# Recent Advances in Landscape Analysis for Optimisation and Learning


**Katherine Malan[1] & Gabriela Ochoa[2]**
[1]University of South Africa, Pretoria, South Africa
[2]University of Stirling, Stirling, UK
malankm@unisa.ac.za, gabriela.ochoa@stir.ac.uk


http://gecco-2021.sigevo.org/



GECCO '21 Companion, Lille, France



https://doi.org/10.1145/3449726.3461396




## Instructors

❖ Katherine M. Malan is an associate professor in the Department of Decision Sciences at the University of South Africa. She holds a PhD from the University of Pretoria. Her research interests include fitness landscape analysis and the application of computational intelligence techniques to real-world problems. She regularly reviews for a number of journals in fields related to evolutionary computation, swarm intelligence and operations research.

❖ Gabriela Ochoa is a Professor in Computing Science at the University of Stirling, Scotland. She holds a PhD from the University of Sussex, UK. Her research interests include evolutionary and heuristic search methods, with emphasis on autonomous search, hyper-heuristics, fitness landscape analysis and visualisation. She is a member of the editorial board editor for Evolutionary Computation and an associate editor of the ACM TELO journal. She served as the EiC for GECCO 2017, and is a member of the SIGEVO board.



## Outline

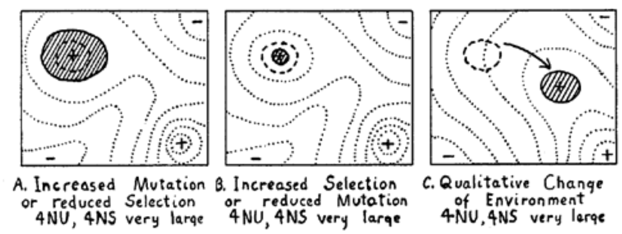
❖ Introduction to Fitness Landscapes
- Motivation for analysing fitness landscapes
- Basics of fitness landscapes

❖ Recent Advances in Landscape Analysis
- Beyond fitness landscapes
- Recent landscape analysis techniques
- Applications of landscape analysis

❖ Local Optima Networks (LONs)
- Basics: complex networks, nodes & edges, visualisation and metrics
- Case study 1: Global structure and characterisation of funnels
- Case study 2: Contrasting two optimisation algorithms

❖ Closing






(Wright, 1932)

## INTRODUCTION TO FITNESS LANDSCAPES

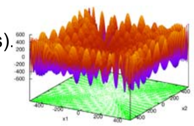
- Motivation for analysing fitness landscapes
- Basics of fitness landscapes
- Survey of 22 fitness landscape analysis techniques

2013 Survey

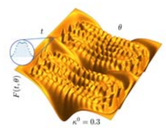
Malan, K.M. and Engelbrecht, A.P. (2013). A survey of techniques for characterising fitness landscapes and some possible ways forward. *Information Sciences*, 241:148-163

# Complex Optimisation

- When classical techniques are not feasible & metaheuristics are needed:
  - Problem complexity is too large (not of the required structure for classical techniques, too many variables).
  - When there is no objective function in mathematical form.
  - Objective function exhibits noise or uncertainty.



- "Massive optimisation"
  - Large scale optimisation (many dimensions)
  - Any-objective optimisation (single-, multi- many-objective)
  - Cross-domain optimisation (continuous / combinatorial / mixed)
  - Expensive optimisation (costly / simulation-based black-box evaluations)

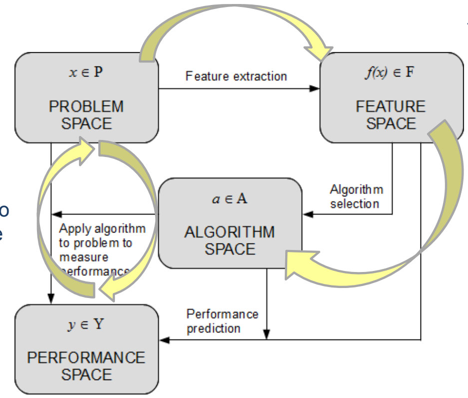


- Many many metaheuristic approaches…

5

# General Algorithm Selection Problem (Rice, 1976)

Trial-and-error approach to finding the best algorithm




Research areas:
- Feature extraction (characterising problems)
- Analysis and design of problems
- Understanding algorithm behaviour
- Performance prediction
- Algorithm selection

6

# Wright's fitness landscape

- Surface of selective values (Wright, 1932).
- No axes, units or labels.
- Commentary 56 years later:
  - *"useless for mathematical purposes"*
  - Aim: provide an intuitive picture of evolutionary processes taking place in higher dimensional space.

1889 - 1988

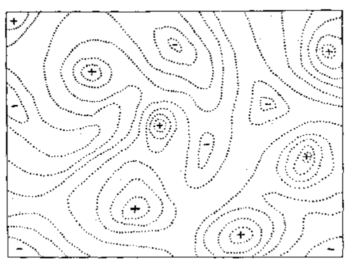
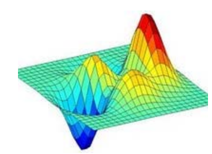


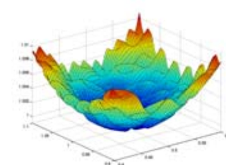
FIGURE 2—Diagrammatic representation of the field of gene combinations in two dimensions instead of many thousands. Dotted lines represent contours with respect to adaptiveness.

7

# Fitness landscapes today

- We now have (useful) formalised mathematical models.
- Three essential elements:
  1. Search space
  2. Fitness function
  3. Notion of neighbourhood or accessibility.
- Intuitively, a fitness landscape is a visualisation of the terrain capturing how fitness changes between neighbouring solutions.
- Idea of "valleys", "peaks", "ridges", "plateaus", "funnels", etc.
- One fitness function, many fitness landscapes (even for real-valued spaces) – depends on the neighbourhood and the sampling.





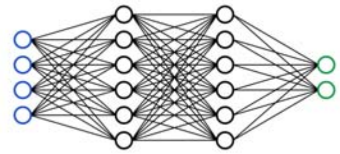
8

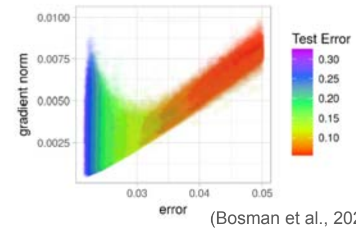
## Fitness landscape analysis techniques (2013 survey)

| | |
|---|---|
| Technique 1 | GA-deception (Goldberg, 1987) |
| … | |
| Technique 7 | Fitness distance correlation (Jones & Forrest, 1995) |
| … | |
| Technique 19 | Dispersion metric (Lunacek & Whitley, 2006) |
| … | |
| Technique 22 | Accumulated escape probability (Lu et al., 2011) |

Malan, K.M. and Engelbrecht, A.P. (2013). A survey of techniques for characterising fitness landscapes and some possible ways forward. *Information Sciences*, 241:148-163








(Bosman et al., 2020b)

## RECENT ADVANCES IN LANDSCAPE ANALYSIS

- Beyond fitness landscapes
- Additional 11 landscape analysis techniques
- Applications of landscape analysis

2021 Survey

Malan, K.M. (2021a). A Survey of Advances in Landscape Analysis for Optimisation. *Algorithms,* 14(2).



## Beyond *fitness* landscapes

- Idea of fitness landscapes have been applied in non-evolutionary contexts, so many are dropping the fitness metaphor.
- Original three elements of fitness landscapes:
  1. Search space
  2. Fitness function
  3. Notion of neighbourhood or accessibility.
- New kinds of landscapes:
  - Multiobjective fitness landscapes
  - Violation landscapes
  - Dynamic and coupled fitness landscapes
  - Error landscapes

Search landscape analysis

Exploratory landscape analysis

Error / Loss landscape analysis

Landscape analysis



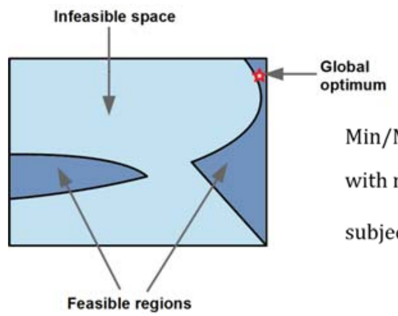
## Multiobjective fitness landscapes

- Multiobjective optimisation (MOO) differs from single-objective optimisation:
  1. Objective-level: Multiple conflicting objectives – what defined the surface of the search landscape?
  2. Solution-level: **Set** of Pareto-optimal solutions.
- Possible approach to defining MOO landscapes (Verel et al., 2011):
  1. Search space: Set of solution-sets.
  2. Fitness function: Multiobjective quality measure (such as hypervolume), which allows a complete order between solution-sets.
  3. Neighbourhood: set-level operators (e.g. replacement, insertion or deletion of a single element).

Verel, S., Liefooghe, A. and Dhaenens, C. (2011). Set-based multiobjective fitness landscapes: A preliminary study. In Proceedings of the 13th annual Conference on Genetic and Evolutionary Computation, Dublin, Ireland, 12–16 July 2011; pp. 769–776.

# Violation landscapes

❖ With constrained optimisation problems, the level of constraint violation can be treated as an additional objective to the minimised.




$$\text{Min/Max} \quad f(x_1, x_2, \ldots, x_n)$$

$$\text{with respect to} \quad x_1, x_2, \ldots, x_n$$

$$\text{subject to} \quad g_k(x_1, x_2, \ldots, x_n) \begin{Bmatrix} \leq \\ = \end{Bmatrix} b_k$$

$$k = 1, \ldots, m$$

Malan, K.M., Oberholzer, J.F. and Engelbrecht, A.P. Characterising Constrained Continuous Optimisation Problems. In *Proceedings of the IEEE Congress on Evolutionary Computation*, May 2015, Sendai, Japan, pp 1351-1358.

13

# Violation landscapes – additional view

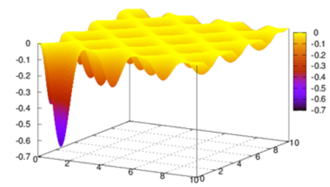
❖ CEC 2010 Benchmark suite, problem C01:

$$\text{Min } f(\mathbf{x}) = -\left|\frac{\sum_{i=1}^{D}\cos^4(z_i) - 2\prod_{i=1}^{D}\cos^2(z_i)}{\sqrt{\sum_{i=1}^{D} i z_i^2}}\right|, \ \mathbf{z} = \mathbf{x} - \mathbf{o}$$

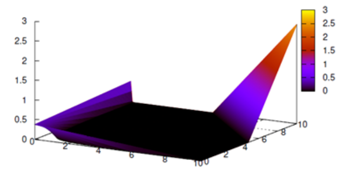
$$g_1(\mathbf{x}) = 0.75 - \prod_{i=1}^{D} z_i \leq 0$$

$$g_2(\mathbf{x}) = \sum_{i=1}^{D} z_i - 7.5D \leq 0$$

$$\mathbf{x} \in [0, 10]^D$$



Fitness landscape



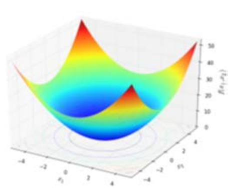
Violation landscape

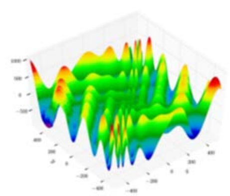
14

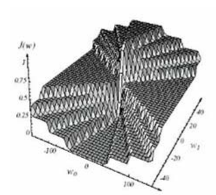
# Error landscapes

❖ Neural network error landscape:
- Every weight vector is associated with an error value.
- The set of all possible weight vectors (neural network instances) with error values defines the error landscape.
- Dimensionality of the search space is equal to the total number of weights.

❖ If these high-dimensional spaces could be visualised, what would they look like?







15

# Contradictory theoretical results

**Abstract**—*We consider the problem of learning from examples in layered linear feed-forward neural networks using optimization methods, such as back propagation, with respect to the usual quadratic error function E of the connection weights. Our main result is a complete description of the landscape attached to E in terms of principal component analysis. We show that E has a unique minimum corresponding to the projection onto the subspace generated by the first principal vectors of a covariance matrix associated with the training patterns. All the additional critical points of E are saddle points (corresponding to projections onto subspaces generated by higher order vectors). The auto-associative case is examined in detail. Extensions and implications for the learning algorithms are discussed.*

Baldi & Hornik (1989)

Auer *et al*. (1996)

We show that for a single neuron with the logistic function as the transfer function the number of local minima of the error function based on the square loss can grow exponentially in the dimension.

This paper presents a case study of the analysis of local minima in feedforward neural networks. Firstly, a new methodology for analysis is presented, based upon consideration of trajectories through weight space by which a training algorithm might escape a hypothesized local minimum. This analysis method is then applied to the well known XOR (exclusive-or) problem, which has previously been considered to exhibit local minima. The analysis proves the absence of local minima, eliciting significant aspects of the structure of the error surface. The

Hamey (1998)

**Loss surface of XOR artificial neural networks**

for potential energy landscapes in molecular science. The number of local minima and transition states (saddle points of index one), as well as the ratio of transition states to minima, grow rapidly with the number of nodes in the network. There is also a strong dependence on the regularization parameter, with the landscape becoming more

Mehta et al. (2018)

16

## Landscape analysis of error landscapes?

Neural network (NN) training differs from benchmark optimisation problems:

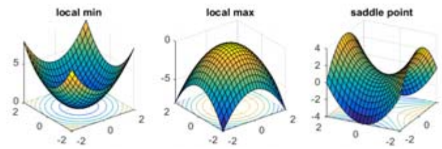
- Very high dimensions (MNIST with 10 hidden neurons gives 7960 weights).
- Evaluation of objective is expensive (involves full run through training data set).
- Search space of weights is unbounded.
- The same solution can evaluate to different error values (depends on the subset of data instances used in training).
- Analytical gradient information is available.
- The global optimum in the training landscape ≠ global optimum in the test landscape.

Training and testing error become increasingly **de-correlated** with the size of the network (Choromanska *et al.*, 2015)

17

## Landscape analysis applied to NNs

- Landscape metrics (ruggedness, neutrality, etc.) can be derived from random walk samples in weight space.
- But, we also have the analytical gradient at each point.
  - Random walk sampling can be biased towards the negative gradient.
  - Stationary points can be identified (gradient = 0).
  - Curvature can be derived from the eigenvalues of the Hessian matrix.
- Stationary points can be:
  - Minima: convex curvature (eigenvalues negative).
  - Maxima: concave curvature (eigenvalues positive).
  - Saddle points: both curvatures (eigenvalues positive and negative).
  - Flat: no curvature (singular Hessian).




18

## Landscape analysis techniques (2021 survey – follow on from 2013 survey)

| | |
|---|---|
| Technique 23 | Local optima networks (LONs) by Ochoa et al. (2008) |
| Technique 24 | Exploratory landscape analysis (ELA) by Mersmann et al. (2011) |
| Technique 25 | Length scale distribution by Morgan & Gallagher (2012) |
| Technique 26 | Codynamic landscape measures by Richter (2014) |
| Technique 27 | Degree of separability by Caraffini et al. (2014) |
| Technique 28 | Constrained landscape metrics by Malan et al. (2015) |

Malan, K.M. (2021a). A Survey of Advances in Landscape Analysis for Optimisation. *Algorithms,* 14(2).

19

## Landscape analysis techniques (2021 survey – follow on from 2013 survey)

| | |
|---|---|
| Technique 29 | Bag of local landscape features by Shirakawa & Nagao (2016) |
| Technique 30 | Maximum entropic epistasis by Sun et al. (2017) |
| Technique 31 | Population evolvability metrics by Wang et al. (2018) |
| Technique 32 | Local multiobjective landscape features by Liefhooghe et al. (2019) |
| Technique 33 | Loss-gradient clouds by Bosman et al. (2020) |

Malan, K.M. (2021a). A Survey of Advances in Landscape Analysis for Optimisation. *Algorithms,* 14(2).

20

# Technique 28:
# Constrained landscape metrics

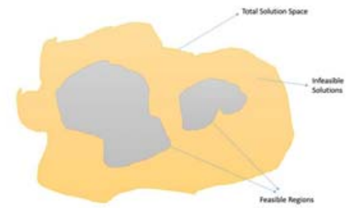
| **Technique 28** | Constrained landscape metrics by Malan et al. (2015). |
|---|---|
| Focus | Constraint violation in relation to fitness |
| Assumptions | Assumes that the extent to which constraints are violated can be quantified for all solutions. |
| Description | Given a sequence of solutions based on random/ hill climbing walks, with associated fitness and level of constraint violation for each solution, the following are estimated:<br>(1) the proportion of feasible solutions in the search space …<br>(2) the level of disjointedness between feasible areas …<br>(3) the correlation between the fitness and violation …<br>(4) the proportion of solutions that are both high in fitness and low in constraint violation … |

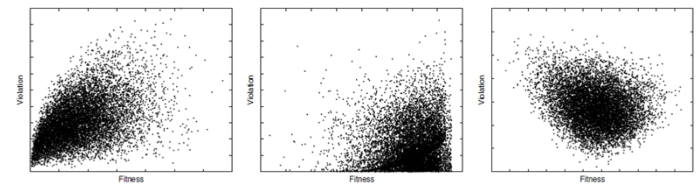
Malan, K.M., Oberholzer, J.F. and Engelbrecht, A.P. (2015) Characterising Constrained Continuous Optimisation Problems. In *Proceedings of the IEEE Congress on Evolutionary Computation*, Sendai, Japan, pp 1351-1358.

21

# Four metrics of constrained landscapes

- What proportion of the sample is feasible?
- How disjoint are the feasible areas?
- How correlated are the fitness and violation landscapes? Do they "pull" in the same direction?
- What proportion of the solutions are both feasible and fit?



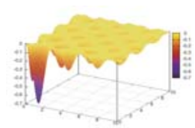



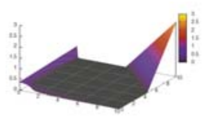

(d) C03 problem in 10D (e) C04 problem in 10D (f) C13 problem in 10D

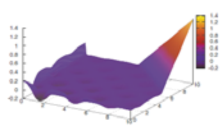
22

# Landscape aware constraint handling

- Metaheuristics do not naturally handle constraints, so a constraint-handling technique has to be added on.
- Many different approaches to handling constraints:
  - Use penalties: adapt fitness function to guide search away from infeasible regions.
  - Feasibility ranking: rules of preference using objectives and constraints.
  - Multi-objective optimisation (constraints treated as objective to be minimised).
- A landscape-aware approach that switches between constraint-handling approaches is more effective than the constituent approaches:
  - With differential evolution (Malan, 2018).
  - With particle swarm optimisation (Malan, 2021b).







23

# Technique 32:
# Local multiobjective landscape features

| **Technique 32** | Local multiobjective landscape features by Liefooghe et al. (2019) |
|---|---|
| Focus | Evolvability for multiobjective optimisation |
| Assumptions | Assumes a discrete search space |
| Description | Given a sequence of solutions obtained through random walks and adaptive walks, features of the walk are derived from the sequence as a whole as well as the neighbourhood of solutions in terms of dominance and hypervolume improvement by neighbours. |
| Result | 26 numerical values representing local features (17 from random walk sampling and 9 from adaptive walk sampling). |

Liefooghe, A., Daolio, F., Verel, S., Derbel, B., Aguirre, H. and Tanaka, K. (2019), Landscape-Aware Performance Prediction for Evolutionary Multi-objective Optimization. *IEEE Transactions in Evolutionary Computation*, 24(6).

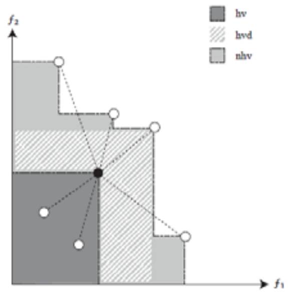
24

## Example local features from random walk sampling (17 features)

1) average proportion of neighbours dominated by the current solution.
2) first autocorrelation coefficient (ruggedness) of the proportion of neighbours dominated by the current solution.

…

15) average neighbourhood's hypervolume-value.
16) …
17) estimated correlation between the objective values




Liefooghe, A., Daolio, F., Verel, S., Derbel, B., Aguirre, H. and Tanaka, K. (2019), Landscape-Aware Performance Prediction for Evolutionary Multi-objective Optimization. *IEEE Transactions in Evolutionary Computation*, 24(6).

25

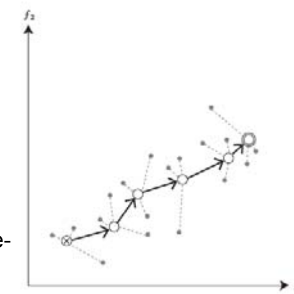
## Example local features from adaptive walk sampling (9 features)

Sampling based on adaptive walks: single solution-based multi-objective Pareto hill climber.

1) average proportion of neighbours dominated by the current solution
2) average proportion of neighbours dominating the current solution

…

8) average neighbourhood's hypervolume-value
9) average length of adaptive walks

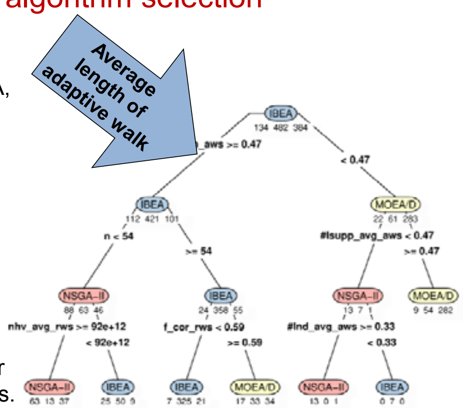



Liefooghe, A., Daolio, F., Verel, S., Derbel, B., Aguirre, H. and Tanaka, K. (2019), Landscape-Aware Performance Prediction for Evolutionary Multi-objective Optimization. *IEEE Transactions in Evolutionary Computation*, 24(6).

26

## Feature-based EMO algorithm selection

- Portfolio of three EMO algorithms: NSGA-II, IBEA, MOEA/D.
- Problems: 1000 quadratic assignment problem instances.
- Neighbourhood sampling: 200 random.
- CART decision tree was able to choose algorithm that was not significantly outperformed by any other in almost 99% of the cases.




Liefooghe, A., Daolio, F., Verel, S., Derbel, B., Aguirre, H. and Tanaka, K. (2019), Landscape-Aware Performance Prediction for Evolutionary Multi-objective Optimization. *IEEE Transactions in Evolutionary Computation*, 24(6).

27

## Technique 33: Loss gradient clouds

| | |
|---|---|
| **Technique 33** | Loss-gradient clouds by Bosman et al. (2020b) |
| Focus | Basins of attraction in neural network error landscapes |
| Assumptions | Requires the numeric gradient of the loss function. |
| Description | A sample of loss values and gradient values is obtained based on random walks. Stationary points in the sample are determined to be local minima, local maxima or saddle points (derived from the eigenvalues of the Hessian matrix). Stagnant sequences on the walk are detected by tracking the deviation in a smoothing of the error. Two quantities are measured:<br>(1) the average number of times that stagnation was observed<br>(2) the average length of the stagnant sequence |
| Result | (1) 2D scatterplot of loss values against gradient values<br>(2) Two metrics to estimate the number and extent of distinct-valued basins of attraction. |

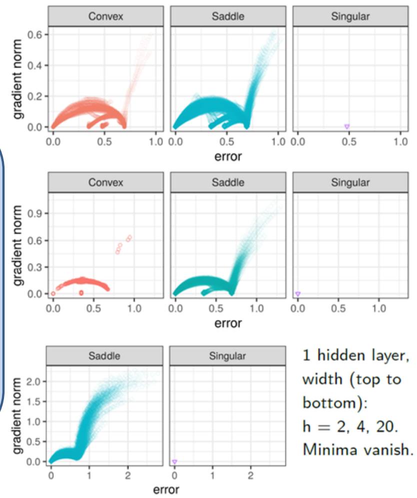
Bosman, A.S., Engelbrecht, A.P. and Helbig, M. (2020b), Visualising Basins of Attraction for the Cross-Entropy and the Squared Error Neural Network Loss Functions. *Neurocomputing*, 400, pp. 113–136.

28

# Loss gradient clouds

**XOR** classification problem with 1 hidden layer:

- Number of attractors decrease as the width (number of neurons) of the network increases.
- **Local minima vanish** with width = 20.




1 hidden layer, width (top to bottom): h = 2, 4, 20. Minima vanish.

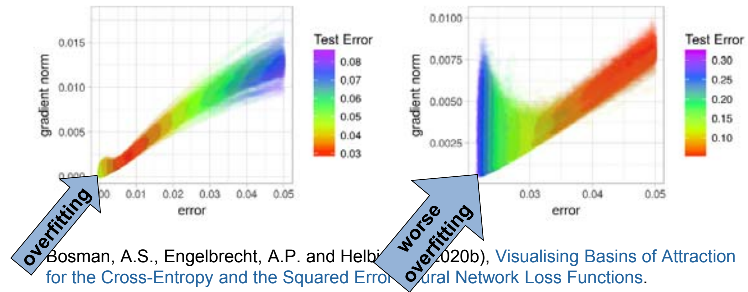
Bosman, A.S., Engelbrecht, A.P. and Helbig, M. (2020a), Loss surface modality of feed-forward neural network architectures. In Proceedings of International Joint Conference on Neural Networks (IJCNN), .



# Loss gradient clouds

**Iris** classification problem:

- Loss-gradient cloud shows a single attractor.
- **Quadratic loss** function (left picture) shows **better generalisation** than entropic loss function (on the right) close to the training global optimum.



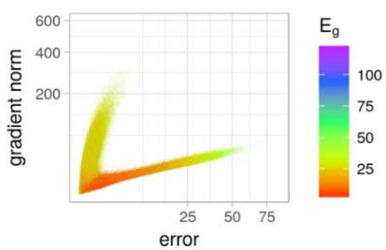

Bosman, A.S., Engelbrecht, A.P. and Helbig, M. (2020b), Visualising Basins of Attraction for the Cross-Entropy and the Squared Error Neural Network Loss Functions. *Neurocomputing*, 400, pp. 113–136.



# Loss gradient clouds

**MNIST** classification problem with 2 hidden layers & width of 10:

- Loss-gradient cloud shows two clusters: steep gradients (narrow valleys) and shallow gradients (wider valleys).
- **Narrower valleys have poorer generalisation.**




Bosman, A.S., Engelbrecht, A.P. and Helbig, M. (2020a), Loss surface modality of feed-forward neural network architectures. In Proceedings of International Joint Conference on Neural Networks (IJCNN), .



# Applications of landscape analysis: 2013 to now

A B S T R A C T

Real-world optimisation problems are often very complex. Metaheuristics have been successful in solving many of these problems, but the difficulty in choosing the best approach can be a huge challenge for practitioners. One approach to this dilemma is to use fitness landscape analysis to better understand problems before deciding on approaches to solving the problems. However, despite extensive research on fitness landscape analysis and a large number of developed techniques, very few techniques are used in practice. This could be because fitness landscape analysis in itself can be complex. In an attempt to make fit-

No longer the case!

Malan, K.M. and Engelbrecht, A.P. (2013). A survey of techniques for characterising fitness landscapes and some possible ways forward. *Information Sciences*, 241:148-163

## Applications of landscape analysis

In the last 10 years, landscape analysis has been widely used in

- understanding complex problems
- understanding and explaining algorithm behaviour
- predicting algorithm performance
- algorithm configuration
- automated algorithm selection

See 2021 survey for ~70 references to studies applying landscape analysis in the last decade.

Malan, K.M. (2021a). A Survey of Advances in Landscape Analysis for Optimisation. *Algorithms,* 14(2).

33

## Understanding complex problems

| Classic problems (18 studies listed) | Real-world problems (11 studies listed) | Machine learning applications (7 studies listed) |
|---|---|---|
| - Quadratic assignment<br>- Maximum satisfiability<br>- Permutation flow-shop scheduling<br>- Packing problems<br>- Travelling salesman<br>- Graph colouring<br>- Number partitioning<br>- Vehicle routing<br>- Travelling thief problem | - Design of wind turbines<br>- University course timetabling<br>- Genetic improvement of software<br>- Automated test case generation<br>- Computational protein design<br>- Design of substitution boxes in cryptography<br>- Hyperparameter optimisation for metaheuristics<br>- Building energy optimisation | - Neural network training for classification<br>- Feature selection for classification<br>- Policy search in reinforcement learning<br>- Machine learning pipeline configuration<br>- Neural architecture search |

34

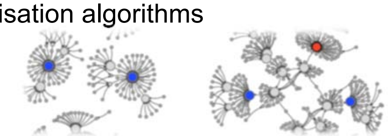
## Summary

- Research in landscape analysis has moved from being a theoretical topic in evolutionary computation,
- to being extensively applied as a practical tool in the wider context of optimisation,
- and has recently also been applied in machine learning.
- A number of new landscape analysis techniques have been developed in the last decade.
- Landscape analysis has been widely applied in the understanding of complex problems, explaining algorithm behaviour, predicting algorithm performance and automatically configuring and selecting algorithms.

35

## Local optima networks (LONs)

- What are complex networks?
  - Visualisation
  - Metrics
- What are local optima networks?
  - Overview and intuition
  - Definition of nodes and edges
- LONs case studies
  - Global structure: characterisation of funnels
  - Contrasting two optimisation algorithms

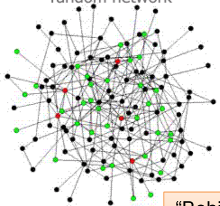


36

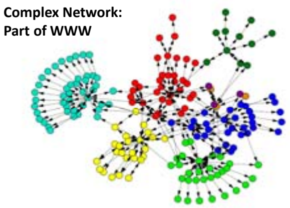
## Complex networks

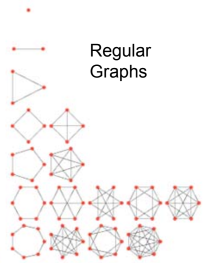
- **Graph** - mathematical object
- **Network** - data-driven instantiation
- **Complex network** - nontrivial topological features
  - Irregular as opposed to regular/simple
  - Not random either
- Can evolve over time

random network



Complex Network: Part of WWW





Regular Graphs

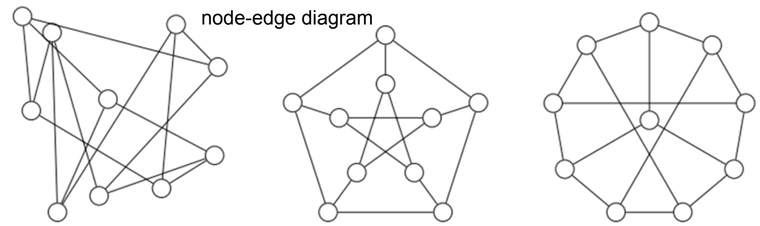
"Behind each complex system, there is an intricate network that encodes the interactions between the system's components."
Albert-László Barabási, Network Science

37

## Graph visualisation

- Graphs are abstract mathematical structures
- They do not have a unique visual representation
- Graph visualisation: art of choosing an appropriate representation that is aesthetically pleasing and highlights important structural properties

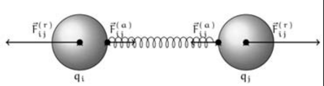
node-edge diagram



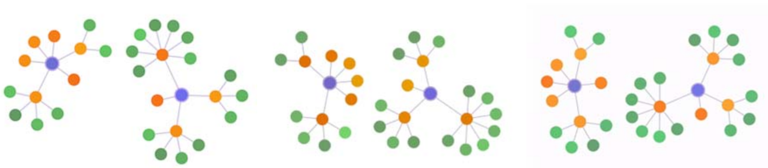
Three layouts of Petersen graph (the 1st is random)

38

## Graph layout algorithms



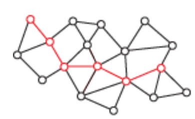
- Force-directed: based on physical analogy (electricity, springs)
- Strive to satisfy accepted aesthetic criteria
  - Vertices are distributed roughly evenly on the plane.
  - The number of edge crossings is minimised.
  - The lengths of edges are approximately uniform.
  - Inherent symmetries in the graph are respected



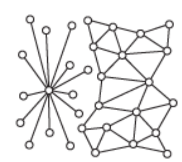
Fruchterman–Reingold (*standard*) | Kamada–Kawai (*organic*) | Reingold-Tilford (*tree layout*)

39

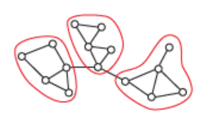
## Network metrics



Distance
- Number of links that make up the path between two points
- "Geodesic" = shortest path



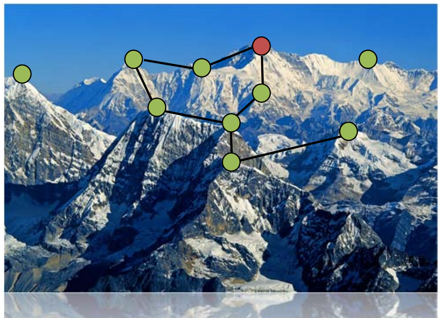
Topology (Degree distribution)
- Gives an idea of the spread in the number of links the nodes have
- $p(k)$ is the probability that a randomly selected node has $k$ links



Cohesion
- Local: clustering coefficient or transitivity
- Global: components, community structure

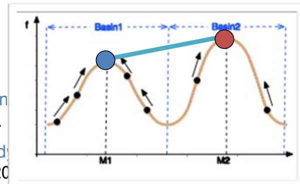
40

# Local optima networks (LONs)



Nodes: local optima according to a hill-climbing heuristic

Edges: possible transitions between optima




P. K. Doye. The network topology of a potential energy lan… a static scale-free network. *Physical Review Letter*, 2002.

G. Ochoa, M. Tomassini, S. Verel, and C. Darabos. A stud… landscapes' basins and local optima networks. *GECCO* 20…



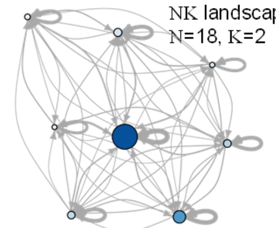
# LON original model

**Algorithm 1:** Best-improvement local search

Choose initial solution $s \in S$
**repeat**
  choose $s' \in N(s)$, $f(s') = max_{x \in N(s)} f(x)$
  **if** $f(s) \leq f(s')$ **then**
    $s \leftarrow s'$
  **end if**
**until** $s$ is a Local optimum

- Space $S$, Neigborhood $N(s)$, fitness $f(s)$
- $h(s)$ stochastic operator that associates each solution $s$ to its local optimum (Alg. 1)
- The basin of attraction of a local optimum $l_i \in L$ is the set $B_i = \{s \in S \mid h(s) = l_i\}$
- Nodes ($L$). A local optima is a solution $l$ such that $\forall\ s \in N(s), f(s) \leq f(l)$
- Basin Edges ($E$). Two local optima are connected if their basins of attraction intersect. At least one solution within a basin has a neighbour within the other basin.
- LON Model. Directed graph LON = ($L$, $E$)

NK landscape
N=18, K=2



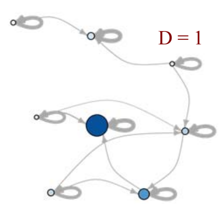
$w_{ij}$ proportion of transitions from solutions $s \in Bi$ to solutions $s' \in Bj$

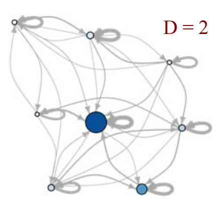


# Escape edges

- Account for the chances of escaping a local optimum after a controlled mutation (e.g. 1 or 2 bit-flips in binary space) followed by hill-climbing
- Given a distance function $d$ and integer value $D$, there is and edge $e_{ij}$ between $l_i$ and $l_j$ if a solution s exists such that $d(s,l_i) \leq D$ and $h(s) = l_j$
- $w_{ij}$ cardinality of $\{s \in S \mid d(s,l_i) \leq D \text{ and } h(s) = l_j\}$
- Sampled networks. There is an edge $e_{ij}$ between $l_i$ and $l_j$ if $l_j$ can be obtained after applying a *perturbation* to $l_i$ followed by hill-climbing. Weights are estimated by the sampling process.

NK landscape
N=18, K=2



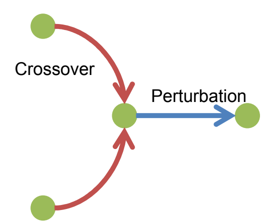



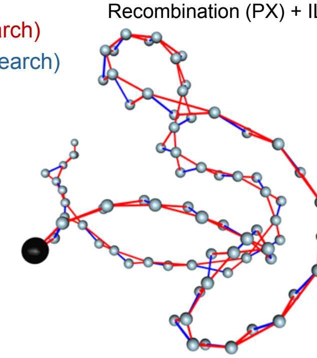



# LONs for hybrid EAs

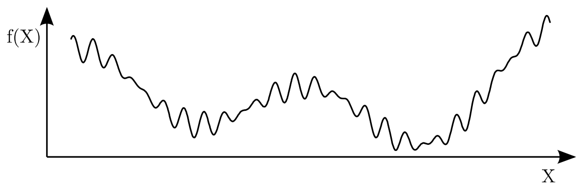
- Hydrid EAs which incorporate a local search component to generate local optima.
- Two types of edges
  - Crossover (followed by local search)
  - Perturbation (followed by local search)

Grey-box Optimisation
DRILS Deterministic Recombination (PX) + ILS



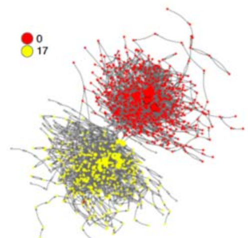



(Chicano, Whitley, Ochoa, Tinos. GECCO 2017)




- What is a funnel?
- Characterization of funnels with LONs
  - Number partitioning problem and the phase transition
  - Travelling salesman problem
  - Computational protein design

# FUNNELS

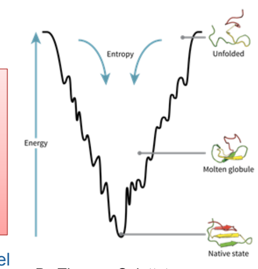



45

## What is a funnel?

"A key concept that has arisen within the protein folding community is that of a *funnel* consisting of a set of downhill pathways that converge on a single low-energy minimum."

Doye, J. P. K., Miller, M. A., & Wales, D. J . The double-funnel energy landscape of the 38-atom Lennard-Jones cluster. *Journal of Chemical Physics*, 1999




By Thomas Splettstoesser (link)(www.scistyle.com) - Own work

Funnels in continuous optimisation

- Multilevel global structure (Locatelli, 2005)
- *Dispersion* metric (Lunacek &Whitley, 2006, 2008)
- Feature-based detection of (single) funnel structure (Kerschke et al., 2015)

Funnels in combinatorial optimisation

- Related to the big-valley (central-massif) hypothesis (Boese et al, 1994)
- The big-valley re-visited (Hains, Whitley & Howe, 2011)

46

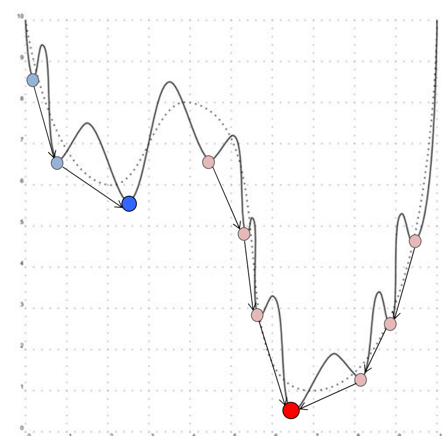
## Characterisation of funnels with LONs

- ❖ Monotonic edges. Keep only non-deteriorating edges $l_s \rightarrow l_e$, if $f(l_e) \leq f(l_s)$
- ❖ Monotonic LON (MLON). LON model where the set of edges is reduced to the non-deteriorating edges
- ❖ Monotonic sequence. Path of connected local optima $l_1 \rightarrow l_2 \rightarrow l_3 \ldots \rightarrow l_s$ ,$f(l_i) \leq f(l_{i-1})$
- ❖ Sink. Natural end of the sequence, when there is no adjacent improving local optima
- ❖ Definition of Funnel
  - Aggregation of all monotonic sequences ending at the same point (sink).
  - Basin of attraction level of local optima

***Sink***. Node without outgoing edges

47

## Characterisation of funnels with LONs

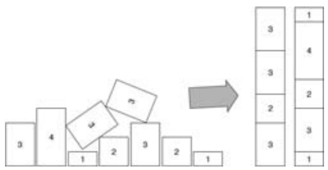


- Global minimum
- Sub-optimal sink
- Local minimum in global optimal funnel
- Local minimum in sub-optimal funnel
- Monotonic edge

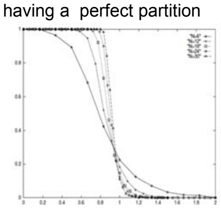
48

# Number partitioning (NPP)

- Given a set of $n$ positive integers $A=\{a_1, a_2, ..,a_n\}$, drawn at random from the set $\{1, 2, .., M\}$, find a disjoint partition $(S_1, S_2)$ of $A$ such that the discrepancy $D$ between their sums is minimised
- A partition is perfect if $D = 0$, where $D = |\ \Sigma_{S1}\ a_i - \Sigma_{S2}\ a_i\ |$
- Easy-hard phase transition, $k = \log_2(M)/n$



Probability of an NPP instance having a perfect partition



(Gent & Walsh, 1996)

$k$ < 1 many perfect partitions
$k$ > 1 very few perfect partitions
$k$ = 1 easy/hard phase transition

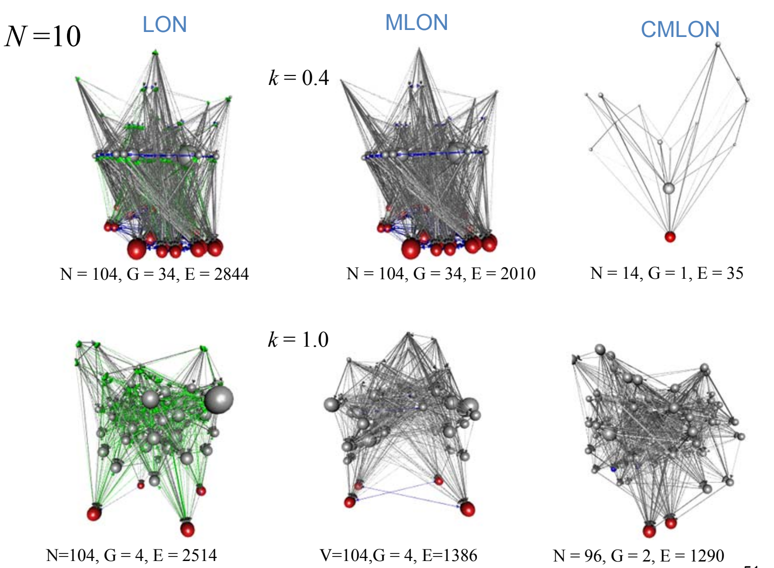


# NPP study - methodology

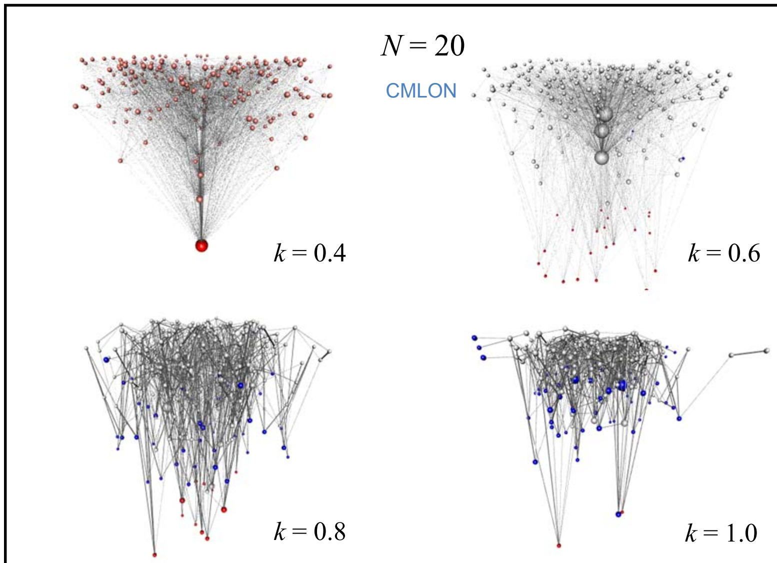
- Full enumeration and extraction of LONs
- $N = \{10, 15, 20\}$, $k$ in $[0.4, 1.2]$ step $0.1$
- 30 instances for each $N$ and $k$
- LON. 1-flip local search, 2-flip perturbation ($D = 2$)
- MLON. Monotonic LON, worsening edges pruned
- CMLON. compressed MLON, LON plateaus contracted in a single node
- Empirical search performance: ILS success rate

(Ochoa, Veerapen, Daolio, Tomassini. EvoCOP 2017)

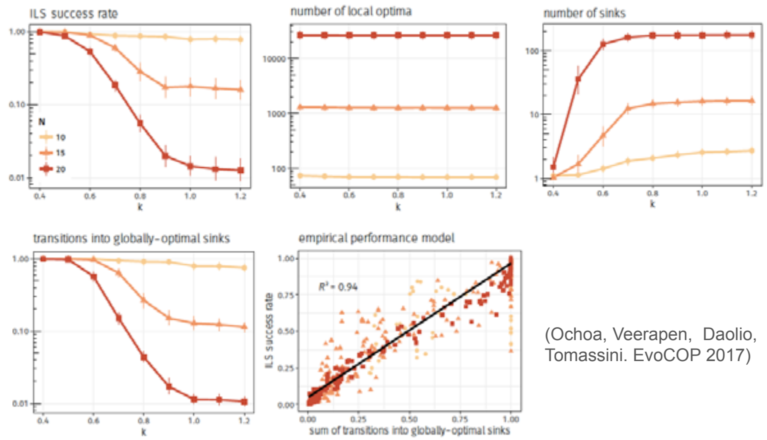






# LON metrics & search performance

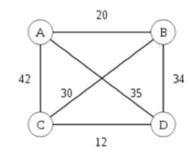



(Ochoa, Veerapen, Daolio, Tomassini. EvoCOP 2017)



# Travelling salesman problem (TSP)

- A salesperson must visit number of cities minimising the total cost of traveling
- *n* cities and distance between each pair is given
- Goal: find a shortest route to visit each city exactly once and come back to the starting point.
- Example solutions: permutation (ordering) of cities
  - s1= (A B C D), f(s1)= 20+30+12+35= 97
  - s2= (A B D C), f(s2)= 20+34+12+42=108
  - s3= (A C B D), f(s3)= 42+30+34+35= 141





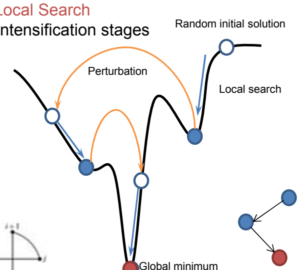

# TSP study - methodology

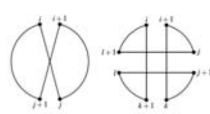
Sampling and constructing LONs with escape edges

**Data:** $I$, TSP instance
**Result:** $L$, set of local optima,
$E$, set of escape edges
$L \leftarrow \{\}; E \leftarrow \{\}$
**for** $i \leftarrow 1$ **to** 1000 **do**
  $s_{start} \leftarrow$ initialSolution()
  $s_{start} \leftarrow \mathrm{LK}(s_{start})$
  $L \leftarrow L \cup \{s_{start}\}$
  **while** $j < 10000$ **do**
    $s_{end} \leftarrow \mathrm{applyKick}(s_{start})$
    $s_{end} \leftarrow \mathrm{LK}(s_{end})$
    $j \leftarrow j + 1$
    **if** $Objective(s_{end}) \leq Objective(s_{start})$ **then**
      $L \leftarrow L \cup \{s_{end}\}$
      $E \leftarrow E \cup \{(s_{start}, s_{end})\}$
      $s_{start} \leftarrow s_{end}$
      $j \leftarrow 0$
    **end**
  **end**
**end**

TSP heuristic, Chained Lin-Kernighan
(Martin, Otto, Felten, 1992)

- Form of Iterated Local Search
- Diversification & Intensification stages



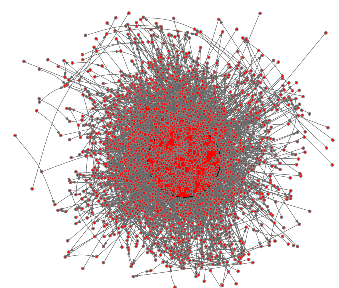




- Nodes. Lin-Kernighan
- Edges. Double-bridge

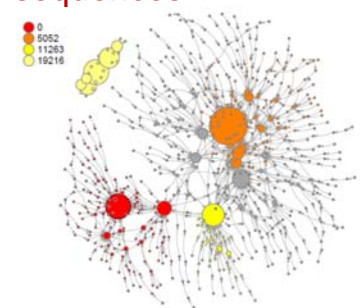
(Ochoa & Veerapen, EvoCOP2016, JoH 2018)



# TSP Synthetic Instances
# Funnels as monotonic sequences



F:88#Foxvwhuhg#Flwlhv
Ixqqhov=4/#Vxfflhvv=433(

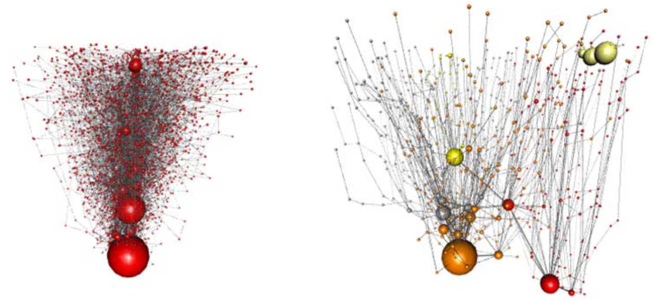



H:88#Xqliruup#Flwlhv
Ixqqhov=7/#Vxfflhvv=46(

DIMACS random instances

(Ochoa & Veerapen, JoH 2018)

## TSP synthetic instances



F:88#Foxvwhuhg#Flwhv
Ixqqhov=#4/#Vxffhvv=#433(

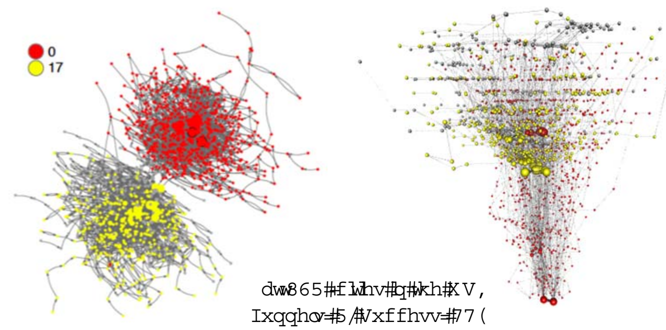
H:88#Xqlirup#Flwhv
Ixqqhov=#7/#Vxffhvv=#46(

DIMACS random instances

Same layout, 3D projection where $z$ coordinate is fitness

57

## TSPLIB city Instance att532

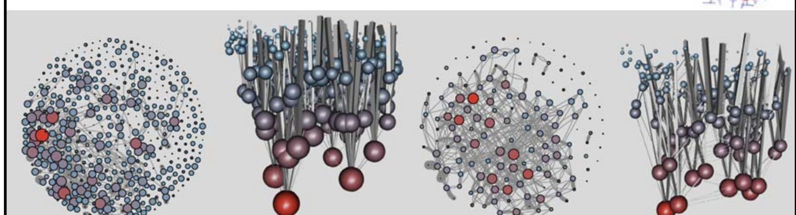



dw865#flwhv#lq#wkh#XV,
Ixqqhov=#5/#Vxffhvv=#77(

2D layout and 3D projection where $z$ coordinate is fitness

58

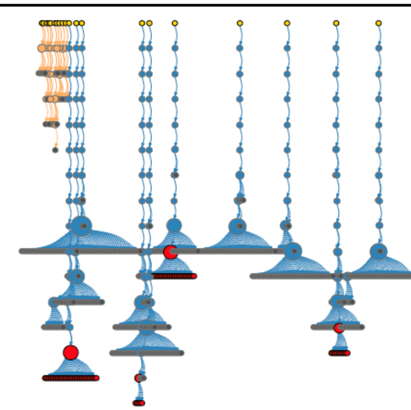
## Computational protein design



- Two protein instances 2ckx (difficult, left) 2gkt (easy, right),
- Minimisation of energy algorithm based in Simulated Annealing
- LONs provided an explanation! There are no connections between the funnels in the harder instance while there are some connections in easier instance

(David Simoncini et al, GECCO 2018)

59



- Hybrid GA vs ILS on the asymmetric TSP
- Two hybrid metaheuristics (matheuristics) on

## CONTRASTING ALGORITHMS

60

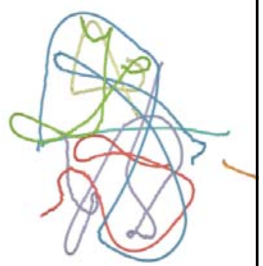
## Contrasting LONs from two solving methods

Hybrid GA vs ILS on asymmetric travelling salesman problem

Asymmetric TSP
Instance rbg323 LONs.

- 100 runs per algorithm
- Hybrid GA has crossover and mutation edges
- ILS has perturbation edges

Hybrid GA
Partition Crossover (PX)
Success: 100%



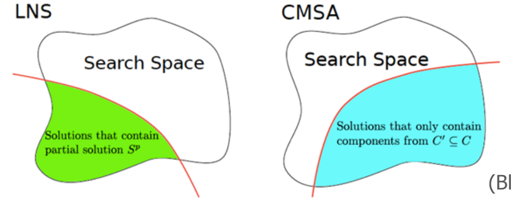
Chained LK
Success: 0%

(Veerapen, Ochoa, Tinós, Whitley. PPSN 2016)

61

## CMSA vs. LNS - matheuristics


- Construct, Merge, Solve & Adapt (CMSA)
- Large Neighbourhood Search (LNS)
- Both incorporate ILP (CPLEX) to solve *sub-instances*
- How is the sub-instance of the next iteration generated?
  - LNS - Partial destruction of the incumbent solution
  - CMSA - Generating new solutions and removing old components



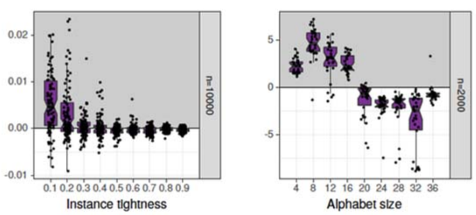

How is the original problem instance reduced?

(Blum & Ochoa, EJOR 2021)

62

## Two subset selection problems

- MDKP: Multidimensional knapsack
- MCSP: Minimum common string partition




(a) Problem: MDKP (b) Problem: MCSP

Intuition: CMSA outperforms LNS for instances for which solutions contain rather few items.

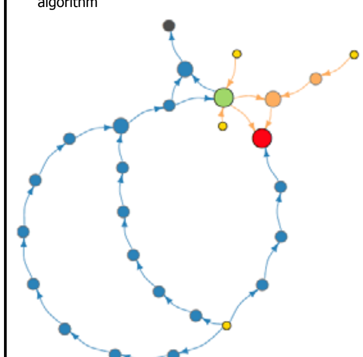
Instances with different characteristics
Improvement of CMSA over LNS (in percent)
Above zero: CMSA better, Below zero: LNS Better

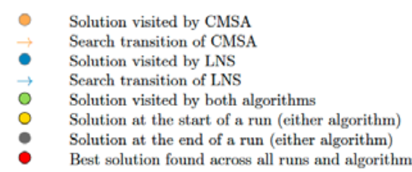
(Blum & Ochoa, EJOR 2021)

63

## Example merged LONs for an MKDP instance

Instance n = 30
3 runs for each algorithm

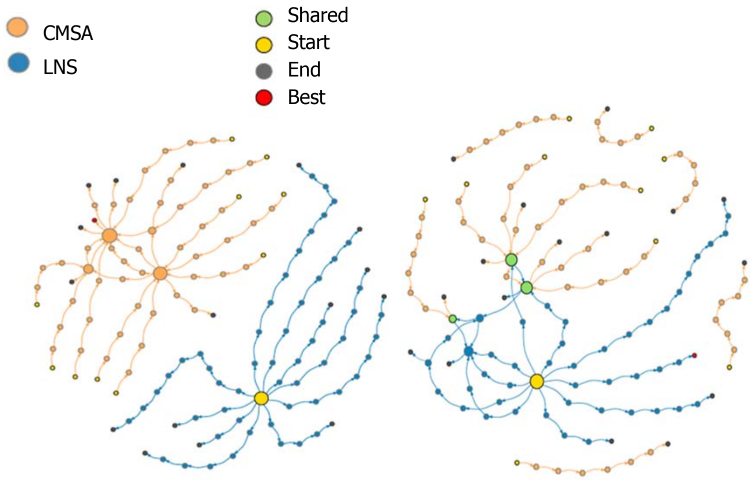





**Nodes**: (near-)optimal solutions to the sub-instances, after applying the exact solver CPLEX

**Edges**: Connect two consecutive solutions in a search trajectory of the studied algorithms. Monotonic edges.

64

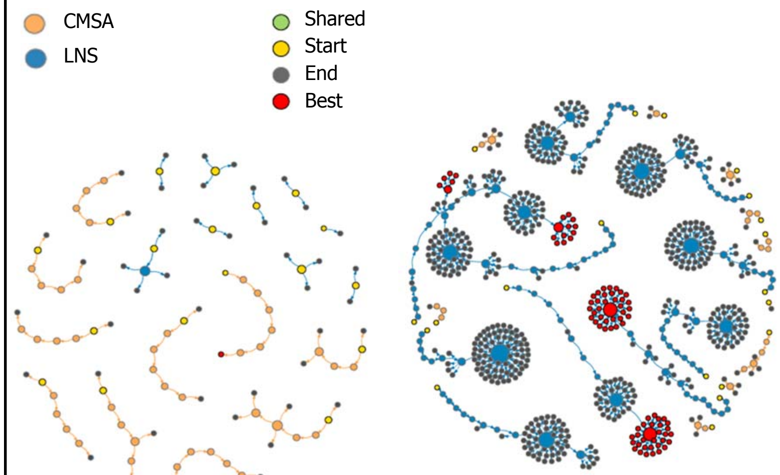



(a) $\alpha = 0.1$ (very tight).

(c) $\alpha = 0.9$ (non-tight).

MDKP $n$ = 10,000.  10 runs

65



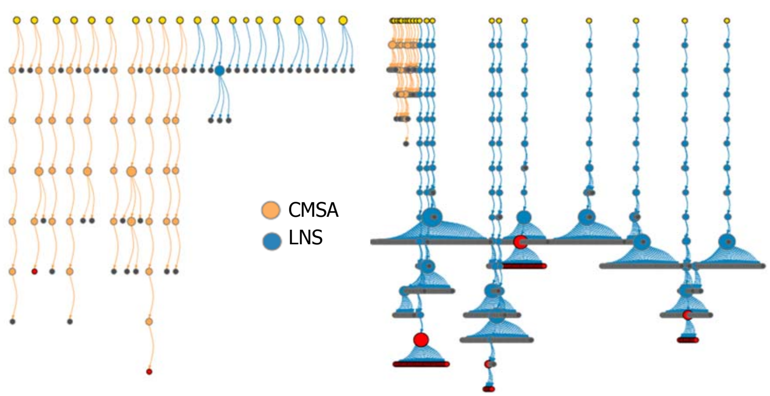

(c) $|\Sigma| = 36$ (large solutions).

(a) $|\Sigma| = 4$ (small solutions).

MCSP $n$ = 1600

66

Same networks as previous slide, but using a tree layout



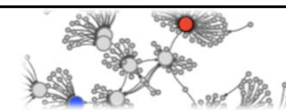

$|\Sigma| = 4$ (small solutions).

$|\Sigma| = 36$ (large solutions).

MCSP $n$ = 1600

67

# LONs



## Conclusions

- More accessible (visual) approach to heuristic understanding
- Global structure impacts search
- Real-world problems are neutral and have multiple-funnels

## Contributions

- Sampling & visualisation
- Characterisation of funnels
- Contrasting algorithms
- New LON metrics can improve performance prediction
- Using knowledge to select/configure algorithms

lonmaps.com - Website with LON resources

68

# Conclusion

- The concept of fitness landscapes originated in an evolutionary context.
- The last decade has seen the application of landscape analysis to the wider fields of optimisation and machine learning.
- New techniques have been developed:
  - Constrained landscape metrics can be used to implement landscape-aware constraint handling for metaheuristics.
  - Local multiobjective landscape features can be used to perform landscape-aware multi-objective evolutionary search.
  - Loss-gradient clouds can be used to understand the nature of neural network error surfaces for training and generalisation.
- We showed how local optima networks can be used
  - To visualise the global structure and characterise funnels
  - To contrast the behaviour of two algorithms

69

# References (1)


- Auer, P., Herbster, M. and Warmuth, M.K. (1996), Exponentially many local minima for single neurons. In *Advances in Neural Information Processing Systems*, pp. 316-322.
- Baldi, P. and Hornik, K. (1989), Neural networks and principal component analysis: Learning from examples without local minima, *Neural Networks*, 2(1), pp. 53-58.
- A. L. Barabási and M. Pósfai, *Network Science*. Cambridge University Press, 2016.
- K. D. Boese, A. B. Kahng, and S. Muddu, A new adaptive multi-start technique for combinatorial global optimizations, *Operations Research Letters*, vol. 16, no. 2, pp. 101–113, Sep. 1994.
- C. Blum, G Ochoa. A comparative analysis of two matheuristics by means of merged local optima networks, *European Journal of Operational Research* 290 (1), 36-56, 2021.
- Bosman, A.S., Engelbrecht, A.P. and Helbig, M. (2020a), Loss surface modality of feed-forward neural network architectures. In *Proceedings of International Joint Conference on Neural Networks* (IJCNN).
- Bosman, A.S., Engelbrecht, A.P. and Helbig, M. (2020b), Visualising Basins of Attraction for the Cross-Entropy and the Squared Error Neural Network Loss Functions. *Neurocomputing*, 400, pp. 113–136.
- F. Chicano, D. Whitley, G. Ochoa, R. Tinos, Optimizing One Million Variables NK Landscapes by Hybridizing Deterministic Recombination and Local Search. Genetic and Evolutionary Computation Conference, GECCO 2017: 753-760, 2017


70

# References (2)


- Choromanska, A., Henaff, M., Mathieu, M., Arous, G.B., LeCun, Y. (2015), The Loss Surfaces of Multilayer Networks. In *Proceedings of the 18th International Conference on Artificial Intelligence and Statistics*, São Paulo, Brazil, 21–25 June 2015; pp.192-204.
- J. P. K. Doye, Network Topology of a Potential Energy Landscape: A Static Scale-Free Network, *Phys. Rev. Lett.*, vol. 88, no. 23, p. 238701, 2002.
- J. P. K. Doye, M. A. Miller, and D. J. Wales, The double-funnel energy landscape of the 38-atom Lennard-Jones cluster, *The Journal of Chemical Physics*, vol. 110, no. 14, pp. 6896–6906, Apr. 1999.
- I. Gent and T. Walsh, Phase Transitions and Annealed Theories: Number Partitioning as a Case Study, in *Proceedings of the 12th European Conference on Artificial Intelligence (ECAI-96)*, 1996, pp. 170–174
- D. R. Hains, L. D. Whitley, and A. E. Howe, Revisiting the big valley search space structure in the TSP, *J Oper Res Soc*, vol. 62, no. 2, pp. 305–312, Feb. 2011.
- Hamey, L.G.C. (1998), XOR has no local minima: A case study in neural network error surface analysis. *Neural Networks*, 11(4), pp. 669-681.
- K. Helsgaun, An effective implementation of the Lin–Kernighan traveling salesman heuristic, *European Journal of Operational Research*, vol. 126, no. 1, pp. 106–130, Oct. 2000.


71

# References (3)


- P. Kerschke, M. Preuss, S. Wessing, and H. Trautmann, Detecting Funnel Structures by Means of Exploratory Landscape Analysis, *Genetic and Evolutionary Computation Conference* , GECCO, 2015, pp. 265–272.
- E. D. Kolaczyk, G. Csárdi, Statistical Analysis of Network Data with R, Springer, 2014
- Liefooghe, A., Daolio, F., Verel, S., Derbel, B., Aguirre, H. and Tanaka, K. (2019), Landscape-Aware Performance Prediction for Evolutionary Multi-objective Optimization. *IEEE Transactions in Evolutionary Computation*, 24(6).
- M. Locatelli, On the Multilevel Structure of Global Optimization Problems, *Comput Optim Applic*, vol. 30, no. 1, pp. 5–22, Jan. 2005.
- M. Lunacek and D. Whitley, The Dispersion Metric and the CMA Evolution Strategy, *Genetic and Evolutionary Computation Conference*, GECCO, 2006, pp. 477–484.
- M. Lunacek, D. Whitley, and A. Sutton, The Impact of Global Structure on Search, *Parallel Problem Solving from Nature – PPSN X*, Springer Berlin Heidelberg, 2008, pp. 498–507.
- Malan, K.M. and Engelbrecht, A.P. (2013). A survey of techniques for characterising fitness landscapes and some possible ways forward. *Information Sciences*, 241:148-163
- Malan, K.M.,Oberholzer, J.F. and Engelbrecht, A.P. (2015) Characterising Constrained Continuous Optimisation Problems. In *Proceedings of the IEEE Congress on Evolutionary Computation*, Sendai, Japan, pp 1351-1358.


72

# References (4)


- K.M. Malan (2018). Landscape-aware Constraint Handling Applied to Differential Evolution, TPNC 2018, Lecture Notes in Computer Science, vol 11324, Springer, pp 176-187.
- Malan, K.M. (2021a). A Survey of Advances in Landscape Analysis for Optimisation. *Algorithms*, 14(2), DOI:10.3390/a14020040.
- Malan, K.M. (2021b). Online Landscape Analysis for Guiding Constraint Handling in Particle Swarm Optimisation, In: Kulkarni, AJ et al. (Eds) *Constraint Handling in Metaheuristics and Applications* , Chapter 5, Springer.
- Mehta, D., Zhao, X., Bernal, E.A. and Wales, D.J. (2018), Loss surface of XOR artificial neural networks. *Physical Review E*, 97(5):052307.
- G. Ochoa, M. Tomassini, S. Vérel, and C. Darabos, A Study of NK Landscapes' Basins and Local Optima Networks, *Conference on Genetic and Evolutionary Computation,* GECCO, 2008, pp. 555–562.
- G. Ochoa and N. Veerapen, Deconstructing the Big Valley Search Space Hypothesis, *Evolutionary Computation in Combinatorial Optimization* EvoCOP, Springer International Publishing, 2016, pp. 58–73.
- G. Ochoa and N. Veerapen, Mapping the global structure of TSP fitness landscapes, *J Heuristics* J. Heuristics 24(3): 265-294, 2018.




# References (5)


- G. Ochoa, N. Veerapen, F. Daolio, and M. Tomassini, Understanding Phase Transitions with Local Optima Networks: Number Partitioning as a Case Study, *Evolutionary Computation in Combinatorial Optimization,* EvoCOP, 2017, Springer vol. 10197, pp. 233–248.
- Rice, J.R. (1976), The Algorithm Selection Problem, *Advances in Computers*, Vol 15, Elsevier, pp. 65-118.
- D. Simoncini, S. Barbe, T. Schiex, S. Verel. Fitness landscape analysis around the optimum in computational protein design. Genetic and Evolutionary Computation Conference, GECCO 2018.
- N. Veerapen, G. Ochoa, R. Tinós, and D. Whitley, Tunnelling Crossover Networks for the Asymmetric TSP, *Parallel Problem Solving from Nature – PPSN XIV*, Springer International Publishing, 2016, pp. 994–1003.
- Verel, S., Liefooghe A. and Dhaenens, C. (2011). Set-based multiobjective fitness landscapes: A preliminary study. In *Proceedings of the 13th annual Conference on Genetic and Evolutionary Computation*, Dublin, Ireland, 12–16 July 2011; pp. 769–776.
- Wright, S. (1932), The Roles of Mutation, Inbreeding, Crossbreeding, and Selection in evolution, In *Proceedings of the Sixth International Congress on Genetics*, pp. 356–366.