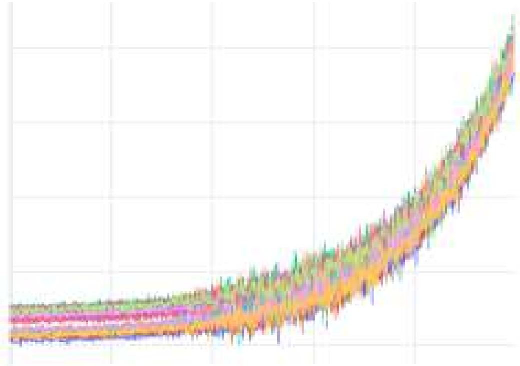
v3



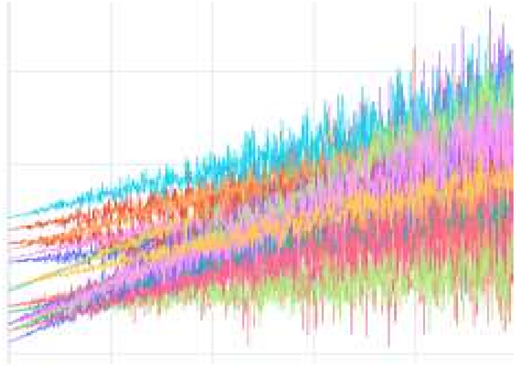
v4



v5

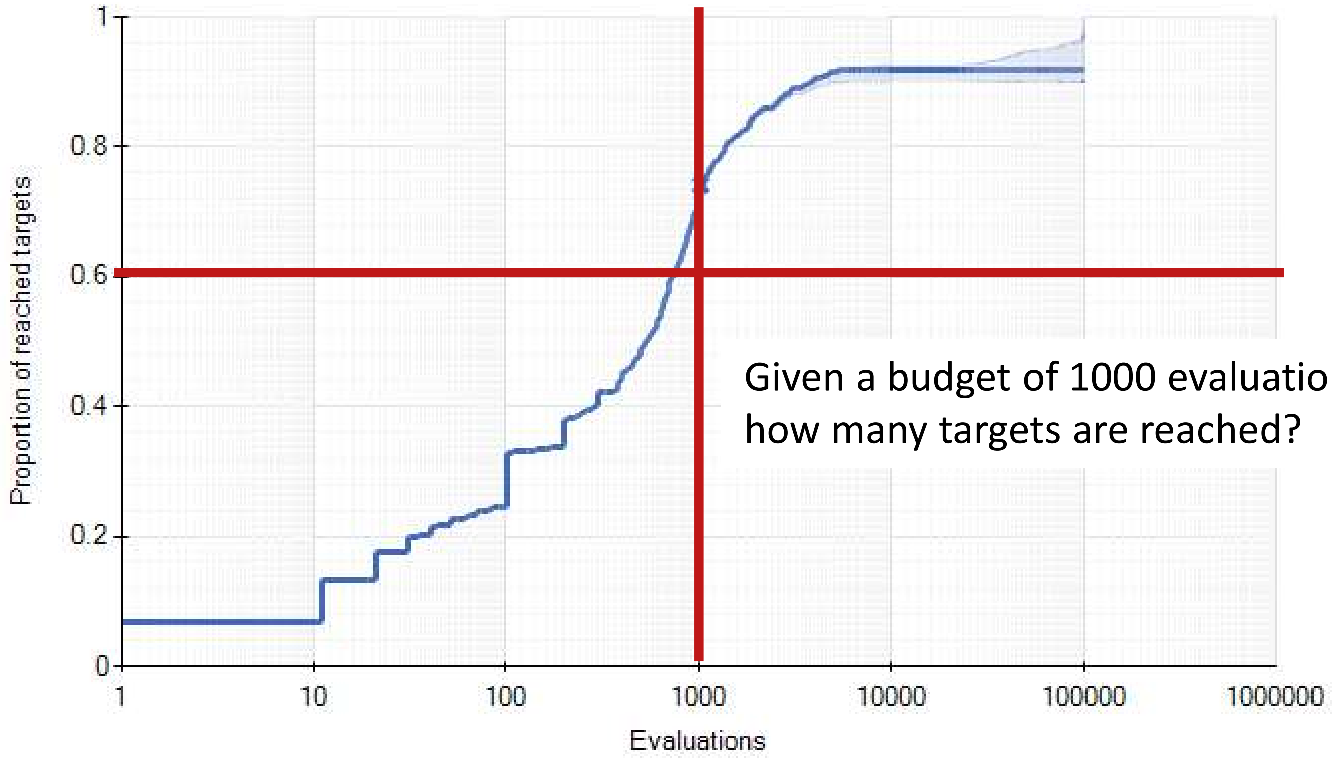


v6



Run-Length Distributions

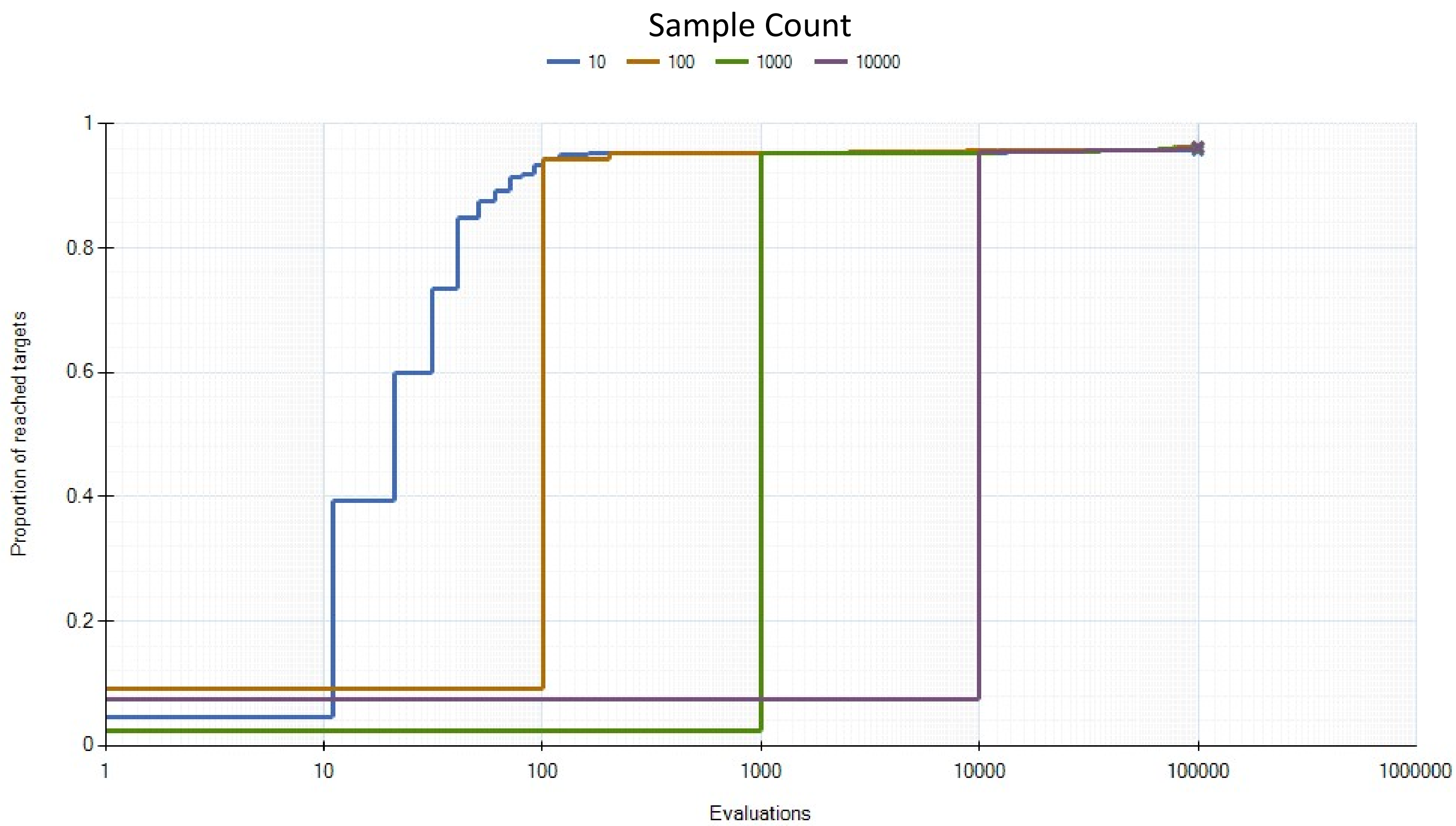
- **Used at GECCO's Black-Box Optimization Benchmarking**
- **Objective target (e.g. MSE)**
  - 0.10, 0.09, … , 0.01, 0.00
  - % of reached target after x evaluations



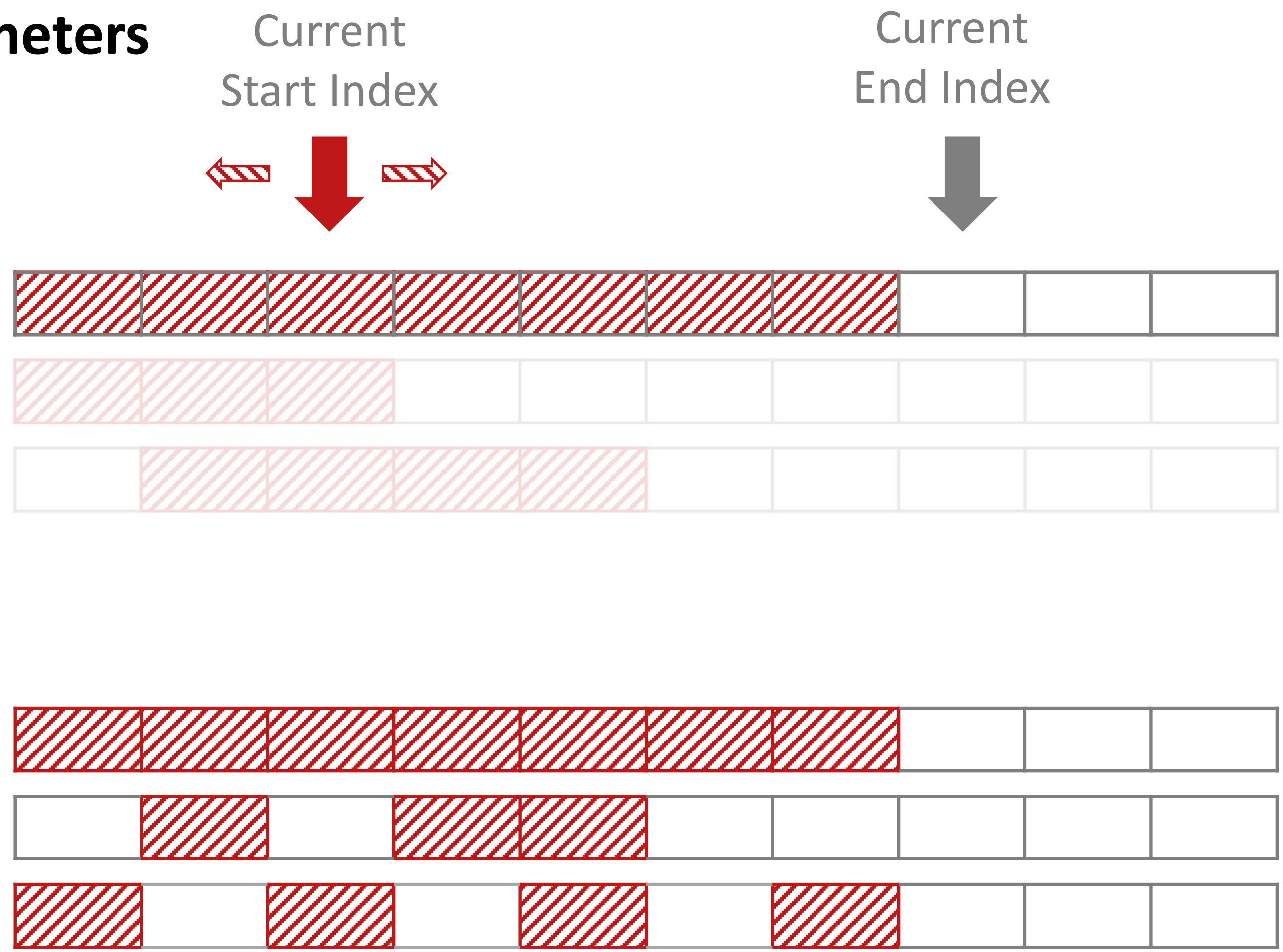
Given a budget of 1000 evaluations, how many targets are reached?

N. Hansen, A. Auger, R. Ros, S. Finck, and P. Pošík, "**Comparing results of 31 algorithms from the black-box optimization benchmarking BBOB-2009**," In Proceedings of the 12th annual conference companion on Genetic and evolutionary computation (pp. 1689-1696). 2010

Parameter Analysis – Sample Count



Search Space Parameters



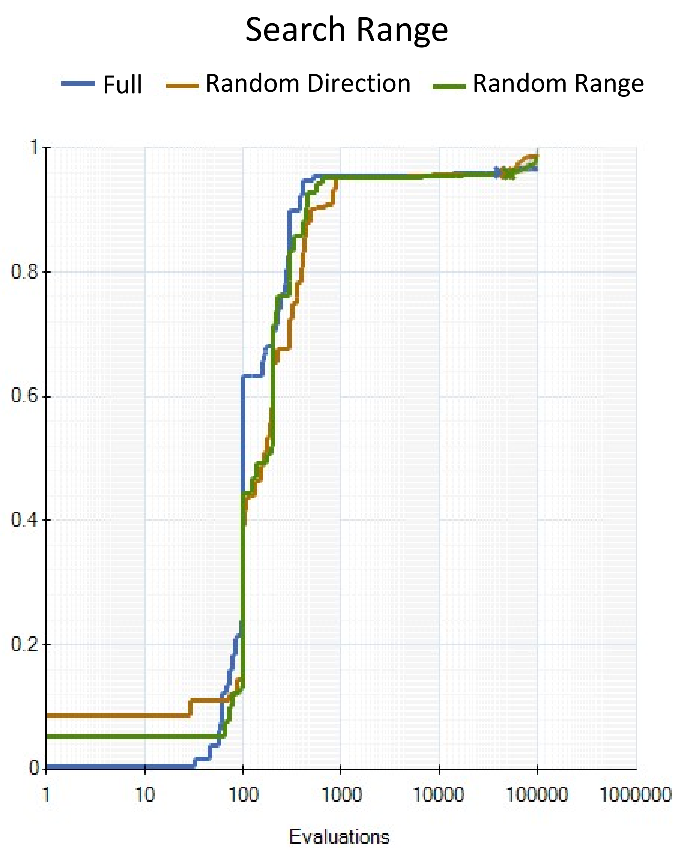
Search range

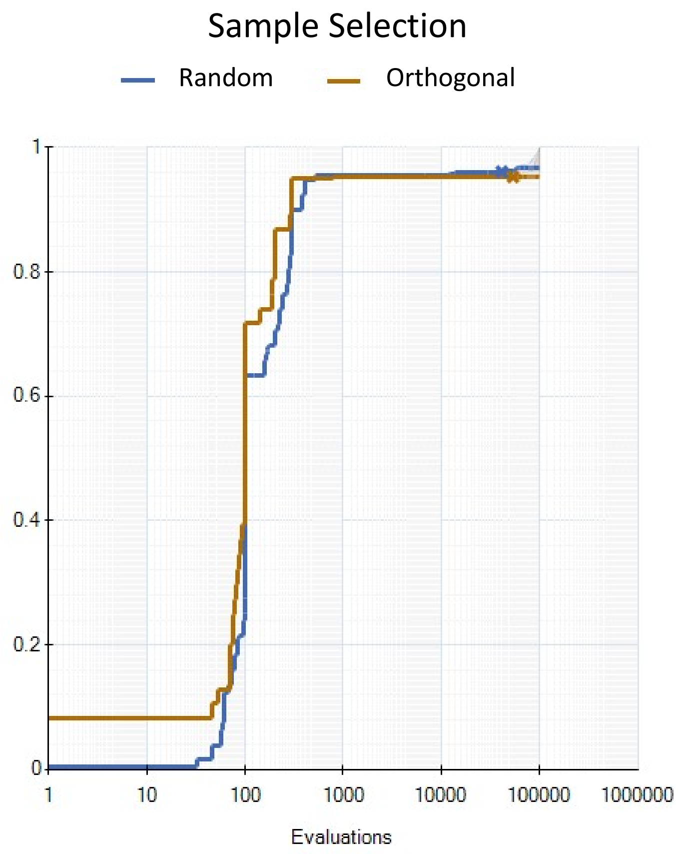
- Full
- Random direction
- Random range

Sample selection

- Exhaustive
- Random
- Orthogonal

Parameter Analysis – Search Space Configuration

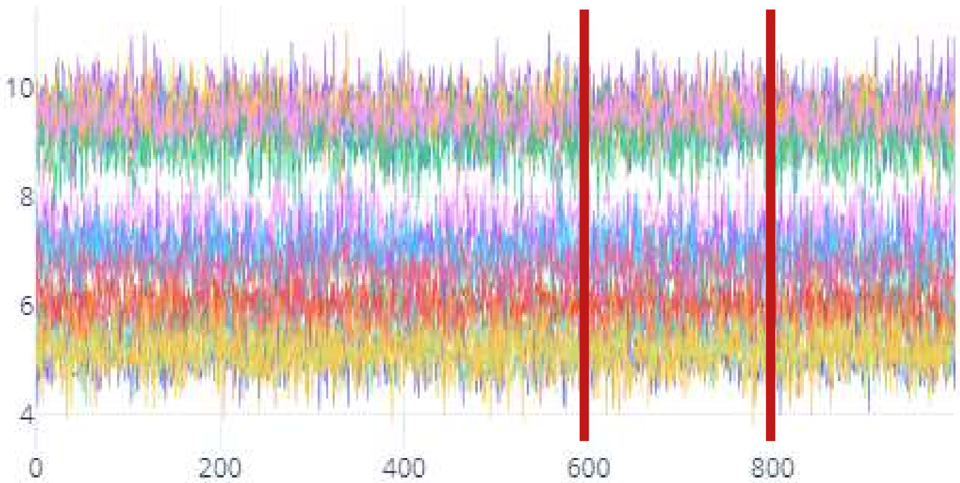


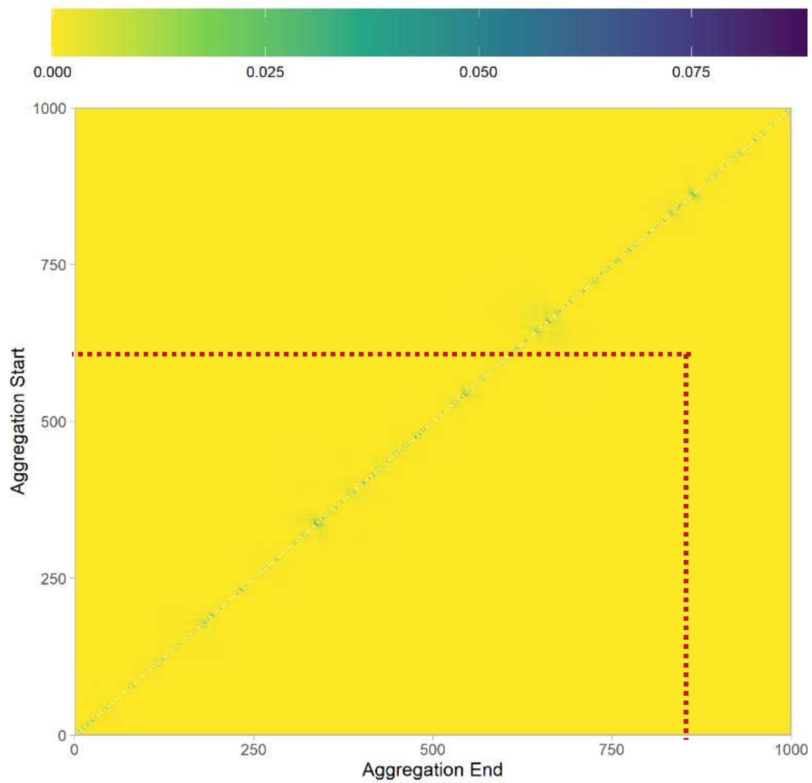


UNIVERSITY OF APPLIED SCIENCES UPPER AUSTRIA

Segment Optimization Fitness Landscape

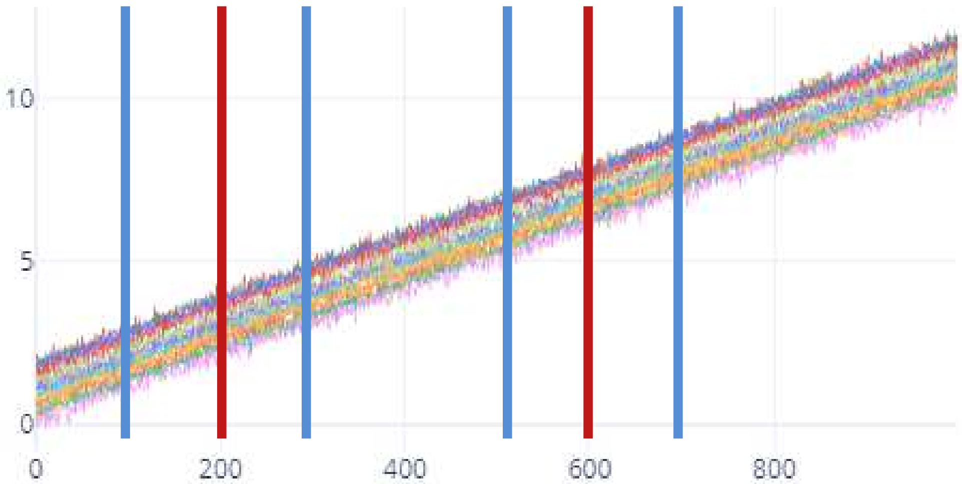
– **v1 instance (control group)**



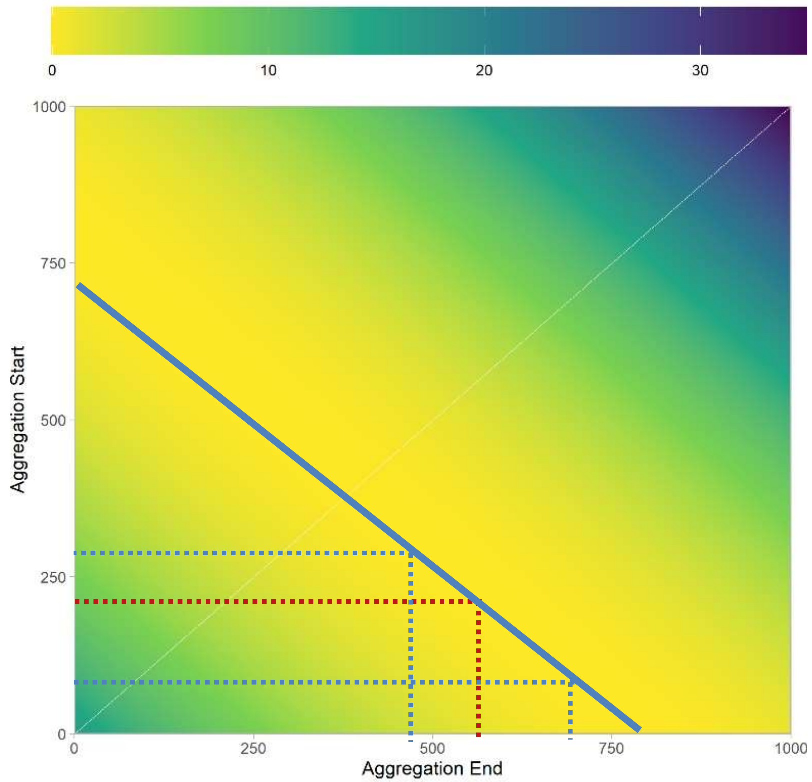


Segment Optimization Fitness Landscape

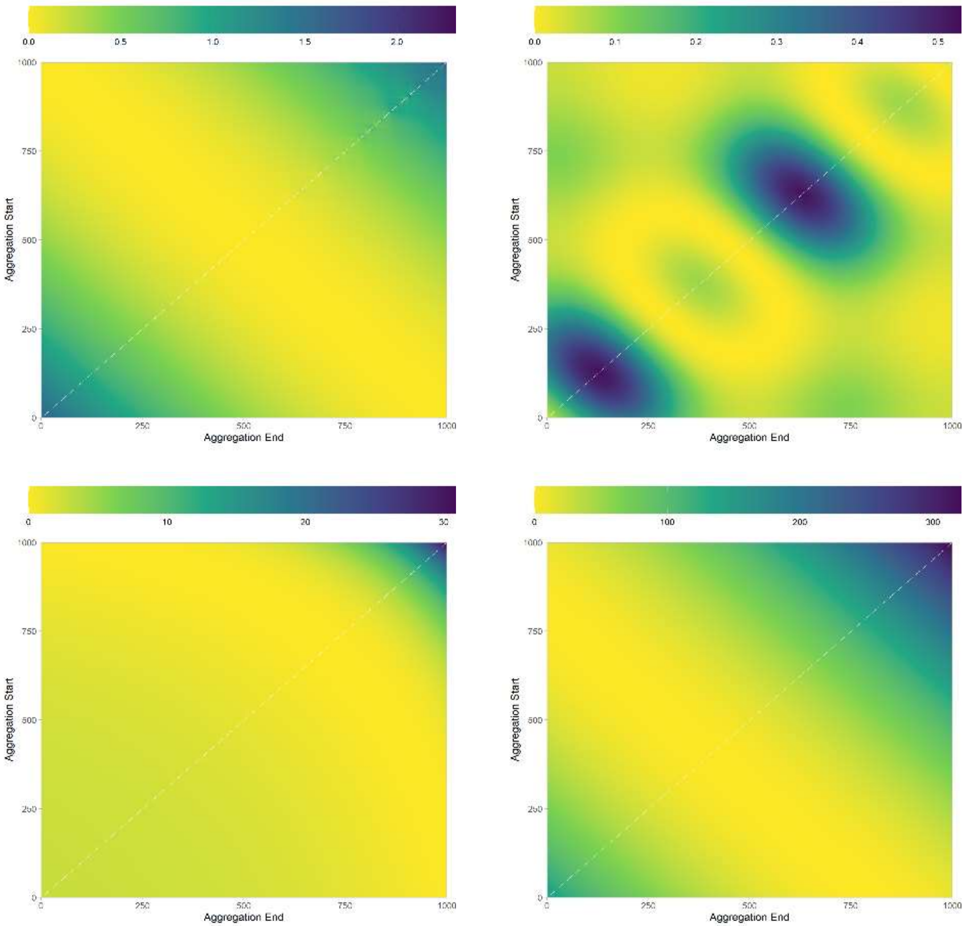
- **v2 instance**



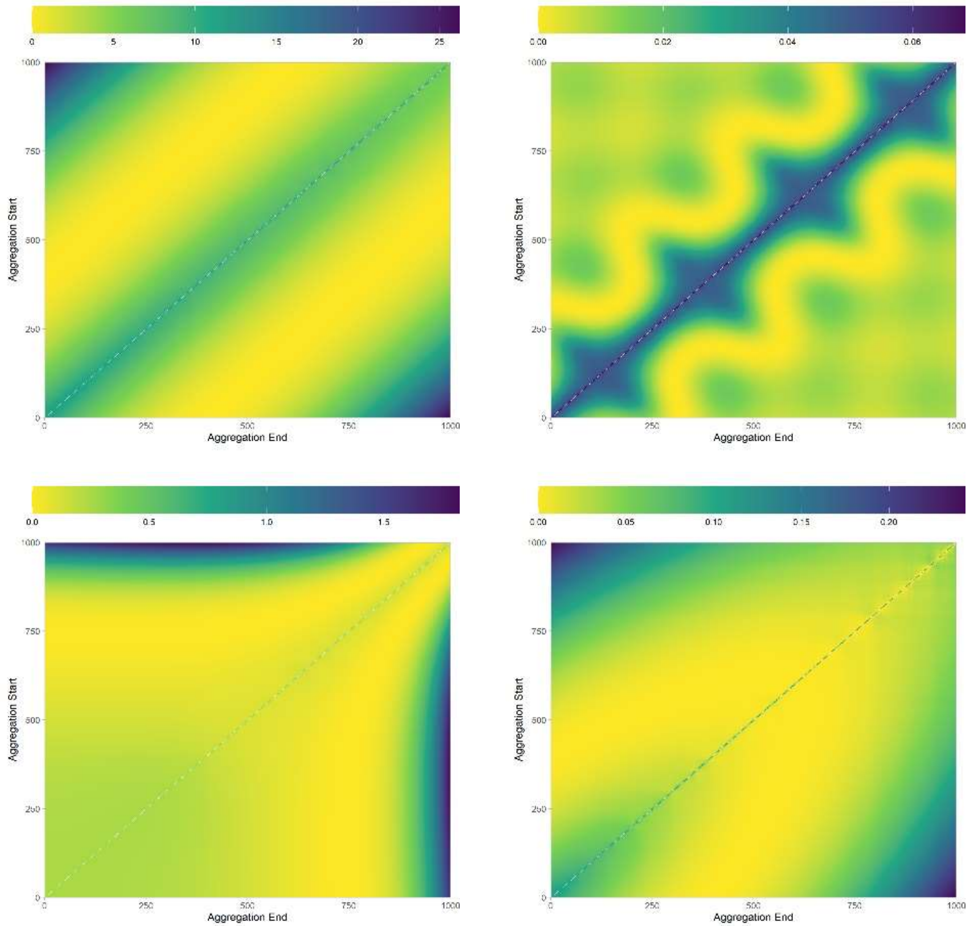
Known bounds



Aggregation: Mean

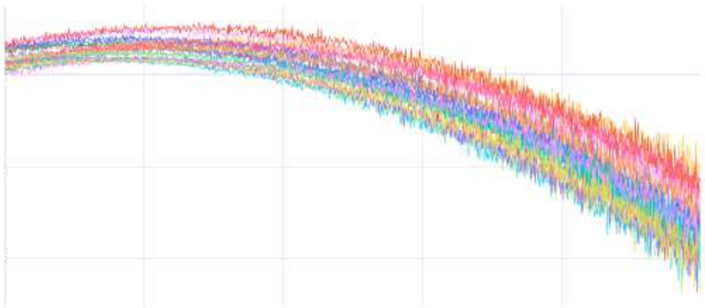


Aggregation: Standard deviation

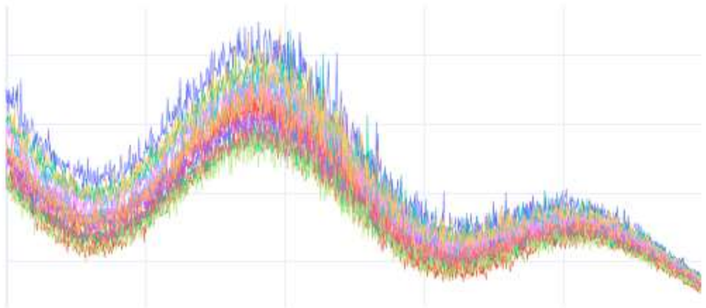


New Instances based with Vector Interactions

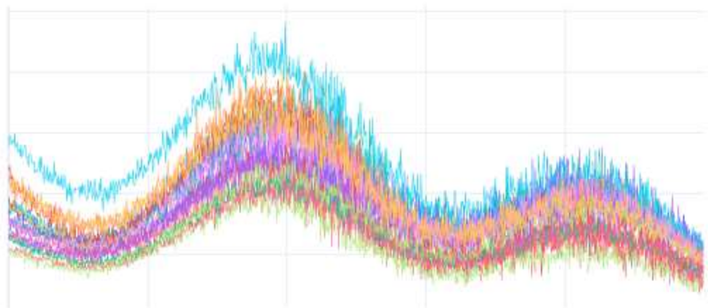
$x1 = v2 * v3$



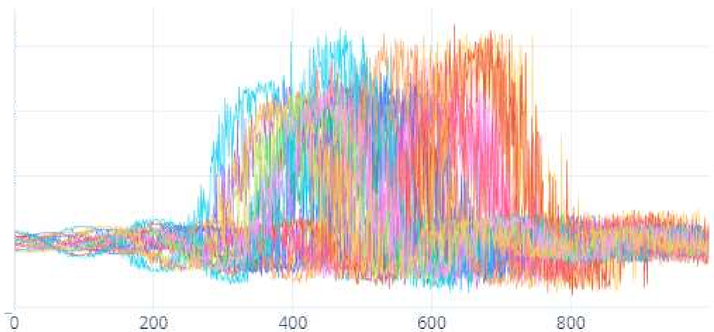
$x2 = v4/v5$



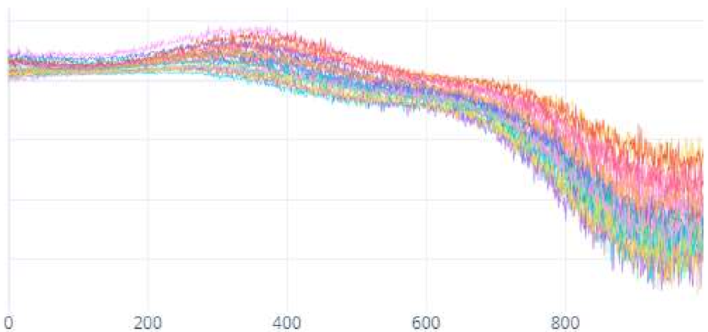
$x3 = v4 * v6/v5$



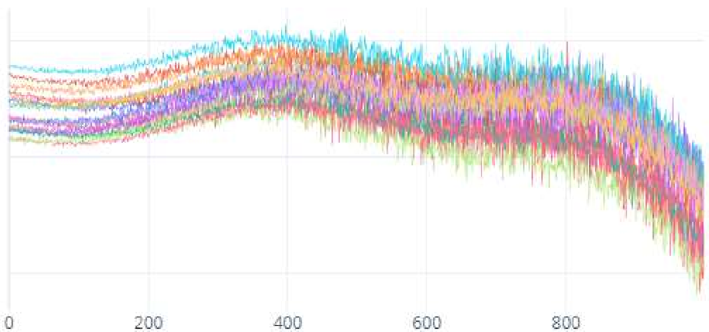
$x4 = \sin(v3 * v6)/v3$



$x5 = v2 * v3 * v4$



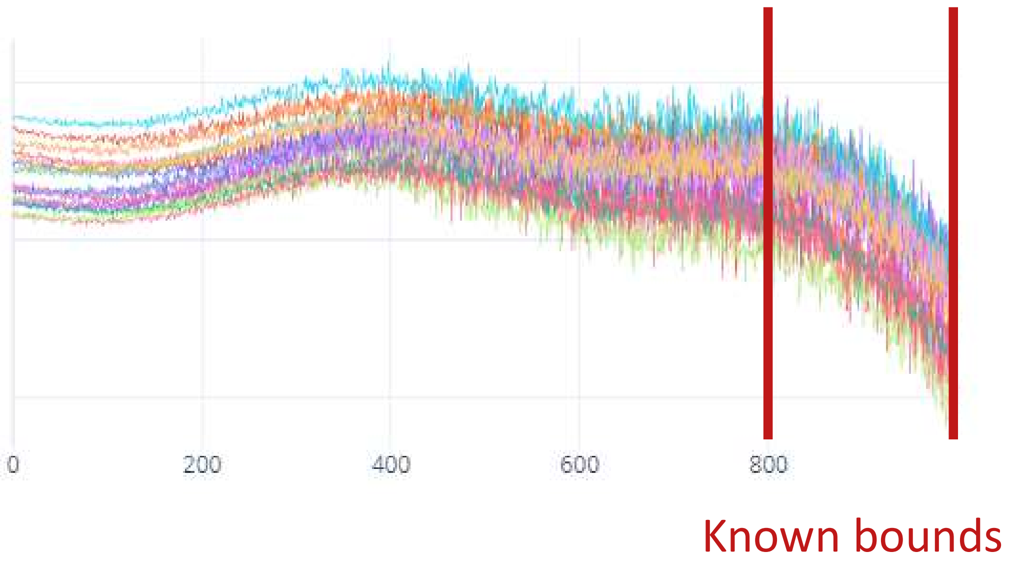
$x3 = v4 + v6 - v5$

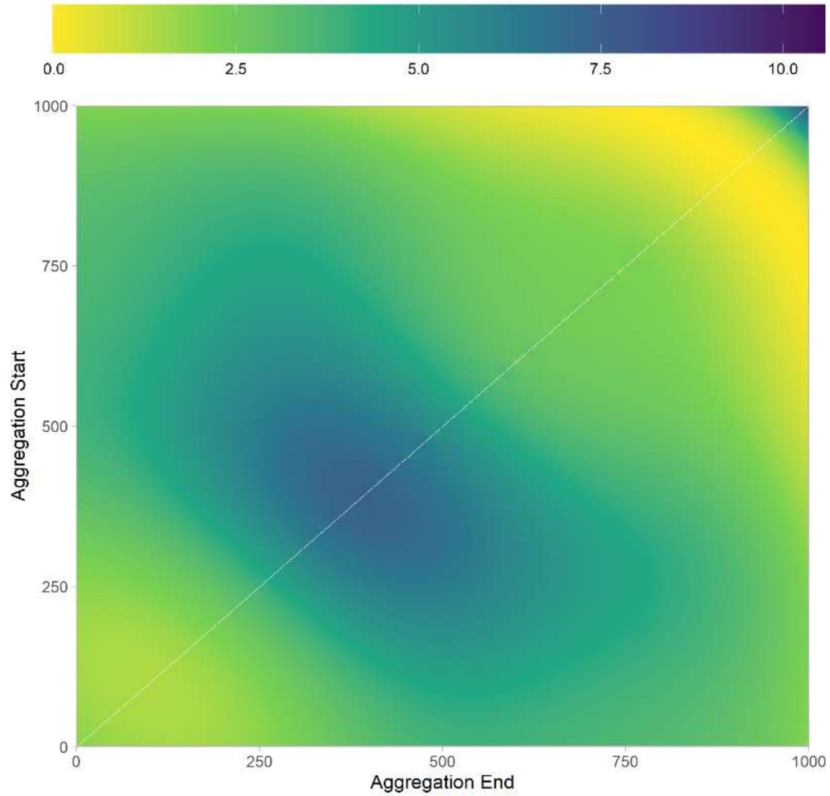


UNIVERSITY OF APPLIED SCIENCES UPPER AUSTRIA

Segment Optimization Fitness Landscape

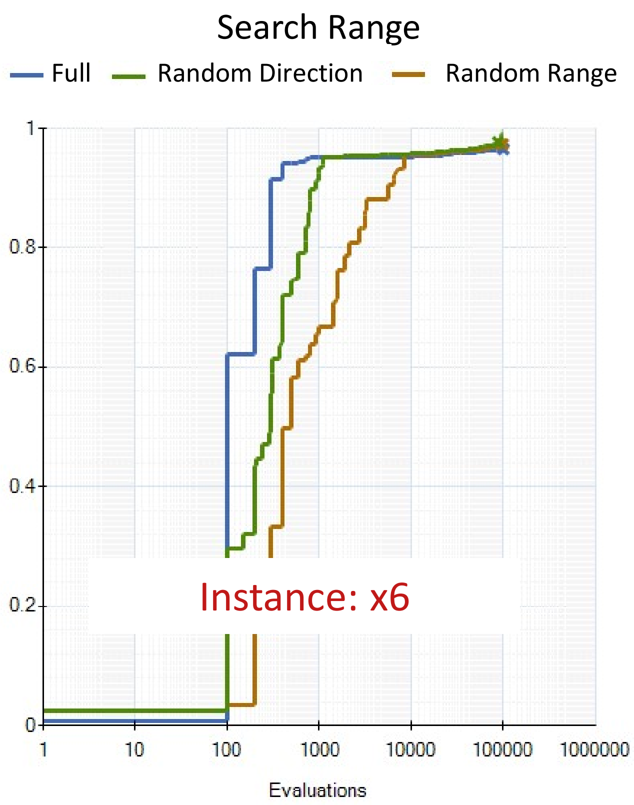
- x6 instance

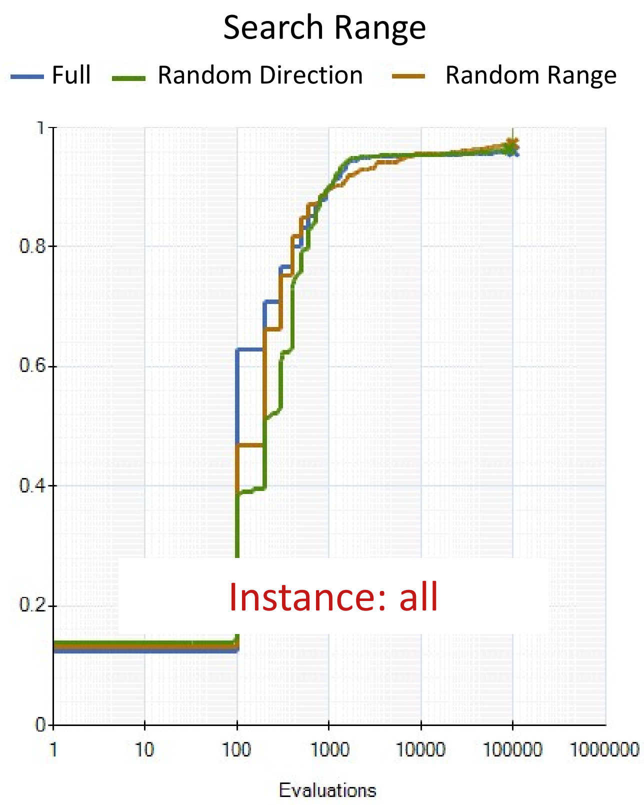




UNIVERSITY OF APPLIED SCIENCES UPPER AUSTRIA

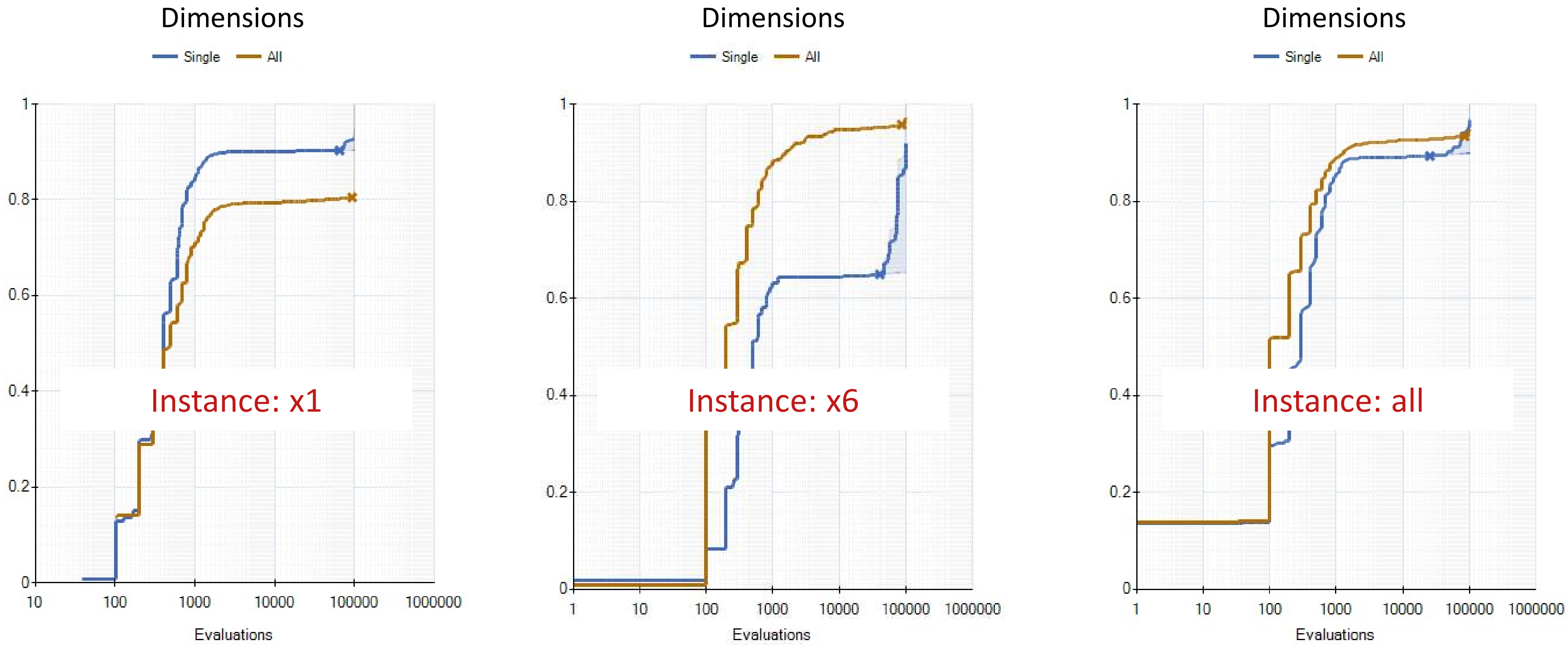
Parameter Analysis – Search Space





UNIVERSITY OF APPLIED SCIENCES UPPER AUSTRIA

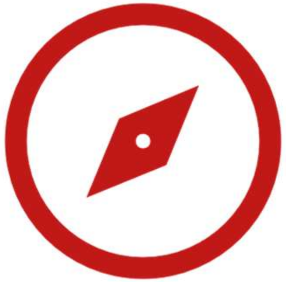
Parameter Analysis – 1-dim vs multi-dim



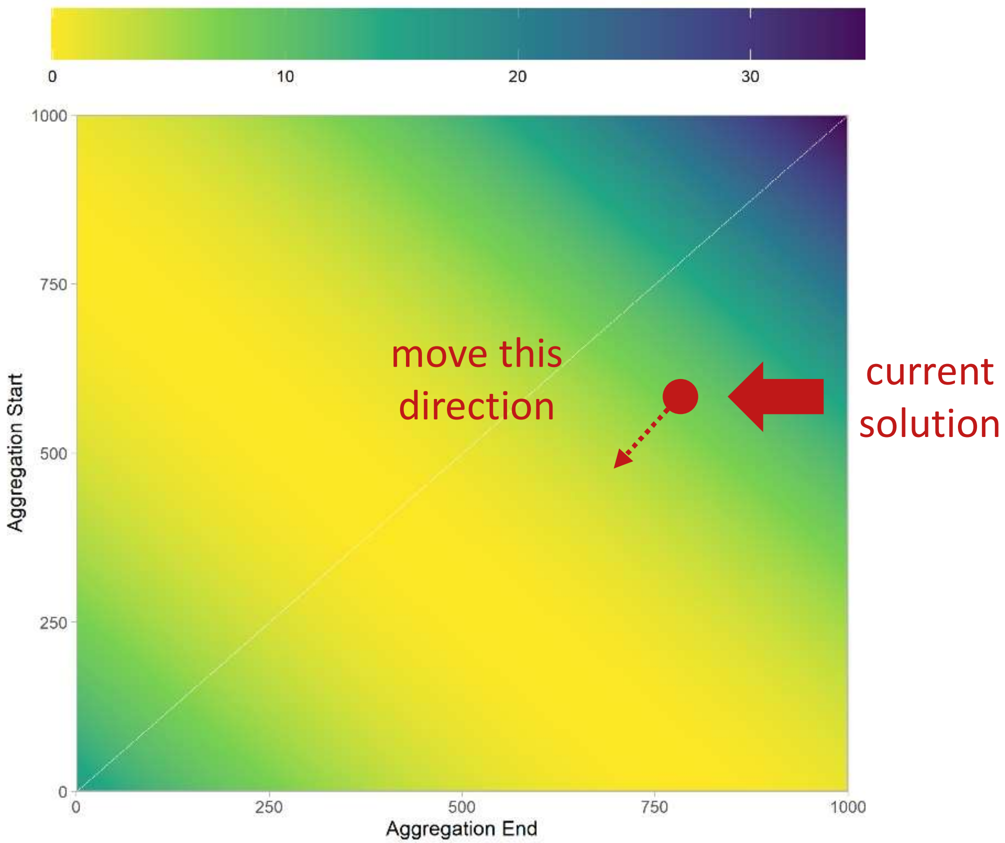
Summary Parameter Analysis

- **Iterative Random Sampling**
  - Fewer samples
    → faster convergence
  - 1-dim vs mutli-dim
    → instance dependent
  - Narrowing the search space
    → no impact
    → slower convergence

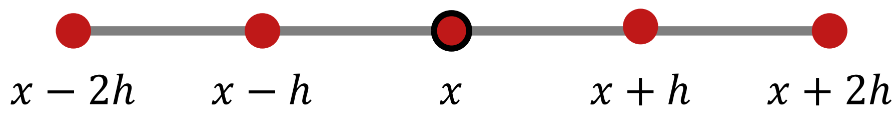
- **Next: Iterative guided sampling**
  - Using an approximate gradient



Gradient-Based Optimization



- **Objective function**
  - $f(x)$ … mean squared error for single dimension (start or end)
- **Five-point stencil**



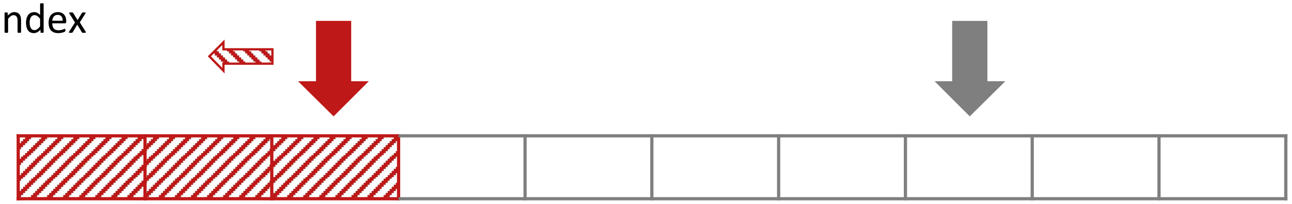
  - $f'(x) \approx \frac{f(x-2h)- \quad (x-h)+8f(x+h)-f(x+2h)}{12h}$
  - $h = 1$

UNIVERSITY OF APPLIED SCIENCES UPPER AUSTRIA

Random Direction → Guided Direction

- **Positive slope**

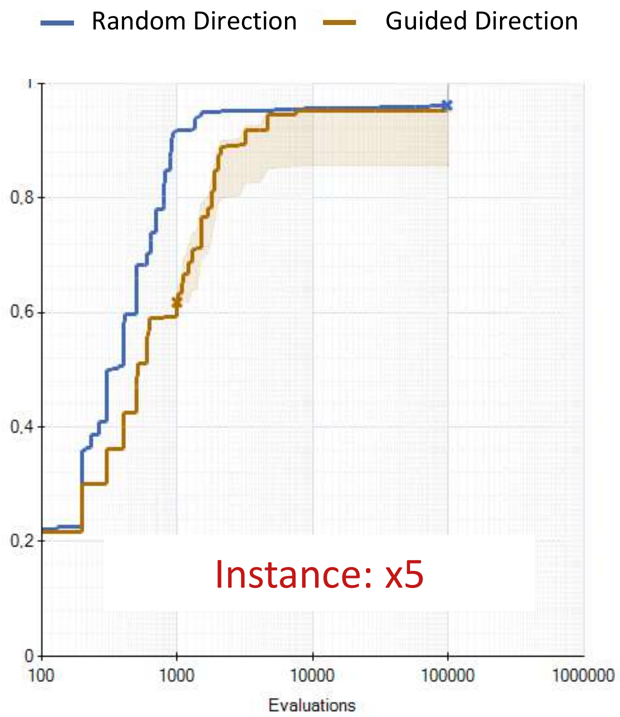
  → Try lower index

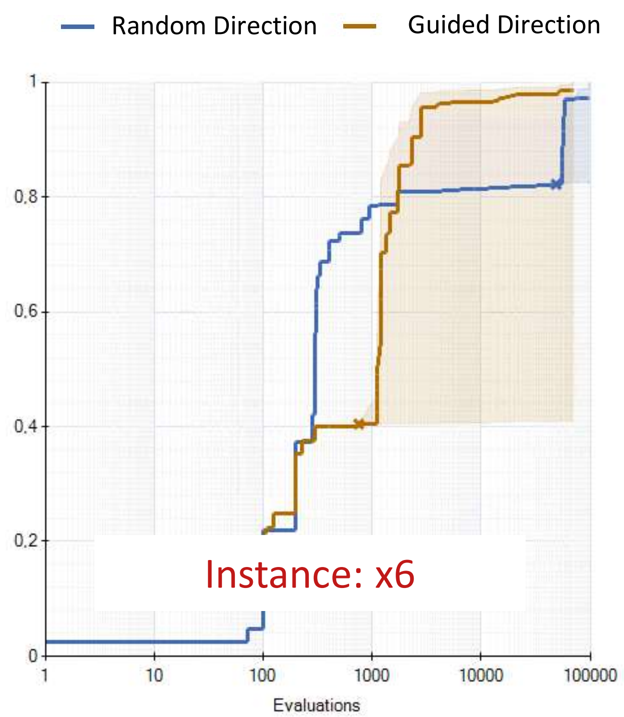


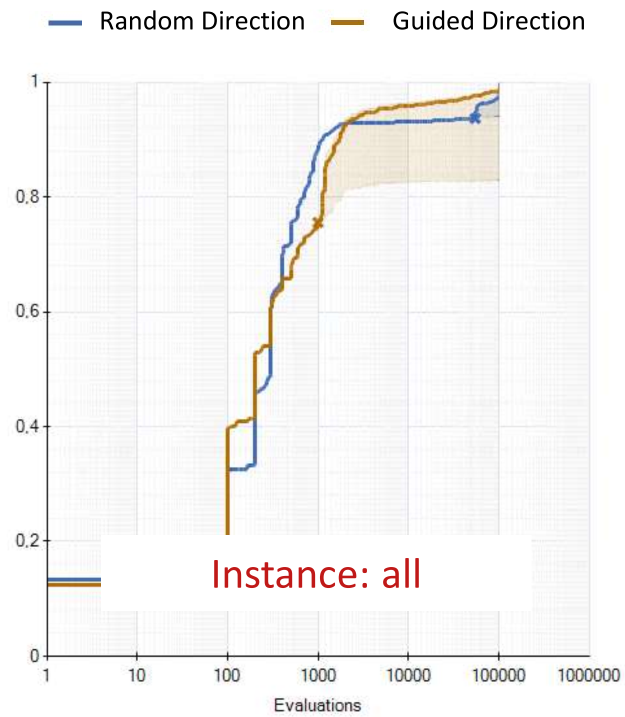
- **Negative slope**

  → Try higher index

Parameter Analysis – Guided Direction







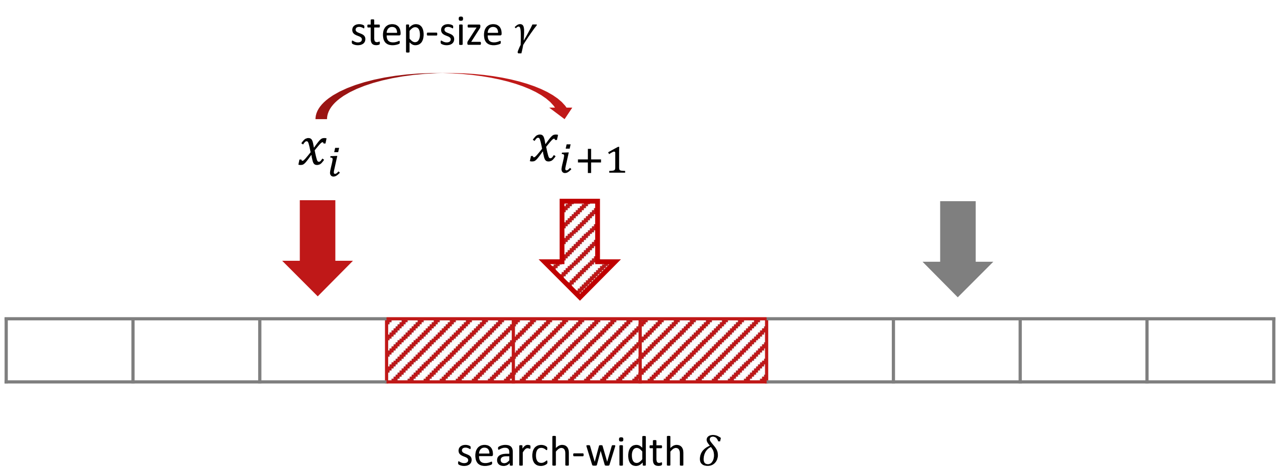
UNIVERSITY OF APPLIED SCIENCES UPPER AUSTRIA

Random Range → Guided Range

1. Go to target x_{i+1}
2. Search area around target

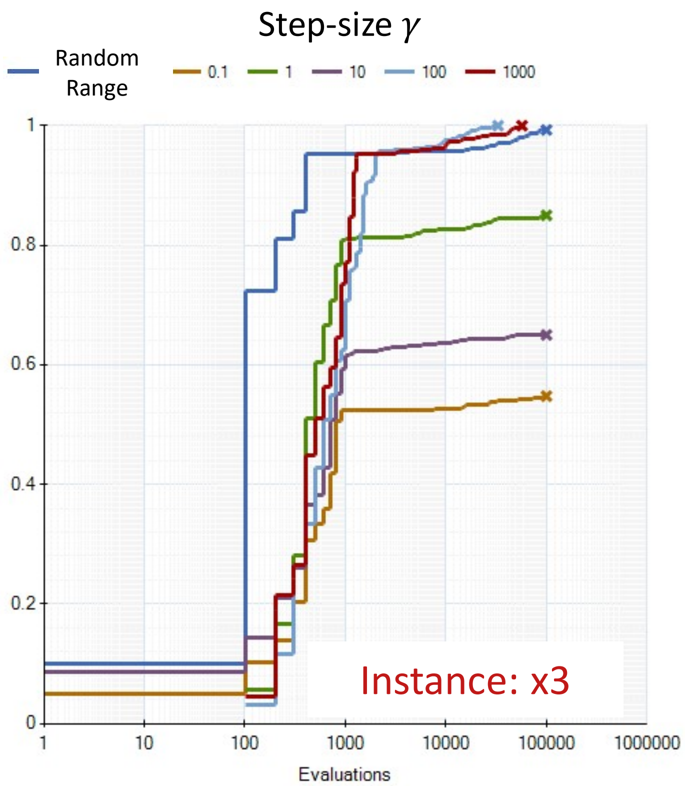
$$x_{i+1} = x_{\mathrm{i}} - \mathrm{round}\big(\gamma \cdot \nabla f(x)\big)$$

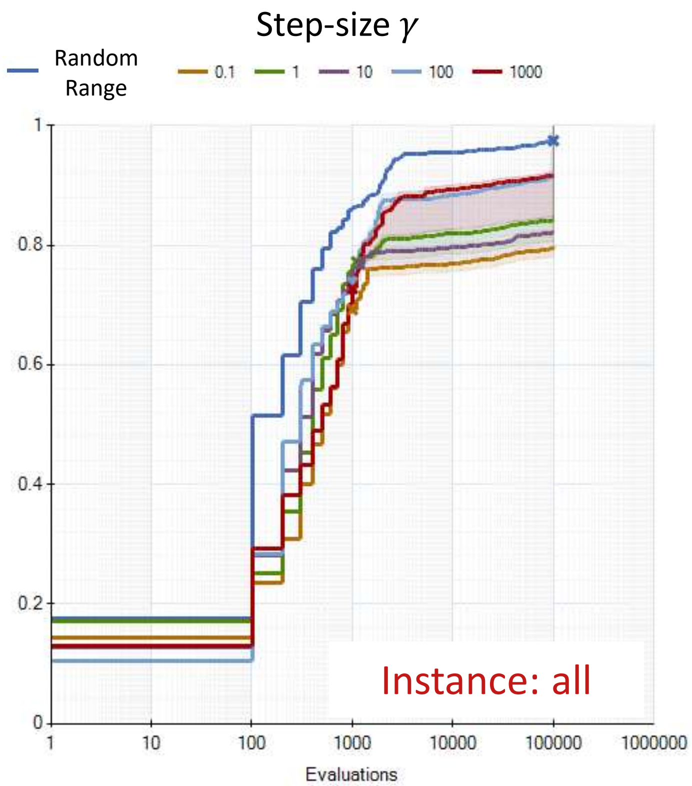
step-size



UNIVERSITY OF APPLIED SCIENCES UPPER AUSTRIA

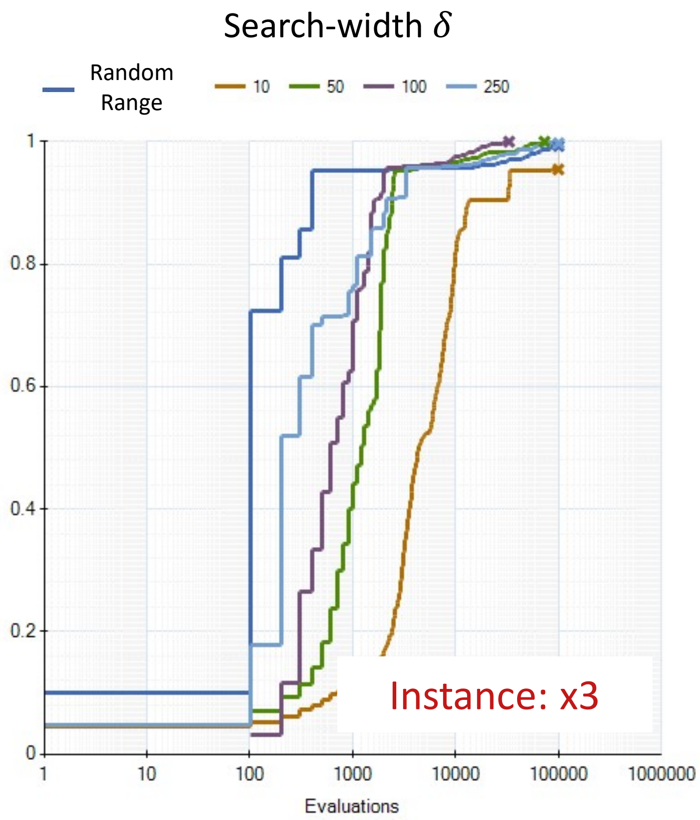
Parameter Analysis – Step-size

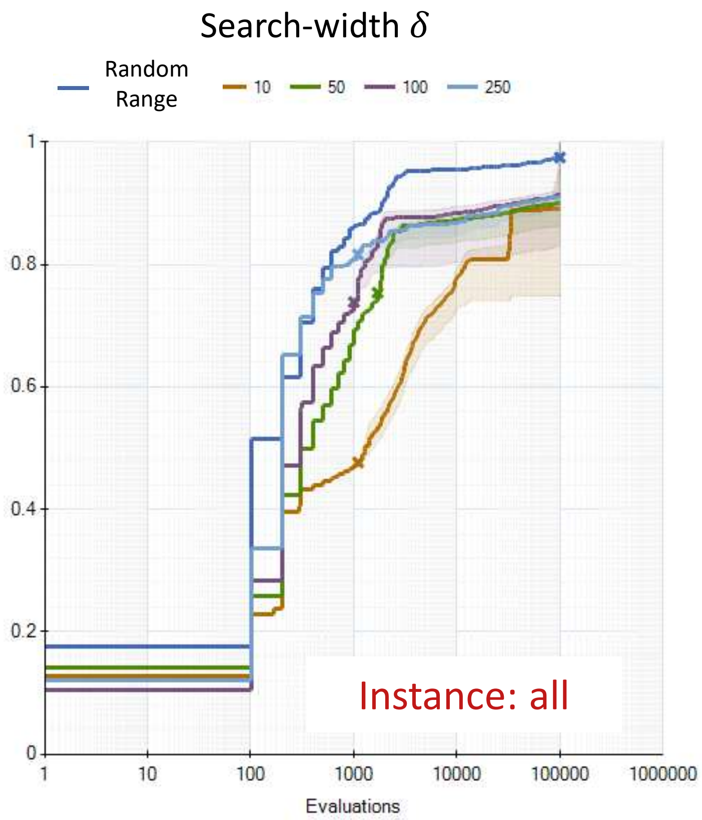




UNIVERSITY OF APPLIED SCIENCES UPPER AUSTRIA

Parameter Analysis – Search-width





Summary Parameter Analysis

- **Iterative Random Sampling**
  - Fewer samples
    → faster convergence ✓
  - 1-dim vs mutli-dim
    → instance dependent ✓
  - Narrowing the search space
    → no impact
    → slower convergence ?

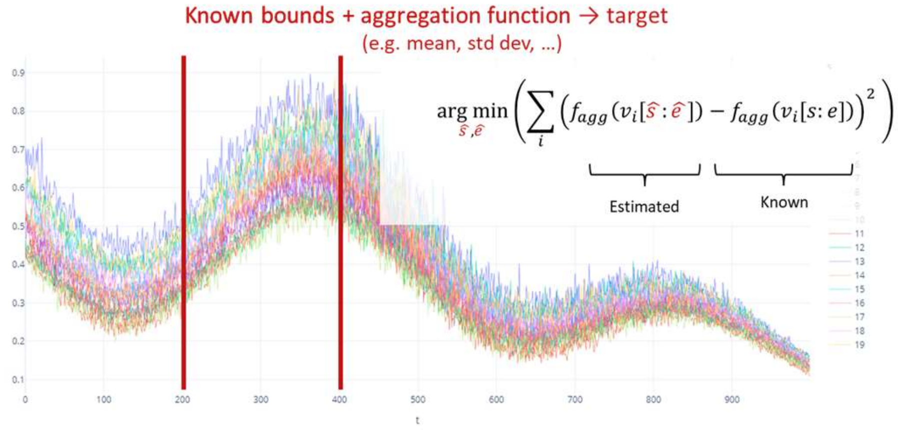
- **Iterative Guided Sampling**
  - Guided direction
    → instance dependent
    → no real improvement ?
  - Guided range
    → minimum step-size required ✓
    → slower convergence than random range ?

  › Stuck in local optima?

Conclusion and Way Forward

- **Seems very simple on paper**

- **Surprising behavior**
  - Narrowing the search space → no effect
  - Guide the search → slower convergence

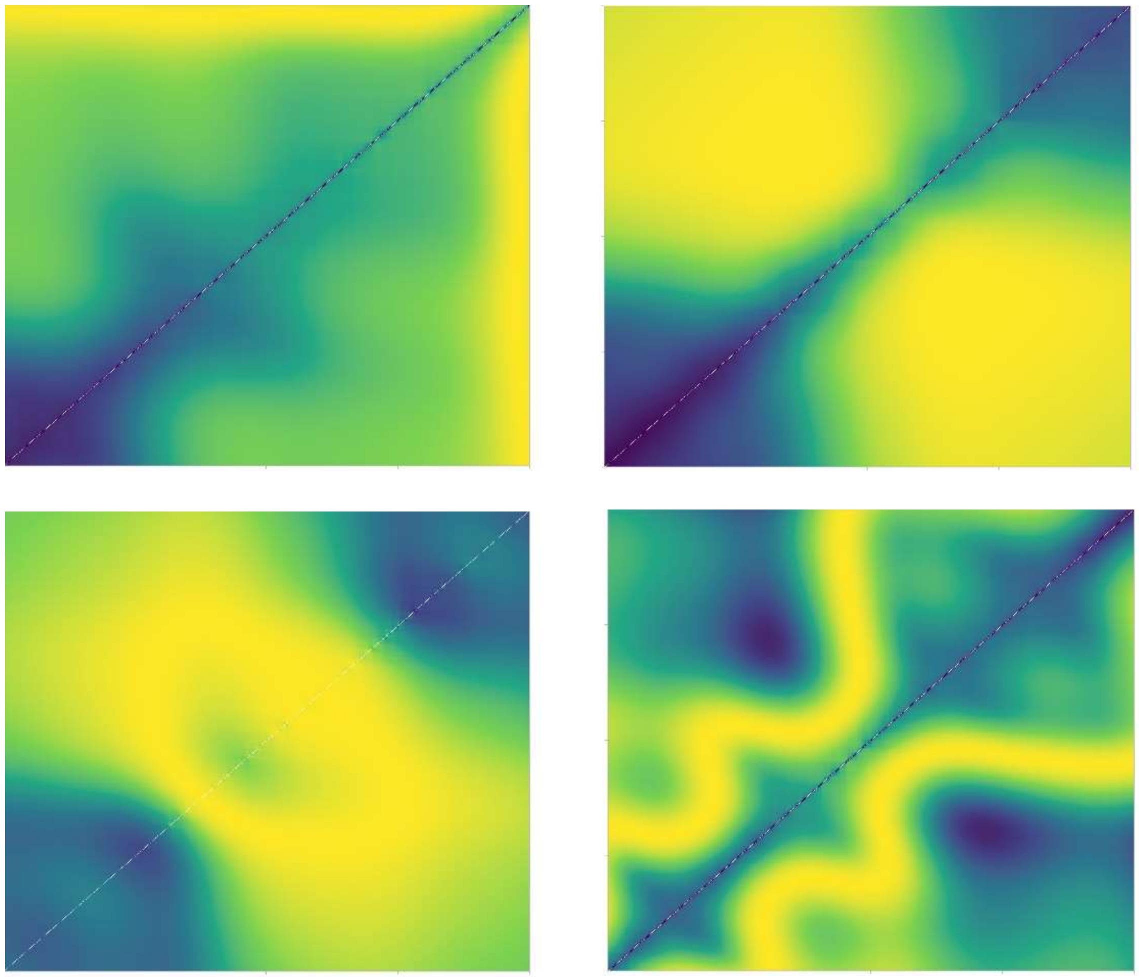


- **Segment optimization in the context of symbolic regression**
  - Will it behave the same?

UNIVERSITY OF APPLIED SCIENCES UPPER AUSTRIA

~~Questions?~~

Answers!



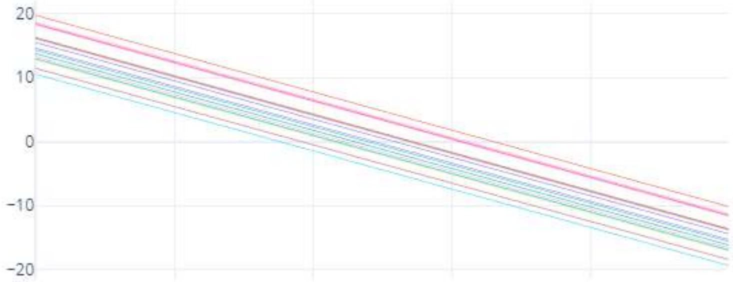
Segment Optimization Problem

- **Given**
  - List of vectors v_i
  - Known aggregation-function f_{agg}
  - Known aggregation indices i and j
- **Find**
  - Aggregation indices $\hat{i}$ and $\hat{j}$
- **So that**
  - The difference between the aggregated values using $\hat{i}$ and $\hat{j}$ and the known indices i and j is minimal

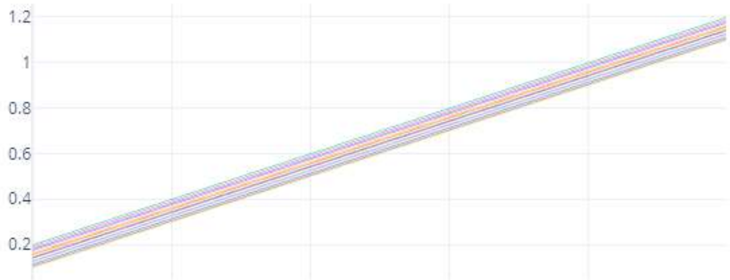
$$\underset{\hat{i},\hat{j}}{\arg\min}\left(\sum_i \Big(\underbrace{f_{agg}(v_i[\hat{i}:\hat{j}])}_{\text{Estimated}} - \underbrace{f_{agg}(v_i[i:j])}_{\text{Known}}\Big)^2\right)$$

Random Vectors

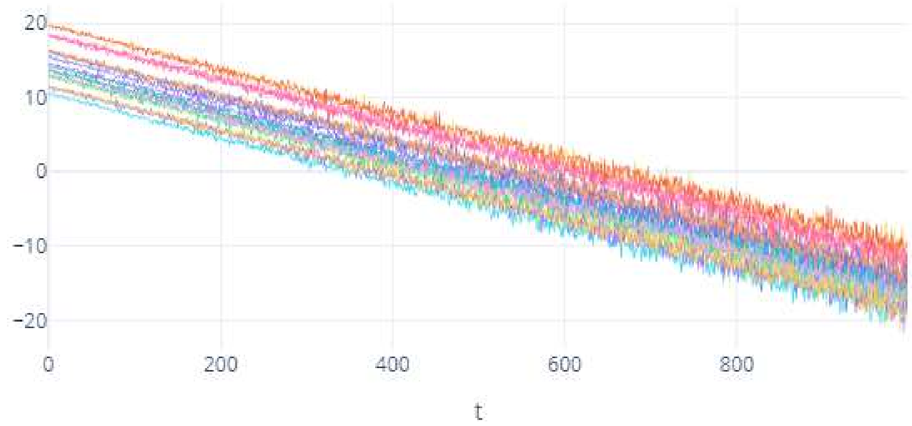
Mean



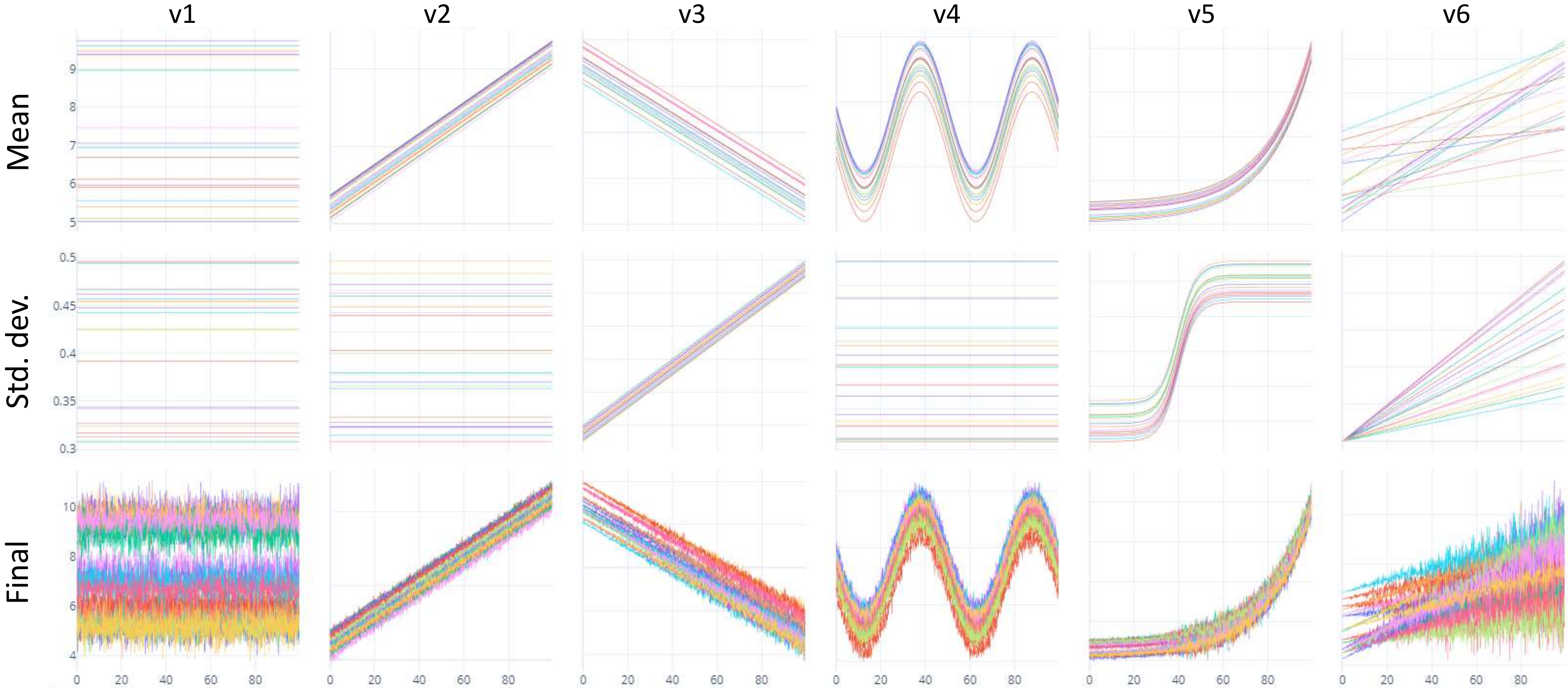
Std. dev.



Final

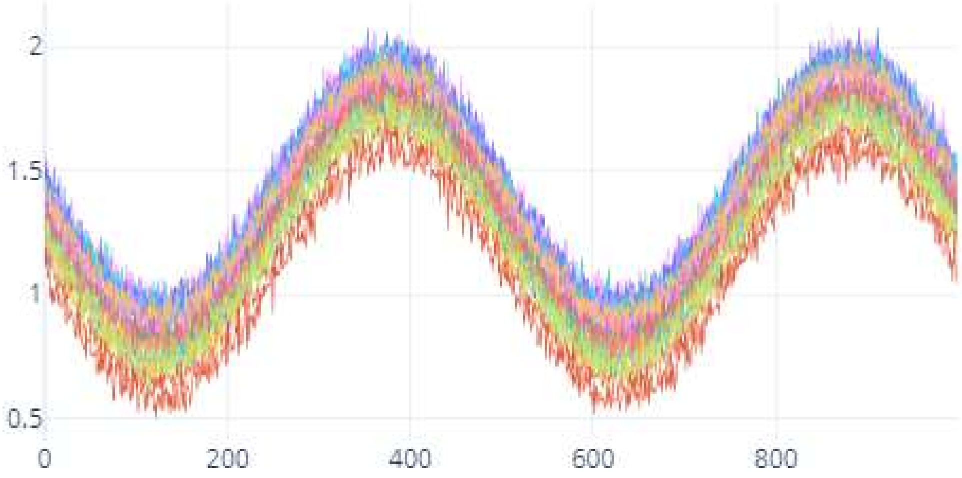


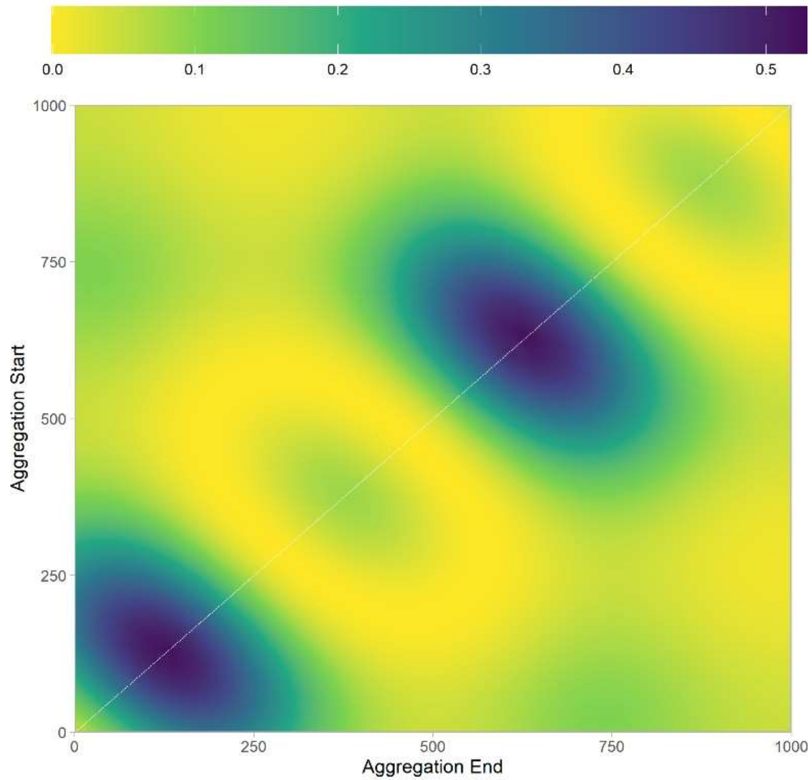
Random Vector Problem Instances



Segment Optimization Fitness Landscape

- **v4 instance**





UNIVERSITY OF APPLIED SCIENCES UPPER AUSTRIA

Experiment Configuration

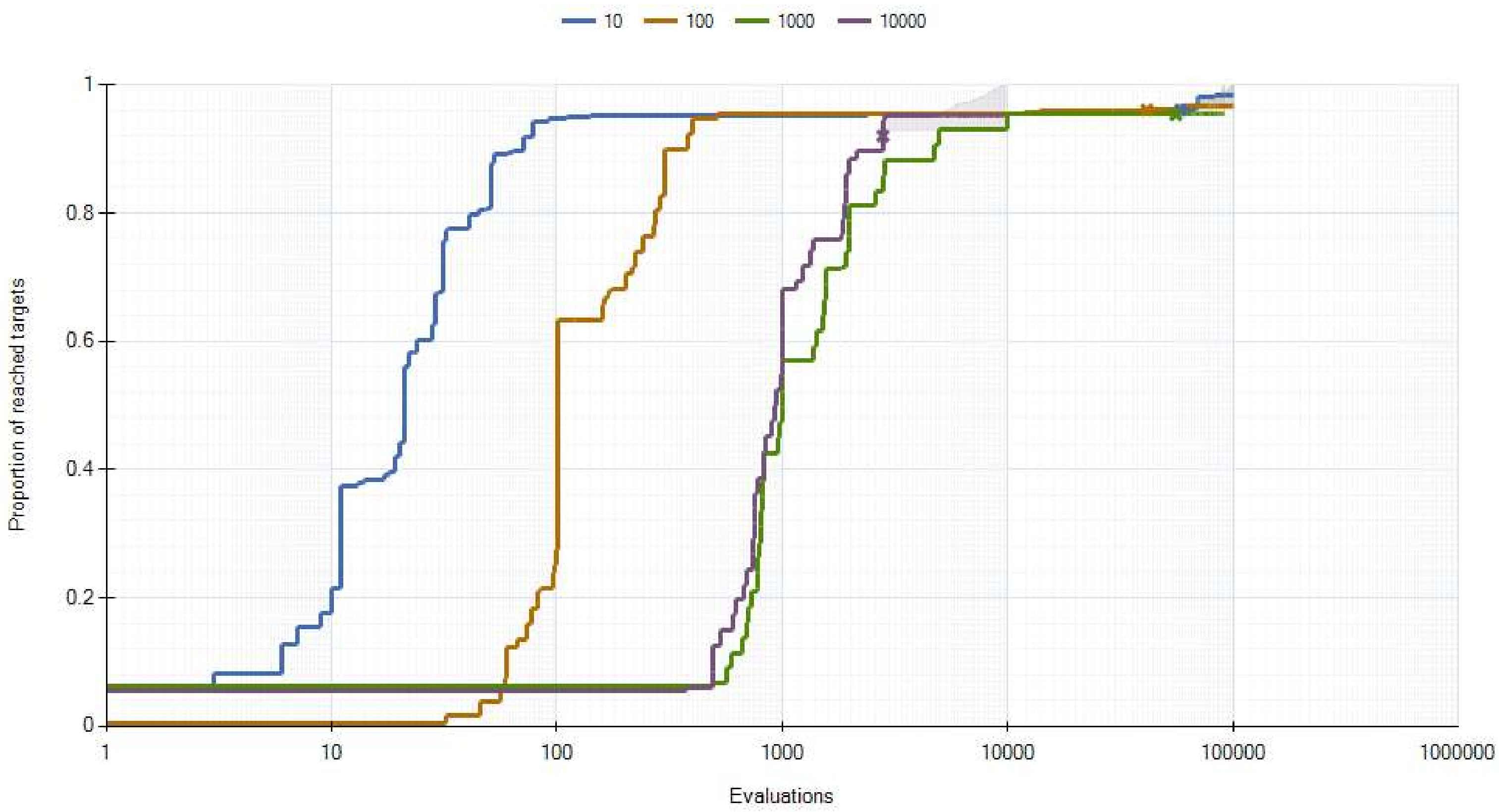
- **1,1 – Evolution Strategy**
  - 100000 maximum evaluations/samples
    - Depending on sample count
  - Mutation = **select best sample**
- **20 Repetitions**
- **6 Instances**
  - Vector length 1000
- **Parameter Variations**
  - Dimension
    - 1-dim, 2-dim
  - Search range
    - Full, random direction, random range
  - Sample selection
    - Exhaustive, random, linear
  - Sample count
    - 1, 10, 100, 1000, 10000
- **Total ~ 50 variations**
  - Some combinations are excluded
    - E.g. sample count for exhaustive

Parameter Analysis

Fixed parameters

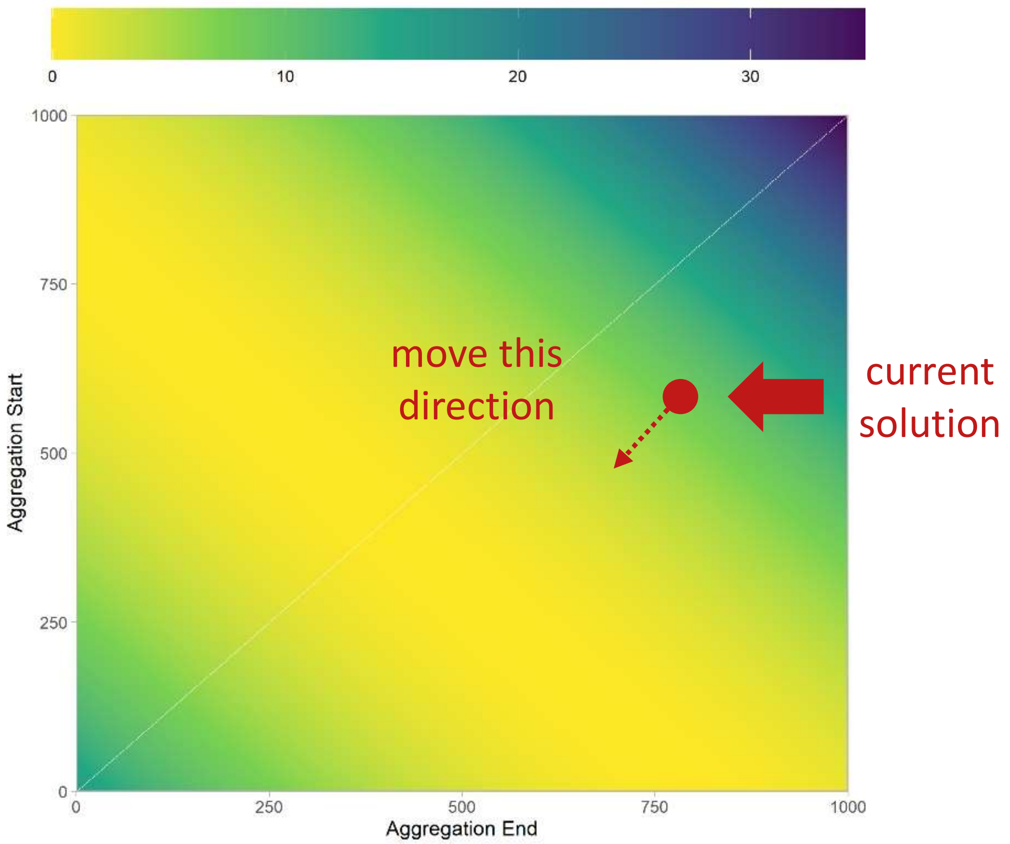
- Dimension
  - 1-dim
- Search Range
  - Full
- Sample Selection
  - Random

Sample Count



UNIVERSITY OF APPLIED SCIENCES UPPER AUSTRIA

Gradient-Based Optimization



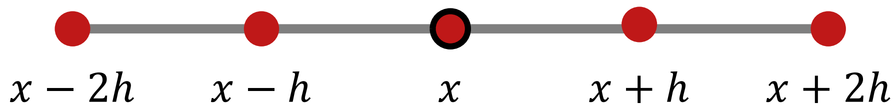
- **Numerical gradient**

$$f'(x) = \lim_{h \to 0} \frac{f(x+h) - f(x)}{h}$$

  - In reality
    - Small $h \in \mathbb{R}$
  - Integer indices
    - Small $h \in \mathbb{N}$

Five-Point Stencil

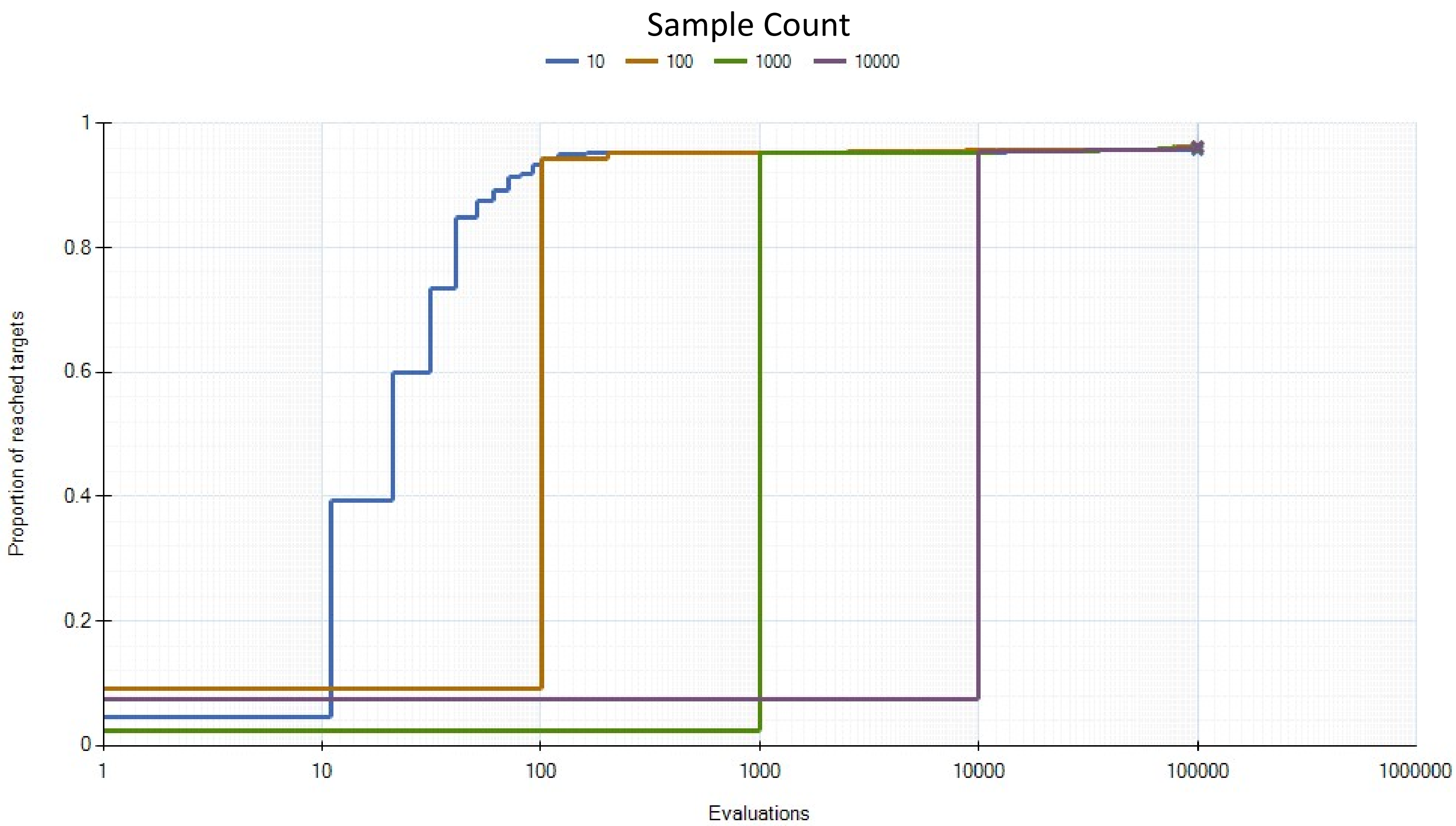
- **Segment optimization**
  - $f(x)$ … mean squared error
  - $h = 1$



$$f'(x) \approx \frac{+f(x-2h) - 8f(x-h) + 8f(x+h) - f(x+2h)}{12h}$$

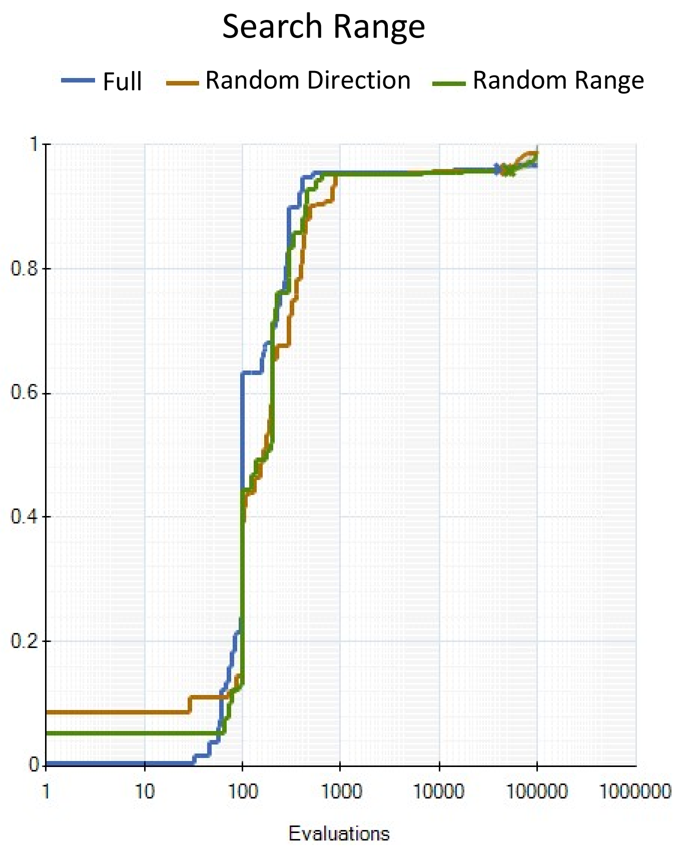
Parameter Analysis – Sample Count

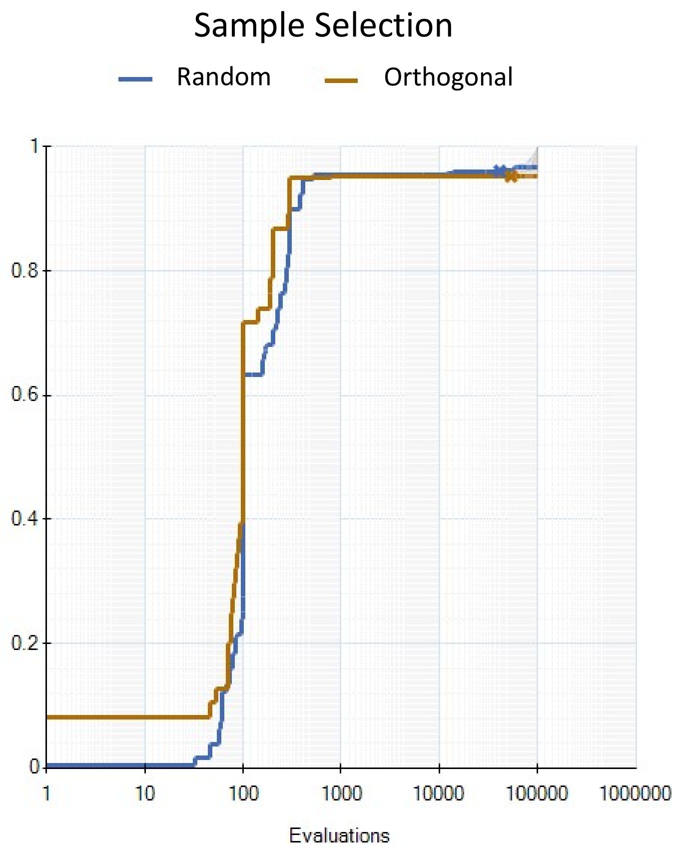
- Fixed parameters
  - Dimension = 2-dim
  - Search Range = Full
  - Sample Selection = Random



Parameter Analysis – Search Space Configuration

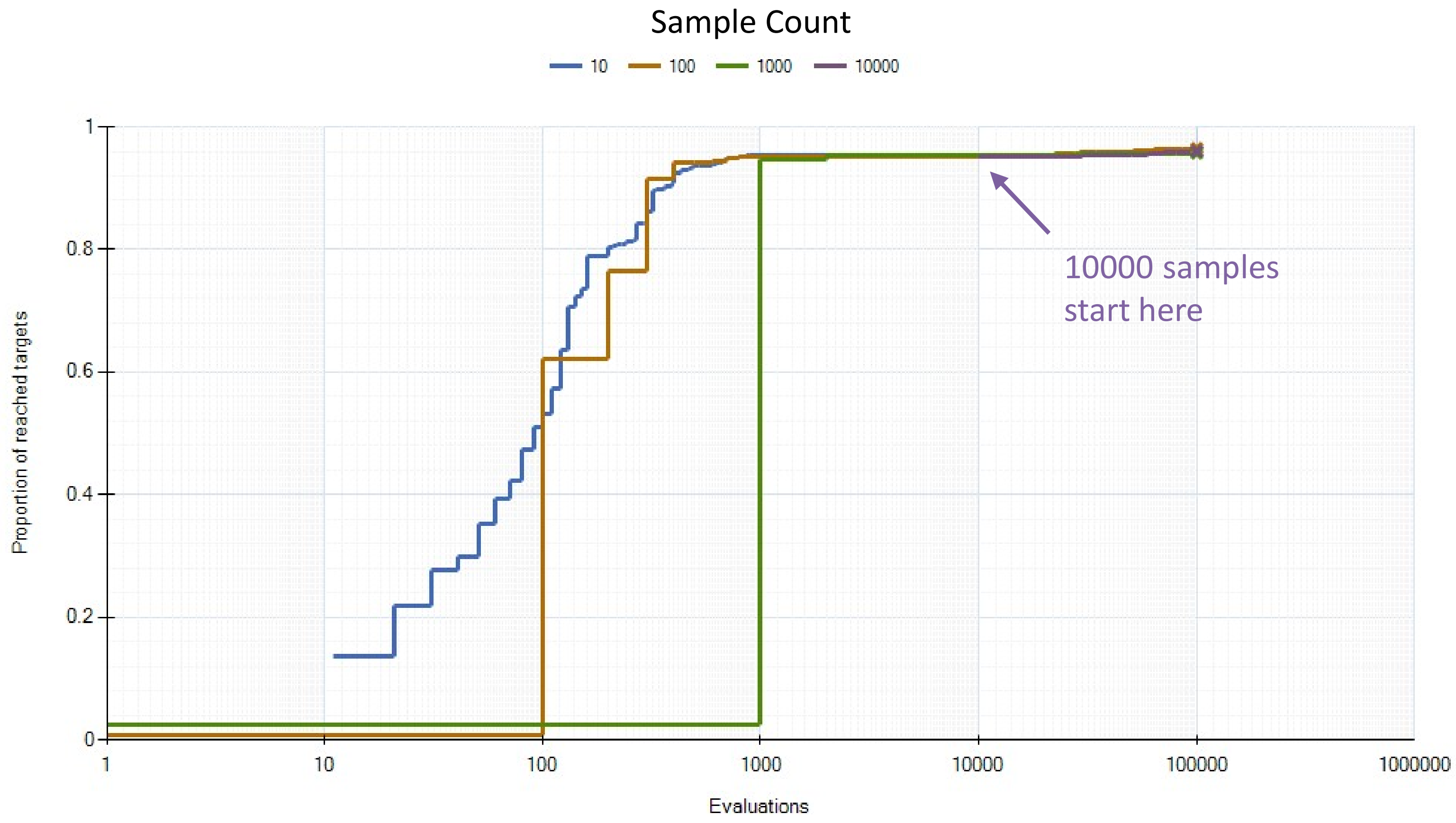
- Fixed parameters
  - Dimension = 1-dim
  - Sample Count = 100
  - Search Range = Full
  - Sample Selection = Random





Parameter Analysis – Sample Count

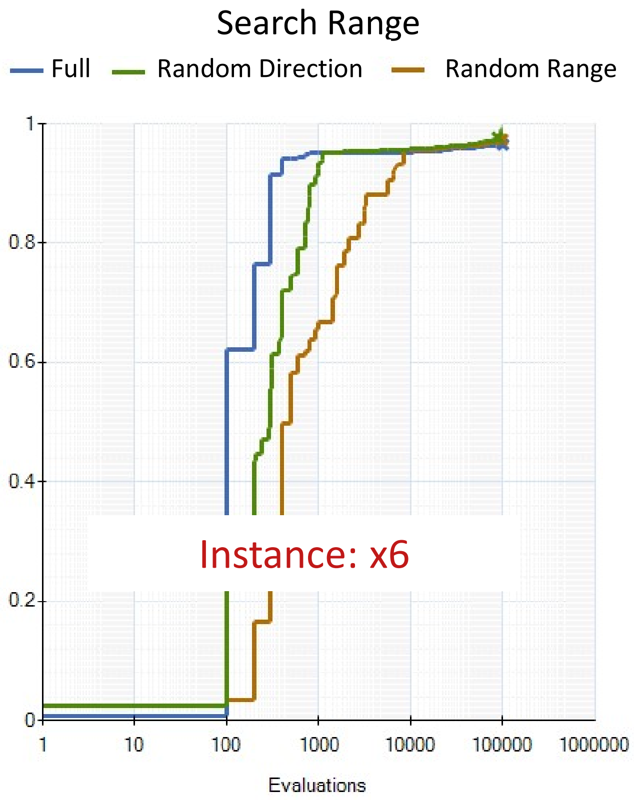
- Fixed parameters
  - Dimension = 2-dim
  - Search Range = Full
  - Sample Selection = Random

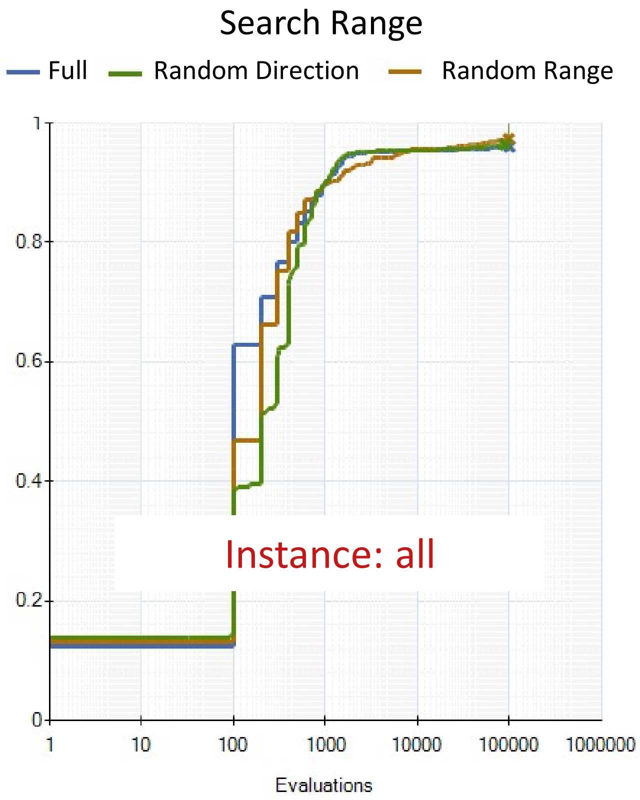


UNIVERSITY OF APPLIED SCIENCES UPPER AUSTRIA

Parameter Analysis – Search Space

- Fixed parameters
  - Dimension = 2-dim
  - Sample selection = Random
  - Sample count = 100

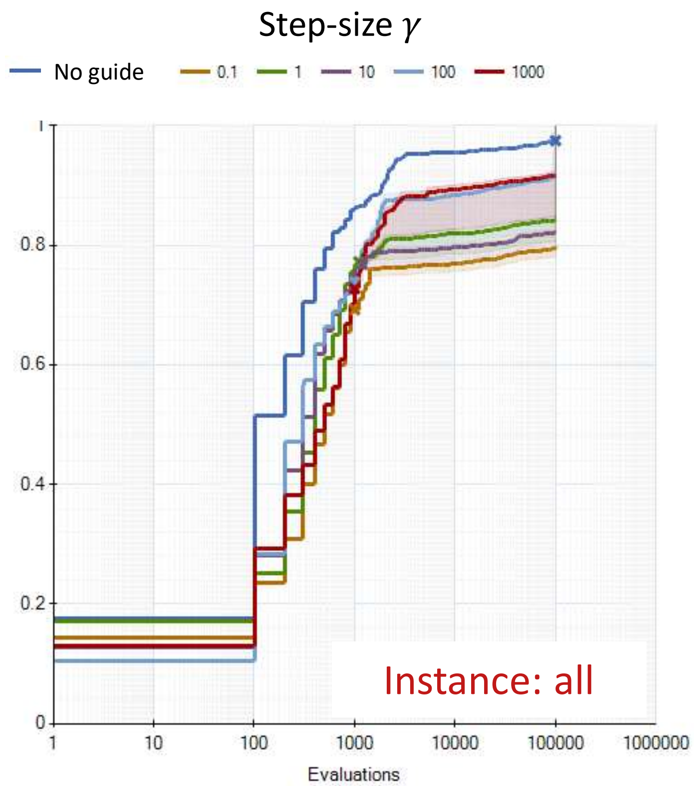




UNIVERSITY OF APPLIED SCIENCES UPPER AUSTRIA

Parameter Analysis – Step-size

- Fixed parameters
  - Dimension = 2-dim
  - Sample selection = Random
  - Sample count = 100
  - Search-width δ = 100

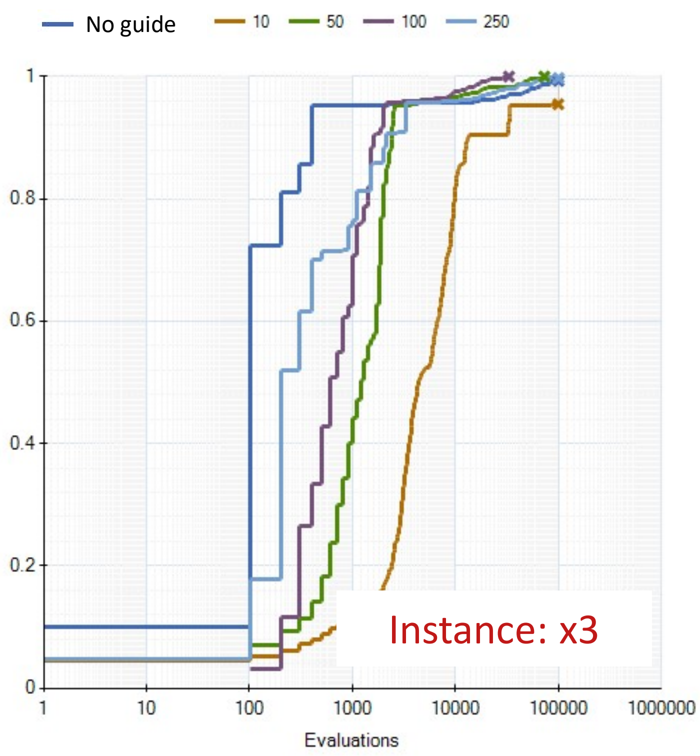


UNIVERSITY OF APPLIED SCIENCES UPPER AUSTRIA

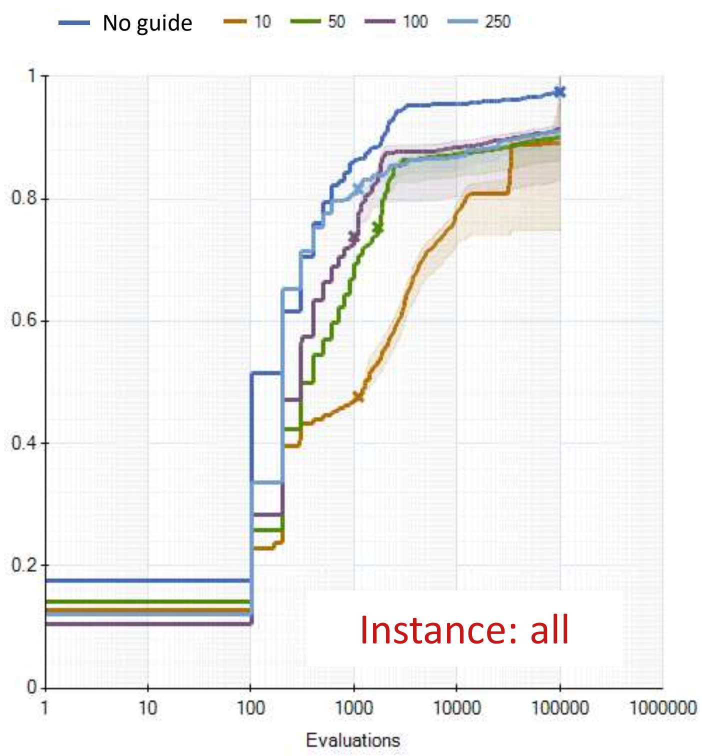
Parameter Analysis – Search-width

- Fixed parameters
  - Dimension = 2-dim
  - Sample selection = Random
  - Sample count = 100
  - Step-size γ = 100

Search-width δ



Search-width δ



UNIVERSITY OF APPLIED SCIENCES UPPER AUSTRIA