Evolutionary Diversity Optimization Using Multi-Objective Indicators

Aneta Neumann, Wanru Gao, Markus Wagner, Frank Neumann
Optimisation and Logistics
School of Computer Science
The University of Adelaide, Australia

GECCO Australia

Motivation

- Diversity plays a crucial role in evolutionary computation
- Diversity
  - prevents premature convergence
  - enables successful recombination/crossover
  - allows to compute set of Pareto optimal solutions for multi-objective problems

Diversity

- Majority of approaches consider diversity in the objective space.
- Ulrich/Thiele considered diversity in the search space (Tamara Ulrich's PhD thesis).
- Diversity with respect to other properties (features) is useful in various domains.

Goal:

- Compute a set of good solutions that differ in terms of interesting properties/features.
  - Think of good designs that vary with respect to important properties.

Application Areas

- Present set of diverse high-quality solutions (instead of single one) to enable discussion for further refinement.
- See how good solutions distribute with respect to important features of solutions
- Understanding of algorithm performance with respect to important features through diverse problem instances
- Construction of diverse set of problem instances for algorithm selection.

Diversity of instances for TSP

- We want to construct a diverse set of TSP instances

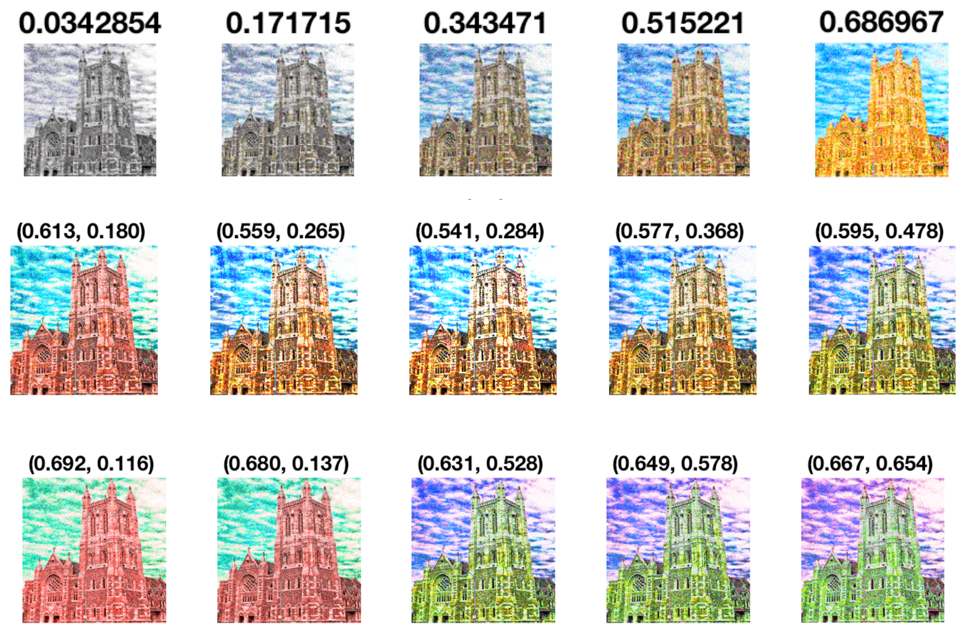
Examples:

- Diverse set where a certain algorithm is performing badly (high approximation ratio) $\alpha_A(I) = A(I)/OPT(I)$
- Diverse set where two solvers are performing differently.

Diversity of Images

- Evolve a diverse set of images that are close to a given image.
- Close means:

RMSE to given image is less than 10.
On the right: either 1 feature or a linear combination of two features as targets
[Alexander, Kortman, A. Neumann, GECCO'17]



Multiple features

- For 2 or more features, weightening of diversity contributions might not lead to good diversity.
- Results depend on chosen weightening.

Questions:

- What is a good diversity measure?
- What is the diversity optimisation goal?

Indicator-based Multi-Objective Optimization

- Let I be a search point
  - $f: X \rightarrow R^d$ a function that assigns to each search point I an objective vector
  - $q: X \rightarrow R^e$ be a function measures constraint violations
- An indicator Ind: $2^X \rightarrow R$ measures the quality of a given set of search points.

Indicator-Based Diversity Optimisation

- Let I be a search point
  - f: $X \rightarrow R^d$ a function that assigns to each search point a feature vector
  - q: $X \rightarrow R$ be a function assigning a quality score to each $I \in X$
    e.g.: require $q(I) \geq \alpha$ for all "good" solutions (constraint)
- Define Ind: $2^X \rightarrow R$ which measures the diversity of a given set of search points.

Goal:

Compute set $P=\{I_1, \ldots, I_\mu\}$ of μ solutions maximizing (minimizing) Ind among all sets of μ solutions under the condition that $q(I) \geq \alpha$ holds for all $I \in P$, where α is a given quality threshold.

Multi-Objective Indicators

Popular indicators in multi-objective optimization:

- Hypervolume (HYP)

$$HYP(S, r) = VOL\left(\cup_{(s_1,\ldots,s_d)\in S} [r_1, s_1] \times \cdots [r_d, s_d]\right)$$

- Inverted generational distance (IGD) (with respect to reference set R)

$$IGD(R, S) = \frac{1}{|R|} \sum_{r\in R} \min_{s\in S} d(r, s),$$

- Additive epsilon approximation (EPS) (with respect to reference set R)

$$\alpha(R, S) := \max_{r\in R} \min_{s\in S} \max_{1\leq i\leq d} (s_i - r_i).$$

How to use Multi-Objective Indicators

- Diversity Optimisation aims to compute a diverse set of solutions for a given single-objective problem
- Multi-Objective indicators guide the search towards a diverse set of Pareto optimal solutions.

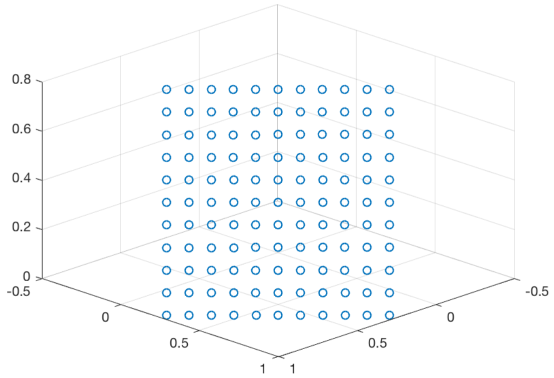
Use of multi-objective indicators:

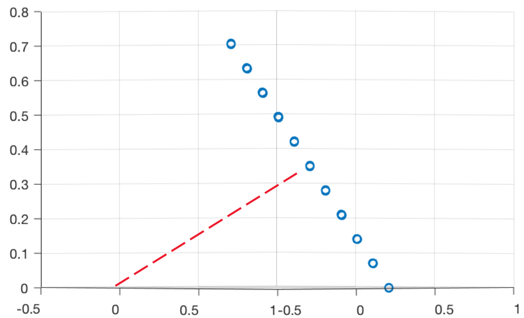
- Transform feature vectors of search points to make them incomparable.
- Apply multi-objective indicators after transformation has occurred.

Transformations (1/2)

For 2 features (transform into 3D) as follows:

- Place the unit square with its original x/y-coordinates in the three- dimensional space using z = 0.
- We rotate it around the x and y axis by 45 degrees each time.
- Translate it such that the centre point of the transformed unit square is at (sqrt(2)/4)



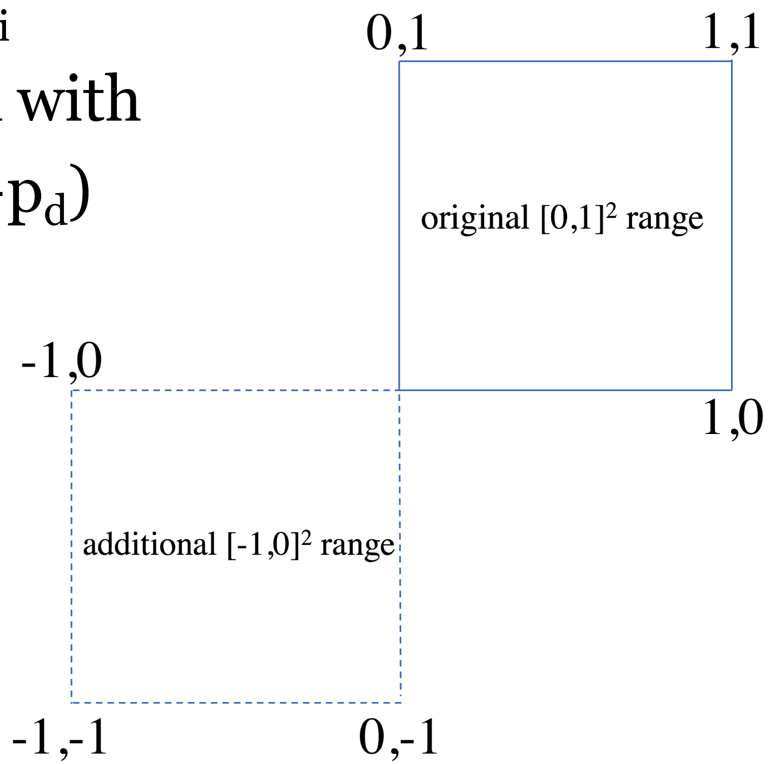


Transformations (2/2)

For d features:

- Double the number of dimensions to make vectors incomparable.
- For feature value p_i, use p_i and $-p_i$
- Instead of $p = (p_1, p_2, \ldots, p_d)$ work with

$$p' = (p_1, p_2, \ldots, p_d, -p_1, -p_2, \ldots, -p_d)$$



Which reference point in the (2*d)-dimensional space?
$(1^d, 0^d)$ would be based on the ranges' extreme values, and
$(2^d, 1^d)$ would put an increased focus on extreme points.

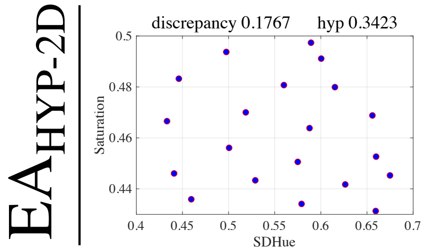
Algorithm

Algorithm 1: $(\mu + \lambda)$-EA_D

1 Initialize the population P with μ instances of quality at least α.
2 Let $C \subseteq P$ where $|C| = \lambda$.
3 For each $I \in C$, produce an offspring I' of I by mutation. If $q(I') \geqslant \alpha$, add I' to P.
4 While $|P| > \mu$, remove an individual with the smallest loss to the diversity indicator D.
5 Repeat step 2 to 4 until termination criterion is reached.

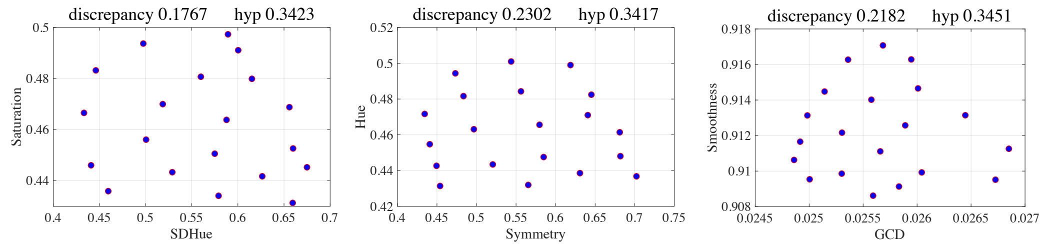
In plain English: it's a population-based EA that
(1) mutates lambda individuals in each generation, and
(2) considers diversity to select the survivors.

Multi-Objective Indicators (Images)

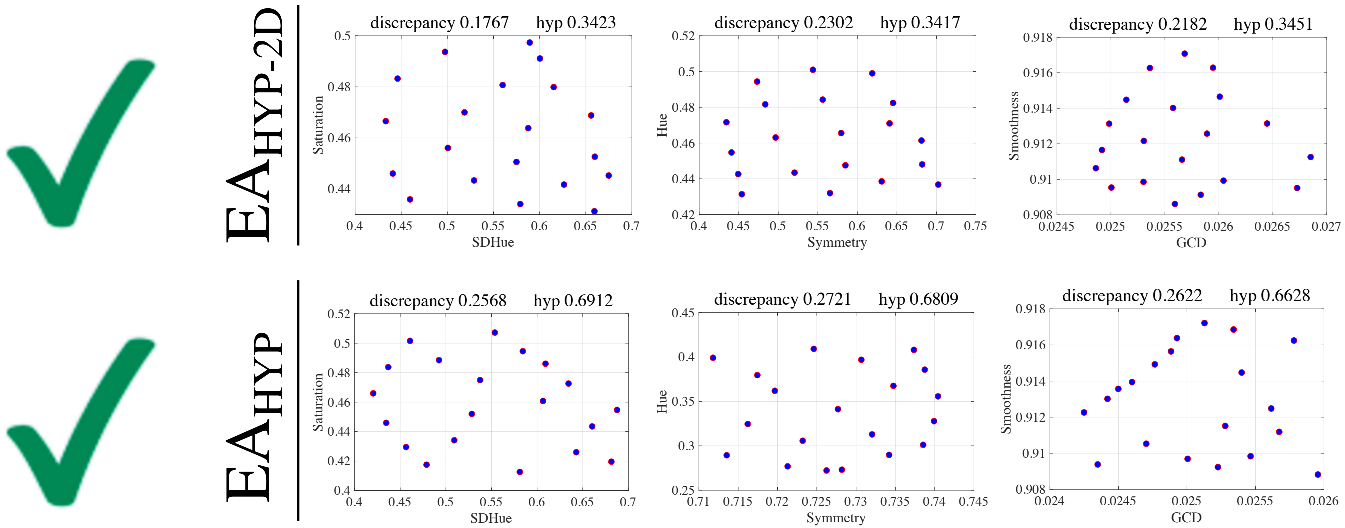


Multi-Objective Indicators (Images)

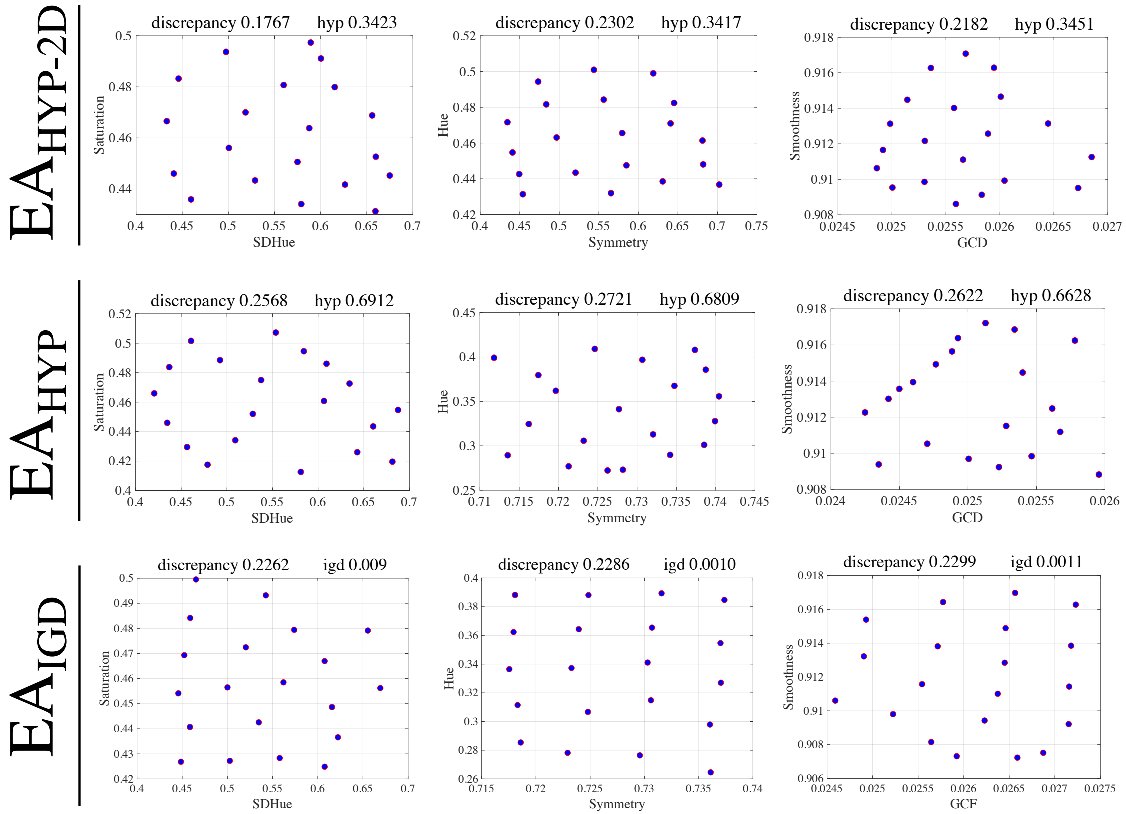
$EA_{HYP\text{-}2D}$



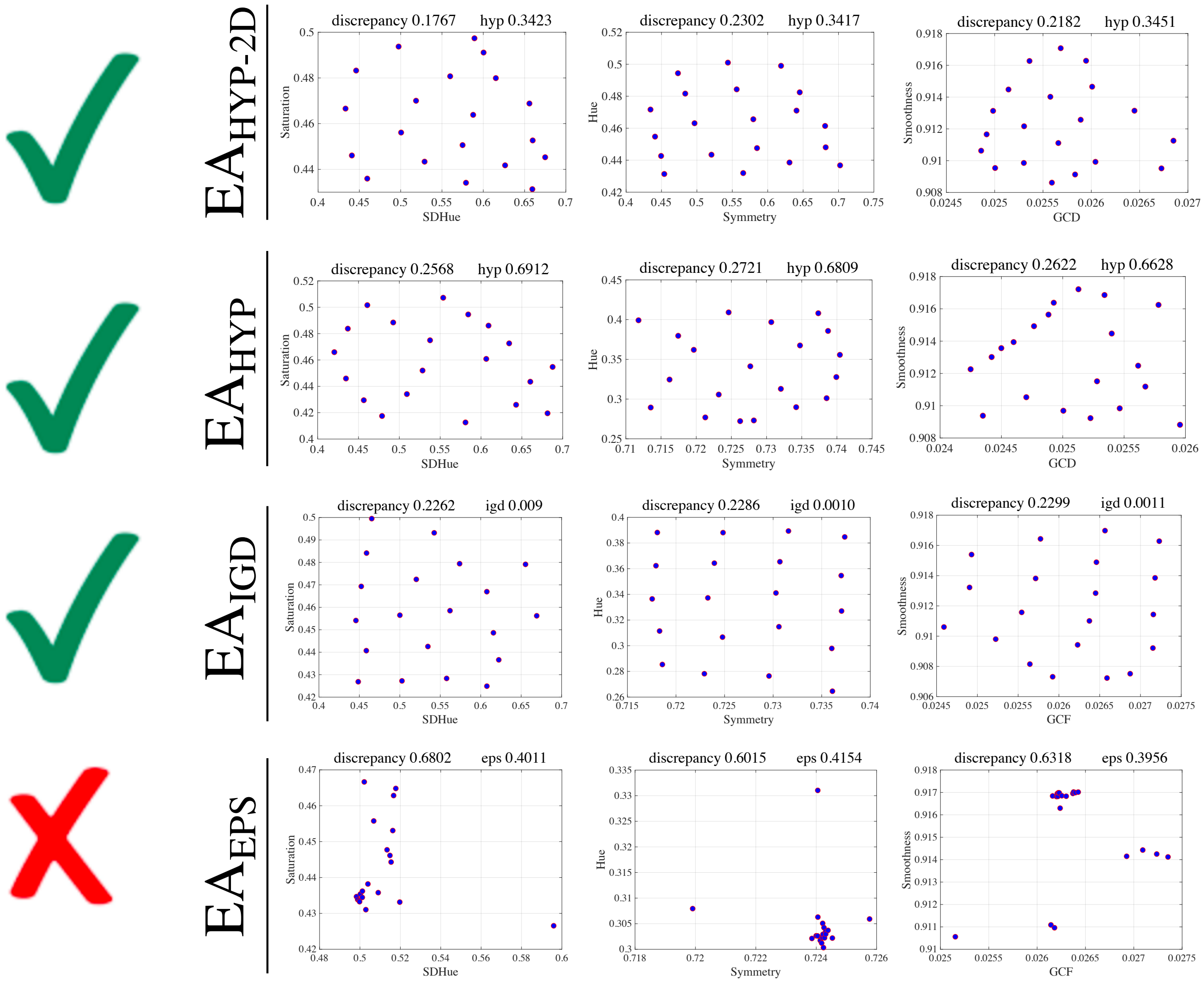
Multi-Objective Indicators (Images)



Multi-Objective Indicators (Images)

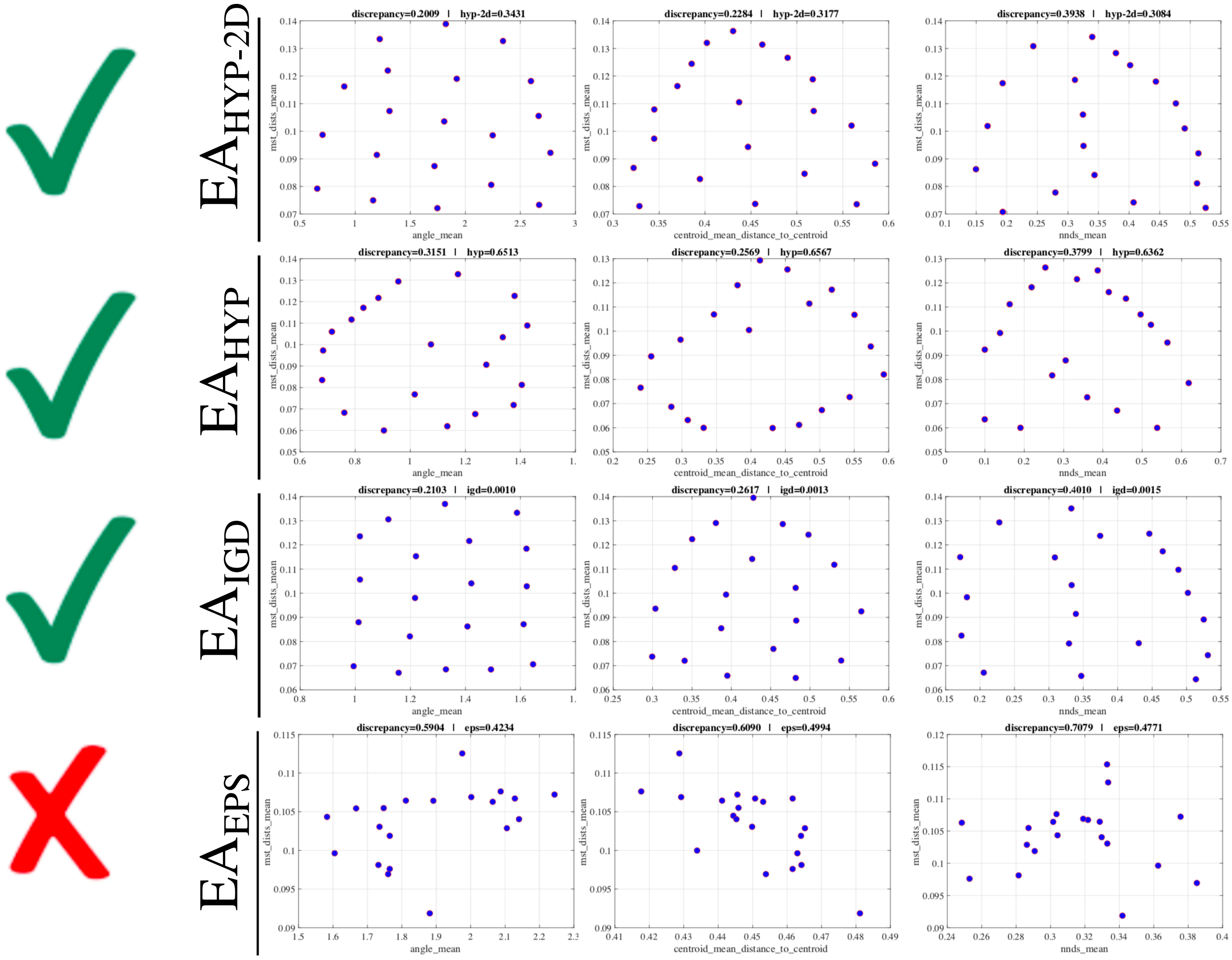


Multi-Objective Indicators (Images)



Not locally sensitive, even when using the vector of all ref.-grid approx.

Multi-Objective Indicators (TSP)



Not locally sensitive, even when using the vector of all ref.-grid approx.

Results TSP

2-feature combinations

		EA$_{\text{HYP-2D}}$ (1)			EA$_{\text{HYP}}$ (2)			EA$_{\text{IGD}}$ (3)			EA$_{\text{EPS}}$ (4)			EA$_{\text{DIS}}$ (5)		
		mean	st	stat	mean	st	stat	mean	st	stat	mean	st	stat	mean	st	stat
HYP-2D	f_1,f_4	0.338	2E-3	$2^{(+)},4^{(+)},5^{(+)}$	0.309	4E-3	$1^{(-)},4^{(+)}$	0.331	3E-3	$4^{(+)},5^{(+)}$	0.190	1E-3	$1^{(-)},2^{(-)},3^{(-)}$	0.256	1E-2	$1^{(-)},3^{(-)}$
	f_2,f_4	0.317	3E-3	$2^{(+)},4^{(+)},5^{(+)}$	0.303	5E-3	$1^{(-)},3^{(-)},4^{(+)}$	0.316	3E-3	$2^{(+)},4^{(+)},5^{(+)}$	0.178	1E-7	$1^{(-)},2^{(-)},3^{(-)}$	0.252	1E-2	$1^{(-)},3^{(-)}$
	f_3,f_4	0.303	2E-2	$2^{(+)},4^{(+)},5^{(+)}$	0.296	5E-3	$1^{(-)},3^{(-)},4^{(+)},5^{(+)}$	0.304	2E-2	$2^{(+)},4^{(+)},5^{(+)}$	0.190	2E-3	$1^{(-)},2^{(-)},3^{(-)}$	0.238	2E-2	$1^{(-)},2^{(-)},3^{(-)}$
HYP	f_1,f_4	0.645	5E-3	$4^{(+)},5^{(+)}$	0.638	7E-3	$4^{(+)},5^{(+)}$	0.639	6E-3	$4^{(+)},5^{(+)}$	0.424	2E-3	$1^{(-)},2^{(-)},3^{(-)}$	0.529	3E-2	$1^{(-)},2^{(-)},3^{(-)}$
	f_2,f_4	0.609	7E-3	$2^{(-)},4^{(+)},5^{(+)}$	0.632	1E-2	$1^{(+)},4^{(+)},5^{(+)}$	0.621	6E-3	$4^{(+)},5^{(+)}$	0.398	1E-6	$1^{(-)},2^{(-)},3^{(-)}$	0.505	2E-2	$1^{(-)},2^{(-)},3^{(-)}$
	f_3,f_4	0.584	3E-2	$2^{(-)},4^{(+)}$	0.621	9E-3	$1^{(+)},3^{(+)},4^{(+)},5^{(+)}$	0.595	4E-2	$2^{(-)},4^{(+)},5^{(+)}$	0.410	2E-3	$1^{(-)},2^{(-)},3^{(-)}$	0.485	3E-2	$2^{(-)},3^{(-)}$
IGD	f_1,f_4	0.001	2E-5	$4^{(+)},5^{(+)}$	0.001	6E-5	$3^{(-)},4^{(+)}$	0.001	4E-5	$2^{(+)},4^{(+)},5^{(+)}$	0.003	2E-5	$1^{(-)},2^{(-)},3^{(-)}$	0.002	2E-4	$1^{(-)},3^{(-)}$
	f_2,f_4	0.001	3E-5	$2^{(+)},4^{(+)},5^{(+)}$	0.002	6E-5	$1^{(-)},3^{(-)},4^{(+)}$	0.001	3E-5	$2^{(+)},4^{(+)},5^{(+)}$	0.003	2E-10	$1^{(-)},2^{(-)},3^{(-)}$	0.002	2E-4	$1^{(-)},3^{(-)}$
	f_3,f_4	0.002	3E-4	$4^{(+)},5^{(+)}$	0.002	6E-5	$3^{(-)},4^{(+)},5^{(+)}$	0.002	3E-4	$2^{(+)},4^{(+)},5^{(+)}$	0.003	3E-5	$1^{(-)},2^{(-)},3^{(-)}$	0.003	3E-4	$1^{(-)},2^{(-)},3^{(-)}$
EPS	f_1,f_4	0.196	2E-2	$2^{(+)},4^{(+)},5^{(+)}$	0.249	2E-2	$1^{(-)},3^{(-)},4^{(+)}$	0.189	2E-2	$2^{(+)},4^{(+)},5^{(+)}$	0.423	1E-3	$1^{(-)},2^{(-)},3^{(-)}$	0.345	4E-2	$1^{(-)},3^{(-)}$
	f_2,f_4	0.226	8E-3	$2^{(+)},4^{(+)},5^{(+)}$	0.256	2E-2	$1^{(-)},3^{(-)},4^{(+)},5^{(+)}$	0.228	1E-2	$2^{(+)},4^{(+)},5^{(+)}$	0.499	2E-16	$1^{(-)},2^{(-)},3^{(-)}$	0.360	5E-2	$1^{(-)},2^{(-)},3^{(-)}$
	f_3,f_4	0.260	4E-2	$4^{(+)},5^{(+)}$	0.278	2E-2	$4^{(+)},5^{(+)}$	0.265	4E-2	$4^{(+)},5^{(+)}$	0.477	3E-3	$1^{(-)},2^{(-)},3^{(-)}$	0.368	5E-2	$1^{(-)},2^{(-)},3^{(-)}$
DIS	f_1,f_4	0.222	2E-2	$2^{(+)},4^{(+)},5^{(+)}$	0.353	2E-2	$1^{(-)},3^{(-)},4^{(+)}$	0.249	2E-2	$2^{(+)},4^{(+)}$	0.589	4E-3	$1^{(-)},2^{(-)},3^{(-)},5^{(-)}$	0.292	5E-2	$1^{(-)},4^{(+)}$
	f_2,f_4	0.230	2E-2	$2^{(+)},4^{(+)},5^{(+)}$	0.274	2E-2	$1^{(-)},4^{(+)},5^{(+)}$	0.252	1E-3	$4^{(+)},5^{(+)}$	0.609	1E-16	$1^{(-)},2^{(-)},3^{(-)},5^{(-)}$	0.336	4E-2	$1^{(-)},2^{(-)},3^{(-)},4^{(+)}$
	f_3,f_4	0.418	6E-2	$4^{(+)}$	0.416	3E-2	$4^{(+)}$	0.401	7E-2	$4^{(+)},5^{(+)}$	0.719	6E-3	$1^{(-)},2^{(-)},3^{(-)},5^{(-)}$	0.448	9E-2	$3^{(-)},4^{(+)}$

3-feature combinations

		EA$_{\text{HYP}}$ (1)			EA$_{\text{IGD}}$ (2)			EA$_{\text{DIS}}$ (3)		
		mean	st	stat	mean	st	stat	mean	st	stat
HYP	f_1,f_2,f_3	0.4511	1E-2	$2^{(+)},3^{(+)}$	0.4261	7E-3	$1^{(-)},3^{(+)}$	0.3385	6E-3	$1^{(-)},2^{(-)}$
	f_1,f_3,f_4	0.4579	8E-3	$2^{(+)},3^{(+)}$	0.4260	6E-3	$1^{(-)},3^{(+)}$	0.3430	6E-3	$1^{(-)},2^{(-)}$
	f_2,f_3,f_4	0.4478	8E-3	$2^{(+)},3^{(+)}$	0.4262	6E-3	$1^{(-)},3^{(+)}$	0.3430	6E-3	$1^{(-)},2^{(-)}$
IGD	f_1,f_2,f_3	0.0083	3E-4	$2^{(-)},3^{(+)}$	0.0075	2E-4	$1^{(+)},3^{(+)}$	0.0110	1E-4	$1^{(-)},2^{(-)}$
	f_1,f_3,f_4	0.0082	2E-4	$2^{(-)},3^{(+)}$	0.0077	1E-4	$2^{(+)},3^{(+)}$	0.0107	1E-4	$1^{(-)},2^{(-)}$
	f_2,f_3,f_4	0.0086	2E-4	$2^{(-)},3^{(+)}$	0.0080	2E-2	$2^{(+)},3^{(+)}$	0.0112	8E-5	$1^{(-)},2^{(-)}$
DIS	f_1,f_2,f_3	0.4115	3E-2	$2^{(+)},3^{(+)}$	0.4839	3E-2	$1^{(-)},3^{(-)}$	0.4399	2E-2	$1^{(-)},2^{(+)}$
	f_1,f_3,f_4	0.5220	4E-2	$3^{(-)}$	0.5474	3E-2	$3^{(-)}$	0.4757	2E-2	$1^{(+)},2^{(+)}$
	f_2,f_3,f_4	0.4669	3E-2	$2^{(+)}$	0.5111	3E-2	$1^{(-)},3^{(-)}$	0.4667	2E-2	$2^{(+)}$

In summary:

- EA$_{\text{HYP}}$ and EA$_{\text{IGD}}$ perform best
- Beats our GECCO'18 results (discrepancy theory)

30 independent runs per setup

Kruskal-Wallis-Test: like a multi-set Mann-Whitney-U-Test

Summary

- We provide two simple and effective transformations to enable the diversity optimisation.

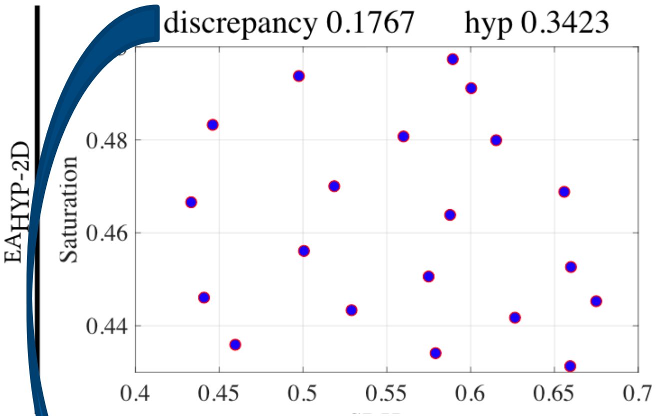
- We use of-the-shelf multi-objective performance indicators; HYP and IGD worked well.

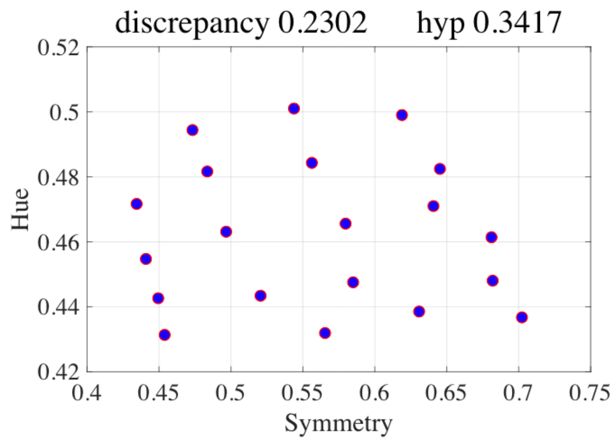
- We provide code: https://tinyurl.com/geccoDiversity (Java code, Matlab wrapper provided)

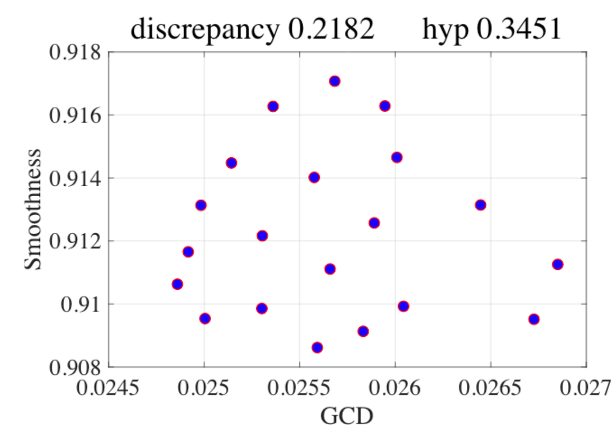
Email: markus.wagner@adelaide.edu.au

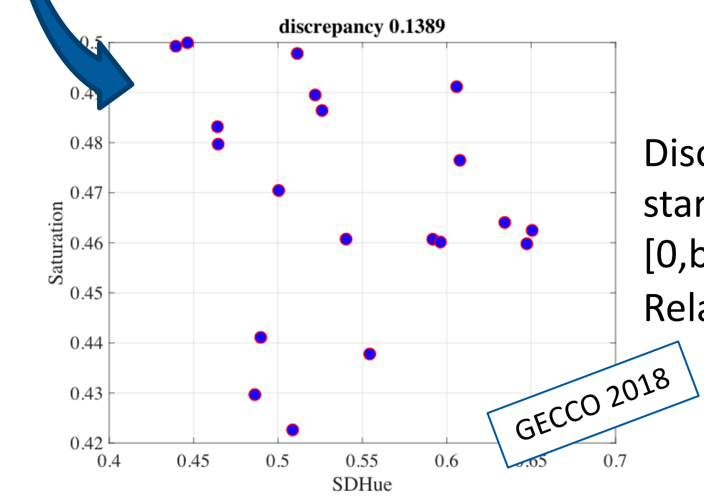
Australia

Multi-Objective Indicators (Images)









Discrepancy theory studies the irregularity of distributions
star discrepancy: regularity w.r.t. all axis-parallel boxes
$[0,b], b \in [0,1]^d$ that are anchored in the origin.
Relatively time consuming ($n^{1+d/2}$), and counter-intuitive distributions.