Evolutionary Search and Fitness Landscapes of Pseudo-Boolean Problems

Doctoral dissertation

Xavier F. C. Sánchez Díaz

Department of Computer Science
Faculty of Information Technology and Electrical Engineering
Norwegian University of Science and Technology

June 11, 2026

Slides available

https://s.ntnu.no/phd-slides

https://s.ntnu.no/phd-slides

Outline

Section 1

What this dissertation is about

In this dissertation...

We deal with descriptive theory of **search landscapes** of **multimodal** **pseudo-Boolean** functions. We study these landscapes using **evolutionary computation**.

What this dissertation is about

We deal with descriptive theory of **search landscapes** of **multimodal pseudo-Boolean** functions. We study these landscapes using **evolutionary computation**.

Descriptive theory: the use of mathematical notation to describe, measure, or quantify observations [Doe24].

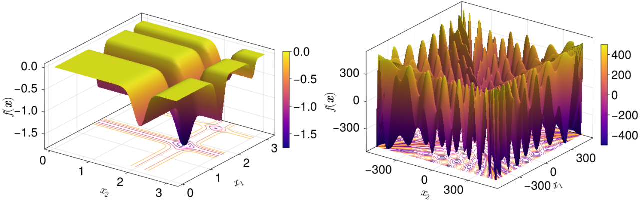
- Example: Fitness landscape analysis, schema theory...

We also use experimentally guided theory: setting up an artificial experiment, or experimentally analyse a particular question [Doe24].

What this dissertation is about

We deal with descriptive theory of **search landscapes** of **multimodal pseudo-Boolean functions**. We study these landscapes using **evolutionary computation**.

A **search landscape** is a visual representation of the solution space as it is navigated by a search method or optimisation algorithm [OM24].



(a) Michalewicz function (b) Rana’s function

Figure: Two optimisation landscapes: 2-dimensional continuous functions. The y-axis is used to plot the objective function, while the x- and z-axes are function parameters (shown here as x_1 and x_2).

What this dissertation is about

We deal with descriptive theory of **search landscapes** of **multimodal pseudo-Boolean** functions. We study these landscapes using **evolutionary computation**.

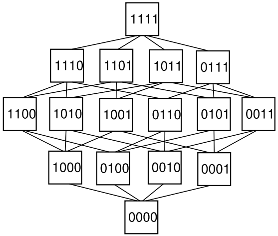
We focus on **pseudo-Boolean** functions (PBFs) [RS02].

- i.e. optimisation and search problems with objective functions of the form $f : \mathbb{B}^n \to \mathbb{R}_{\geq 0}$

With a special interest in those with **multiple solutions** [LEDE17].

So search spaces look more like this $\longrightarrow$

and usually contain many optima



What this dissertation is about

We deal with descriptive theory of **search landscapes** of **multimodal pseudo-Boolean** functions. We study these landscapes using **evolutionary computation**.

We use **evolutionary algorithms** (EAs) as our heuristic search methods [Gol89]. Broadly, an EA can be defined as follows [Leh06]. Given a PBF $f : \mathbb{B}^n \to \mathbb{R}_{\geq 0}$,

1. Initialise a population P with random elements from $\mathbb{B}^n$
2. **While** stopping criterion is not met, **do**
   - 2.1 Select parents from P using fitness function f
   - 2.2 Create a set of offspring C by recombining parents
   - 2.3 Stochastically mutate individuals in C
   - 2.4 Select new individuals from $P \cup C$ using f to form a new population P
3. **return** the best individuals in P according to f.

Section 2

How research was carried out

Research Context

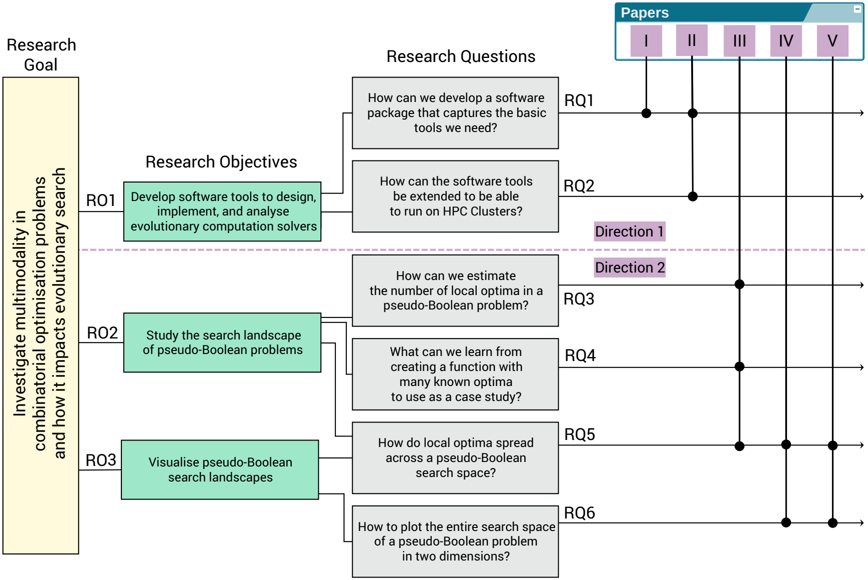
Research Structure

This dissertation encompasses research conducted from October 2021 to September 2025. The work was carried out in two research directions, comprising three research objectives and six research questions in total. Our results lead to five publications, all of which have been presented at both national and international venues.

Research Direction 1
: Aimed at developing a code base for quick prototyping and testing of EAs.

Research Direction 2
: Focused on the analysis of search landscapes of multimodal PBFs.

Overview of Research Structure



Section 3

Summary of Publications

Paper I: EvoLP.jl: A Playground for Evolutionary Computation

Norwegian AI Society Symposium 2023 – Bergen, Norway

EvoLP.jl

- Presents a Julia package for designing and implementing EAs through **building blocks**. *LP* comes from *Lekeplass*—"playground" in Norwegian.
- An extensible computational unit for each step in the evolution process
  - Initialisation → population generator: *ind* → *repr*
  - Selection → selector: $P \to ind$
  - Crossover → recombinator: $ind \times ind \to ind$
  - Mutation → mutator: $ind \to ind$
  - Survival is usually algorithm dependent (but selectors can be used)
- It also includes test functions, results reporting, statistics computing, and a few built-in solvers for toy examples.

Why Julia?

Paper I: EvoLP.jl: A Playground for Evolutionary Computation

- It is fast!
  - Type-inference and JIT compilation using LLVM
  - **1.54 petaFLOPs** on 1.3M threads over 9K KNL Nodes [Jul17]
- Solves the *two-language* problem [BEKS17]—one languange instead of:
  - a simple language for scripting
  - another performant language for production
- Multiple dispatch allows for a **natural mapping** of evolutionary computation concepts to code

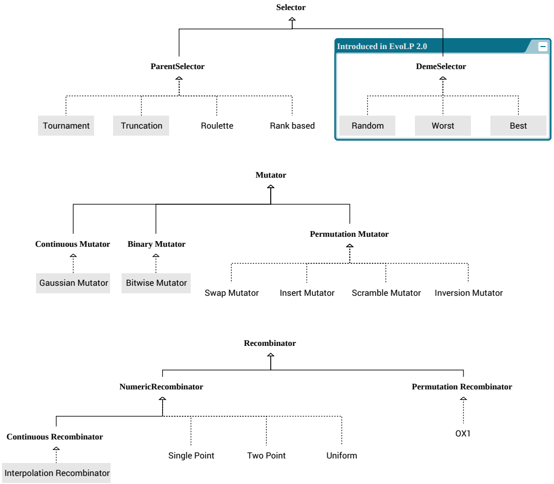
A high-level, high-performance programming language for scientific computing

Paper I: EvoLP.jl: A Playground for Evolutionary Computation

Norwegian AI Society Symposium 2023 – Bergen, Norway

RQ1: How can we develop a software package that captures the basic tools we need?

- By **formalising** a type hierarchy [KW19] that closes the gap between modelling and implementation.
- Taking advantage of Julia's multiple dispatch—one function, several implementations depending on parameter types [Kwo].



Paper II: Evolutionary Computation with Islands: Extending EvoLP.jl for Parallel Computing

Norsk IKT-konferanse for forskning og utdanning 2023 – Stavanger, Norway

EvoLP.jl

NIKT
Norwegian ICT conference
for research and education
NIKT2023

- Presents an extension for EvoLP.jl to run island models on **HPC systems** [SM23b]
  - Uses the Message Passing Interface (MPI) communication scheme [Mes21]
- Based on the generalised island model of Izzo et al. [IRB12]. Given an archipelago (set of islands and their topology), each island will:
  1. Run an EA with its parameters
  2. Select a deme $\mathbb{M}$ from P using a selection policy
  3. **Send** $\mathbb{M}$ to adjacent islands w.r.t the topology
  4. **Receive** $\mathbb{M}'$ individuals from adjacent islands
  5. **Reinsert** $\mathbb{M}'$ into the island's P using a replacement policy
  6. Continue runing until termination/sync

Steps 3 and 4 are mapped directly with MPI directives!

Paper II: Evolutionary Computation with Islands: Extending EvoLP.jl for Parallel Computing

Norsk IKT-konferanse for forskning og utdanning 2023 – Stavanger, Norway

RQ2: How can the software tools be extended to be able to run on HPC Clusters?

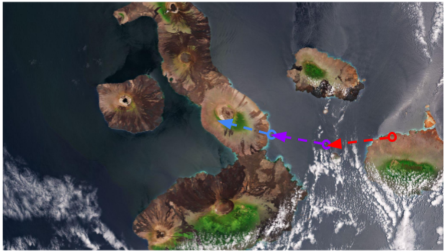
- By **formalising** the communication and **implementing** appropriate blocks [SM23a].
- Exploiting convergence and diversity preservation benefits from communication [ES15].
- Tested on both **local** hardware and NTNU's **HPC system**, Idun [SJTR22]

drift operator
Send individuals from one island to another

strand operator
Receive individuals from another island

reinsert! operator
Reinsert individuals into population according to a given policy

The island model



The Galapagos Islands, Copernicus Sentinel (2020), ESA

Paper II: Evolutionary Computation with Islands: Extending EvoLP.jl for Parallel Computing

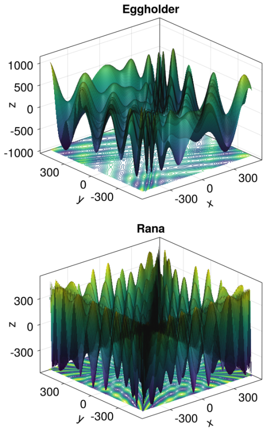
Norsk IKT-konferanse for forskning og utdanning 2023 – Stavanger, Norway

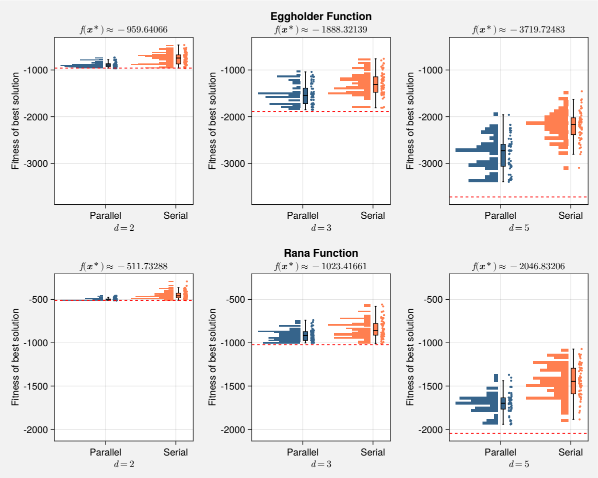
- **Test 1**: Empirical comparison of serial vs parallel island model approaches, using a unidirectional ring topology with 64 islands. Each island with a steady state GA with random deme selection policy, and replace-worst policy.
  - Ackley, Rosenbrock, and Michalewicz test functions, for dimensions $d \in \{2, 5, 10\}$
  - Migration every 10 iterations (5 for Michalewicz)
  - Results hint at better convergence to the known optima when run **on parallel**.

- **Test 2**: Empirical comparison of the impact of exploration on island models on highly multimodal test functions. Same topology.
  - Eggholder and Rana test functions, for dimensions $d \in \{2, 3, 5\}$
  - Migration every 10 generations
  - Results show smaller spread of fitnesses across populations.

Implemented as an **extension** in EvoLP.jl in Oct. 2025:
`https://ntnu-ai-lab.github.io/EvoLP.jl/stable/man/islands.html`

Paper II: Test 2





- Higher median of best solutions on serial approach
- *Better exploration* overall using the parallel approach

Subsection 2

Research Direction II — Landscape Analysis

Paper III: Estimating the Number of Local Optima in Multimodal Pseudo-Boolean Problems: Validation via Landscapes of Triangles

Genetic and Evolutionary Computation Conference 2024 – Melbourne, Australia

Introduces $\mathrm{Triangle}$, a synthetic PBF with an **adjustable number of optima**.

- Dependent on the interaction of its parameters [SM24]
- We use it to estimate the number of optima in landscapes with different structures

- The estimation procedure is based on the Birthday Paradox [Mat91]
  - First studied by Caruana and Mullin [CM00] for continuous optimization
  - Later studied by Hernando et al. [HML13] for TSP

Given 365 days in a year, how many people in a room do we need to find a repeated birthday with $p \geq 0.5$?

Paper III: Estimating the Number of Local Optima in Multimodal Pseudo-Boolean Problems: Validation via Landscapes of Triangles

Genetic and Evolutionary Computation Conference 2024 – Melbourne, Australia

- RQ3: How can we estimate the number of local optima in a PBF?
- RQ4: What can we learn from creating a function with many known optima to use as a case study?
- RQ5: How do local optima spread across a pseudo-Boolean search space?

By using the *reverse* birthday paradox:

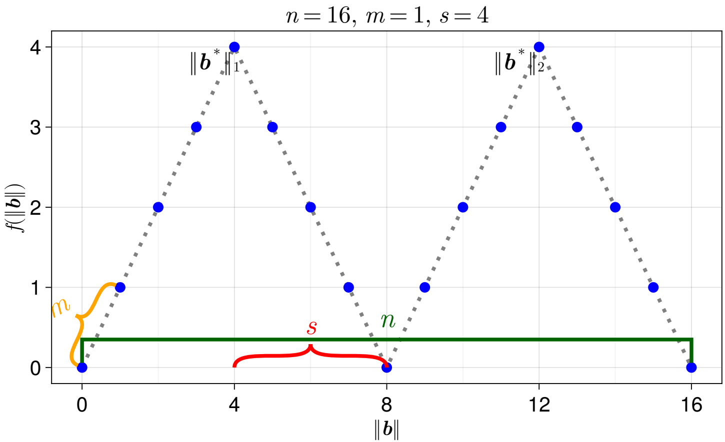
Given this many **people** in a room with a repeated birthday, how many **days** are there in a *year*?

- **Sampling procedures** are essential for landscape analysis
- The size and distribution of **basins of attraction** affect algorithm behaviour

Paper III: Estimating the Number of Local Optima in Multimodal Pseudo-Boolean Problems: Validation via Landscapes of Triangles

Genetic and Evolutionary Computation Conference 2024 – Melbourne, Australia

- n is the bitstring length
  - There are nCk bitstrings at $\|\boldsymbol{b}\| = k$
- s is a segmenting factor
  - or *the step*
  - number of bit flips to optimum
  - determines peak width
- m is a scaling factor
  - or *the slope*
  - scales fitness
  - can introduce deception



Triangle($\boldsymbol{b}, m = 1, s = 2$), $\boldsymbol{b} \in \mathbb{B}^4$

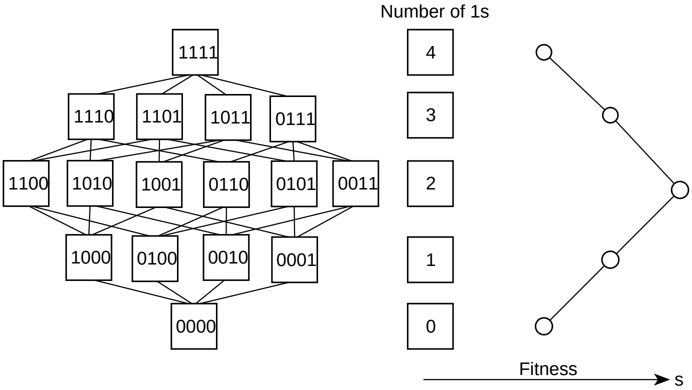
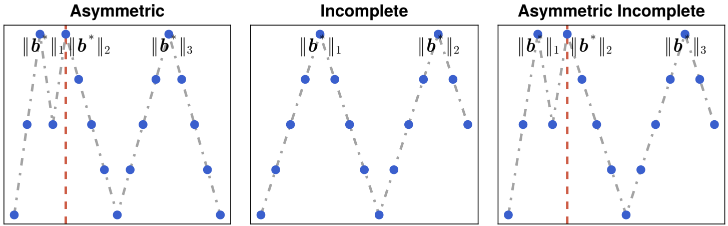


Figure: The search space of a Triangle function with $n = 4$, $m = 1$, and $s = 2$, and how each state relates to its objective function. Note no edge exists between two states with the same fitness.

Paper III: Estimating the Number of Local Optima in Multimodal Pseudo-Boolean Problems: Validation via Landscapes of Triangles

Genetic and Evolutionary Computation Conference 2024 – Melbourne, Australia



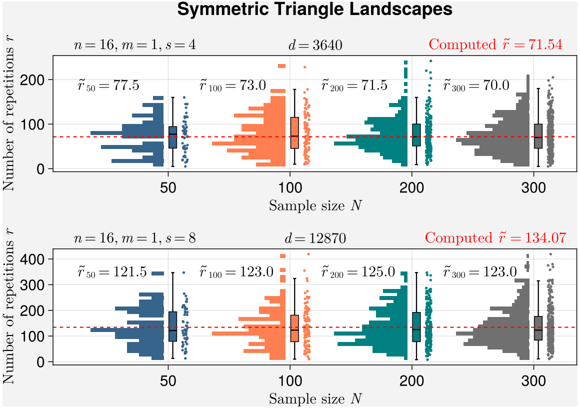
- We can create several landscape structures using parameter interaction
- One landscape from several Triangle functions
  - piecewise definition
  - different *Region* of attraction sizes

Two experiments using a 1+1-EA for sampling, using random initialisation:

1) **Symmetric** landscapes
2) **Asymmetric**, **Incomplete**, and **Asymmetric Incomplete** landscapes

Symmetric Landscapes

Paper III: Experiment 1 — Estimating Local Optima in Symmetric Landscapes



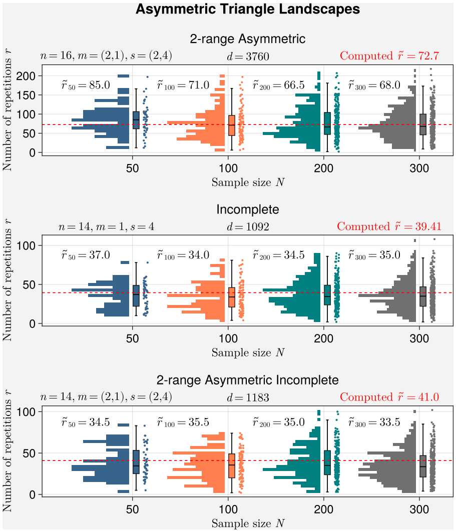
- The estimation procedure is noticeably accurate on **symmetric landscapes** (with uniformly distributed optima)
  - Even with a small number of samples N, empirical estimates (medians) are close to the computed estimate
  - Large landscapes are naturally more difficult to sample evenly

Asymmetric Landscapes

Paper III: Experiment 2 — Estimating Local Optima in Asymmetric Landscapes

- **Incomplete** landscapes ($n \operatorname{div} s \notin \mathbb{Z}$) have larger difference in the size of their regions of attraction
  - Estimates are still relatively accurate, although not as in symmetric landscapes
- Estimates on **asymmetric** landscapes are more accurate than those in incomplete landscapes
  - Less accurate than symmetric landscapes

Useful for **budgetting** purposes!



Paper IV: Regularized Feature Selection Landscapes: An Empirical Study of Multimodality

Parallel Problem Solving from Nature XVIII, 2024 – Hagenberg, Austria

Studies the effect of regularisation on the search landscape of **feature selection** (a real-world, highly multimodal PBF [MMOE19]).

- We study 10 classification datasets from UCI [KRK]
  - Paper IV focuses on two of them: *Glass identification* [Ger87] and *Heart Disease (Cleveland)* [And89]
  - Here we show only a subset of the results for brevity
- Introduces the **Hinged-Bitstring Map (HBM)** [SMM24]
  - Visual overview of the entire pseudo-Boolean landscape
  - Highlights the distribution of optima accross the landscape
  - Works in combination with **Local-Optima Networks** [SMM24; MSM25]

Paper IV: Regularized Feature Selection Landscapes: An Empirical Study of Multimodality

Parallel Problem Solving from Nature XVIII, 2024 – Hagenberg, Austria

- RQ5: How do local optima spread across a pseudo-Boolean search space?
- RQ6: How to plot the entire search space of a PBF in two dimensions?

Precomputing the entire landscape to **count** the optima and **visualise** their distribution using HBMs [SMM24], LONs [OTVD08; OVDT14], and fitness-distance correlation [SM25].

Three experiments:

1. Impact of varying regularisation
2. Visualising regularisation
3. Visualising connectedness and optima distribution

Paper IV: Regularized Feature Selection Landscapes: An Empirical Study of Multimodality

Parallel Problem Solving from Nature XVIII, 2024 – Hagenberg, Austria

Feature selection [SMM24]

Consider a bitstring $\boldsymbol{b} = b_1, \ldots b_n$ indicating which features are included ($b_i = 1$) or not ($b_i = 0$). We model the feature selection problem as an *energy* function to **minimize**:

$$h(\boldsymbol{b}) = h_E(T(\boldsymbol{b})) + \epsilon \cdot h_P(\boldsymbol{b}),$$

where $h_E(T(\boldsymbol{b}))$ is the **classification error** over a given dataset using a model T, $h_P(\boldsymbol{b})$ is a penalty depending on the number of features used for training with feature subset $\boldsymbol{b}$, and ϵ controls the degree of regularisation.

For this paper, T is a **decision tree** trained on all 2^n feature subsets.

Paper IV: Regularized Feature Selection Landscapes: An Empirical Study of Multimodality

Parallel Problem Solving from Nature XVIII, 2024 – Hagenberg, Austria

Exhaustive enumeration - landscape precomputing

1. Enumerate all bitstrings $\boldsymbol{b}_i, i \in \{1, 2, \ldots, 2^n - 1\}$
   - e.g., bitstring $\boldsymbol{b}_7 = 00000111$, with $n = 8$
   - i.e., «*using the first three features, not using the remaining five*»
2. Train a decision tree using $T(\boldsymbol{b}_i)$ and store its classification error
   - using 70% of the data for training
   - testing on the remaining 30%
3. Use the classification error as a surrogate for the *fitness*
4. Calculate regularisation with varying degrees of $\epsilon \in \{0, 1/8, 1/16, 1/32\}$

Massive amounts of parallel compute provided by NTNU's **HPC system**, Idun [SJTR22].
The surrogate tables are available in UiT's Dataverse.NO [Sán24].

Data

Paper IV — Experiment 1: Impact of Varying Regularisation

Table: Datasets used in Paper IV, sorted by number of features (n). We show the number of examples m, and the number of local $|\boldsymbol{L}|$ and global optima $|\boldsymbol{G}|$ for various values of the regularization term ϵ.

Name	n	m	**Number of optima**							
			$\epsilon = 0$		$\epsilon = 1/32$		$\epsilon = 1/16$		$\epsilon = 1/8$	
			$\|\boldsymbol{L}\|$	$\|\boldsymbol{G}\|$	$\|\boldsymbol{L}\|$	$\|\boldsymbol{G}\|$	$\|\boldsymbol{L}\|$	$\|\boldsymbol{G}\|$	$\|\boldsymbol{L}\|$	$\|\boldsymbol{G}\|$
1-seeds	7	210	20	5	14	1	6	1	7	1
2-e-coli	7	336	17	1	7	1	7	1	6	1
3-breast-w	9	699	65	2	6	1	9	1	9	1
4-glass	9	214	65	1	51	2	22	2	7	2
5-heart-c	13	303	700	1	407	1	117	1	13	1
6-wine	13	178	976	58	286	2	29	3	14	3
7-credit-a	15	690	2511	4	351	1	18	1	15	1
8-zoo	16	101	10862	8275	3003	1	134	1	16	1
9-letter-r	16	20000	3964	1	16	1	16	1	16	1
10-hepatitis	19	155	50985	2284	23577	3	7621	3	133	3

The Hinged-Bitstring Map (HBM)

Paper IV — Experiment 2: Visualising Regularisation

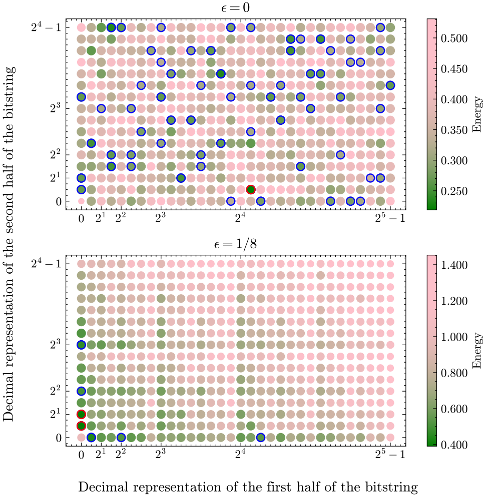
HBM construction

For each bitstring $\boldsymbol{b}_i \in \mathbb{B}^n$

1. Split $\boldsymbol{b}_i$ into halves[a] $\boldsymbol{b}_{i_1} \| \boldsymbol{b}_{i_2}$
   - e.g., $\boldsymbol{b}_{167} = 1010 \| 0111$
2. Use $\mathrm{Dec}(\boldsymbol{b}_{i_1})$, as the x-coordinate
3. Use $\mathrm{Dec}(\boldsymbol{b}_{i_2})$ as the y-coordinate
4. Plot $\boldsymbol{b}_i$ at (x, y) as a scatter plot, using colour for fitness
   - i.e., $\boldsymbol{b}_{167} \mapsto (10, 7)$
5. Highlight global and local optima

[a] in favour of $\boldsymbol{b}_{i_1}$ if n is odd

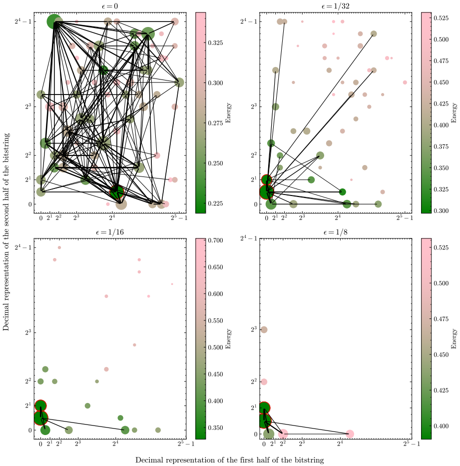
Figure: *Glass Identification* ($\epsilon \in \{0, 1/8\}$)



LONs and HBMs

Paper IV — Experiment 3: Connectedness and Optima Distribution

Figure: *Glass identification* ($\epsilon \in \{0, 1/32, 1/16, 1/8\}$)



Local Optima Network (LON)

A graph representation of a **multimodal** landscape [OTVD08; OVDT14].

- Nodes are **local optima**
- Edges represent connections (or the *probability of a connection*) between basins
- Node size is proportional to basins of attraction size

Regularisation transforms the landscape:

- Optima disappear, basins are split, and connectivity decreases

Paper V: Visualizing Pseudo-Boolean Functions: Feature Selection and Regularization for Machine Learning

25th European Conference on Evolutionary Computation in Combinatorial Optimization, 2025 – Trieste, Italy

evo* 2025

Delves further into visualisation of PBFs through both synthetic and **feature selection** problems.

- We study four additional datasets from UCI [KRK]
  - Paper V highlights *Dry Bean* [KÖ20] and *Bank Marketing* [MRC12]
  - In contrast to Paper IV, here we use a random forest classifier (with $\texttt{n\_trees} = 10$)
- This paper also formalises the **Hinged-Bitstring Map (HBM)** [SMM24]
  - Proof of the impossibility of keeping neighbourhood structure in $\mathbb{R}^2$
    - Based on the work by Masson [Mas24]
  - Introduces the LON+HBM [MSM25; SM25]

Paper V: Visualizing Pseudo-Boolean Functions: Feature Selection and Regularization for Machine Learning

25th European Conference on Evolutionary Computation in Combinatorial Optimization, 2025 – Trieste, Italy

- RQ5: How do local optima spread across a pseudo-Boolean search space?
- RQ6: How to plot the entire search space of a PBF in two dimensions?

In a similar fashion to Paper IV, we precompute the entire landscape to **count** the optima and **visualise** their distribution using HBMs [SMM24] and LONs [OTVD08; OVDT14].

Two case studies:

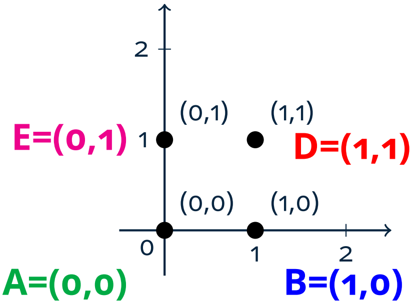
1. Toy problem in $\mathbb{B}^6$
2. Real-world case: **feature selection** with regularisation

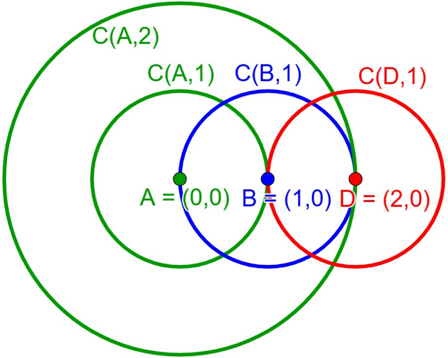
Preserving Neighbourhood Structure

Paper V — Analysis of Visualisation

Lemma

There is no isometric embedding from $\mathbb{B}^2$ into $\mathbb{R}^2$ that preserves the Hamming distance as the Euclidean distance [Mas24].





Enhancing the HBM

Paper V — Analysis of Visualisation

Since there is no way around the neighbourhood distortion in 2D, we may as well enhance our visualisation. An HBM can be enhanced by combining it with a LON [SM25]:

- Using **arrows** to show connectedness between optima (like a LON)
- Using **size** to convey the size of basin of attraction (like a LON)
- Using the **position** resulting from the coordinate system of the HBM

Using the aesthetic and geometric elements of both visualisations [Wil05], they can be merged by superimposing one on top of the other [JE12].

HBM+

Paper V — Case Study 1: A toy problem in $\mathbb{B}^6$: $\sin(2\mathrm{Dec}(b))$

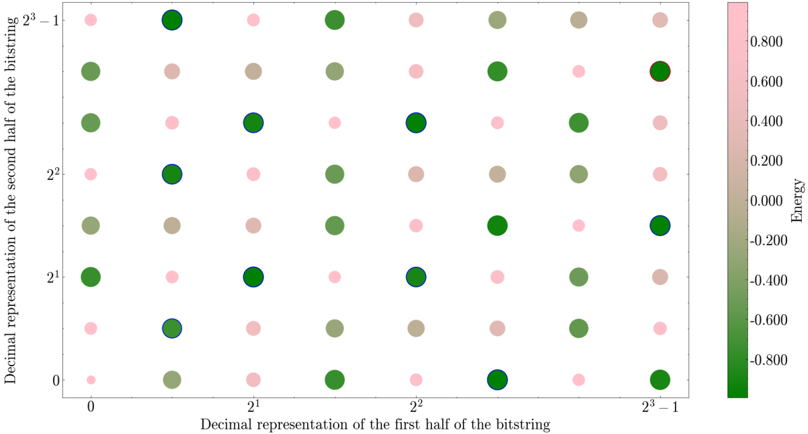


Figure: An HBM for $\sin(2\mathrm{Dec}(b))$, altering the **size** of local and global optima.

HBM+Arrows

Paper V — Case Study 1: A toy problem in $\mathbb{B}^6$: $\sin(2\text{Dec}(b))$

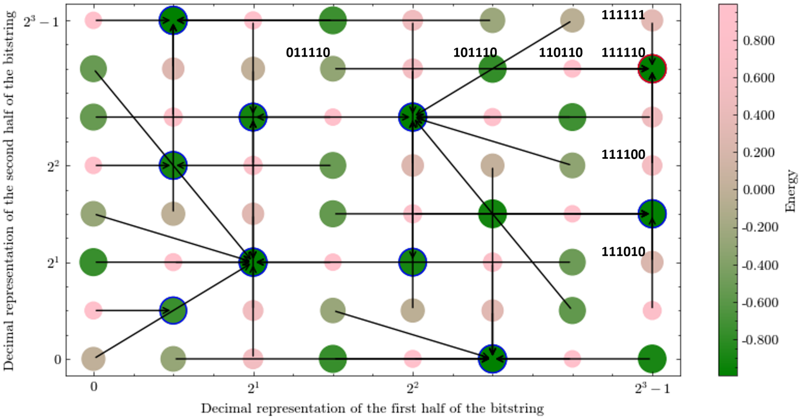


Figure: An HBM for $\sin(2\text{Dec}(b))$, altering the **size** aesthetic and adding **arrow** geometries for connectivity between basins.

HBM+LON

Paper V — Case Study 1: A toy problem in $\mathbb{B}^6$: $\sin(2\mathrm{Dec}(b))$

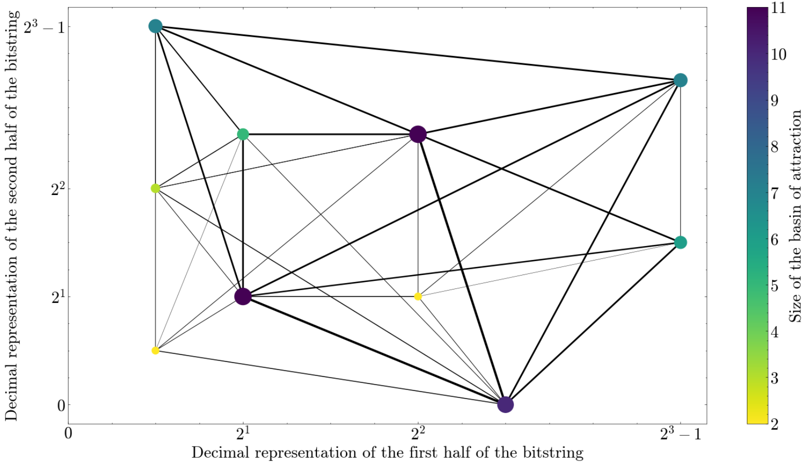


Figure: An HBM+LON for $\sin(2\mathrm{Dec}(b))$. Nodes represent specific bitstrings.

Data

Paper V — Case Study 2: Feature Selection and Regularisation

Table: Datasets used in Paper V. We present number of features (n), number of examples (m) and number of classes.

Name	Subject Area	n	m	Number of Classes
Yeast [Nak96]	Biology	8	1484	10
HTRU2 [Lyo17]	Physics and Chemistry	8	17898	2
Dry Bean [KÖ20]	Biology	16	13611	7
Bank Marketing [MRC12]	Business	16	45211	2

Reminder: Feature Selection

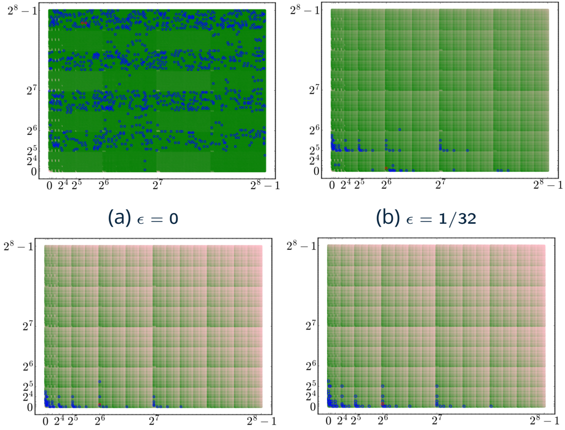
$$h(\boldsymbol{b}) = h_E(T(\boldsymbol{b})) + \epsilon \cdot h_P(\boldsymbol{b})$$

For this paper, T is a **random forest** classifier trained on all 2^n feature subsets, 70%/30% train/test split, using the default parameters in scikit-learn.

Visualising Regularisation

Paper V — Case Study 2: Feature Selection and Regularisation

Figure: Classification error of T on *Dry Bean* with regularisation values $\epsilon \in \{0, 1/32, 1/16, 1/8\}$.



(a) $\epsilon = 0$ (b) $\epsilon = 1/32$

(c) $\epsilon = 1/16$ (d) $\epsilon = 1/8$

Similar findings as in Paper IV!

- **Number of optima** diminishes drastically
- The model prefers bitstrings with fewer 1s
- **Data preprocessing and treatment** affect the number and distribution of optima over the landscape
- Several repetitions are necessary when dealing with stochastic classifiers

Section 4

Concluding Remarks

Contributions

Concluding Remarks

Some of the contributions from our research include:

1. A **software framework**: EvoLP.jl
   - and its extension for HPC!
2. A **new multimodal PBF**: Triangle
3. A novel **visualisation technique** for PBFs: HBM
4. A series of **precomputed fitness landscapes** (*lookup tables*)

Limitations and Future Work

Concluding Remarks

...and there plenty of ways to continue our work!

- **Extending HBMs**
  - Build *sparse* HBMs from samples
  - HBMs and their mapping
  - HBMs to visualise evolutionary operators
  - HBMs to represent formulas or operators in propositional logics
- **Formal analysis** of Triangle
- Testing other **sampling procedures** for local optima estimation
- Using **different ML models** for feature selection landscapes
  - Computing repetitions and logging compute time
  - Consider regularisation as a multiple-objective problem

That's it.

The thesis is available online

at the Norwegian Research Information Repository (NVA)

https://s.ntnu.no/phd-thesis

https://s.ntnu.no/phd-thesis

References I

[And89] W. S. Andras Janosi, *Heart Disease*, UCI Machine Learning Repository, 1989. DOI: 10.24432/C52P4X Accessed: Mar. 19, 2026. [Online]. Available: https://archive.ics.uci.edu/dataset/45

[BEKS17] J. Bezanson, A. Edelman, S. Karpinski, and V. B. Shah, "Julia: A Fresh Approach to Numerical Computing," *SIAM Review*, vol. 59, no. 1, pp. 65–98, Jan. 2017, ISSN: 0036-1445. DOI: 10.1137/141000671 Accessed: Sep. 25, 2025. [Online]. Available: https://epubs.siam.org/doi/10.1137/141000671

[CM00] R. Caruana and M. Mullin, "Estimating the Number of Local Minima in Big, Nasty Search Spaces," in *Proceedings of IJCAI-99 Workshop on Statistical Machine Learning for Large-Scale Optimization*, Stockholm, Sweden, Mar. 25, 2000. [Online]. Available: https://www.researchgate.net/publication/2626107_Estimating_the_Number_of_Local_Minima_in_Big_Nasty_Search_Spaces

[Doe24] B. Doerr, "A Gentle Introduction to Theory (for Non-Theoreticians)," in *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, ser. GECCO '24 Companion, New York, NY, USA: Association for Computing Machinery, Aug. 1, 2024, pp. 800–829, ISBN: 979-8-4007-0495-6. DOI: 10.1145/3638530.3648402 Accessed: Apr. 22, 2025. [Online]. Available: https://dl.acm.org/doi/10.1145/3638530.3648402

[ES15] A. Eiben and J. Smith, *Introduction to Evolutionary Computing* (Natural Computing Series). Berlin, Heidelberg: Springer, 2015, ISBN: 978-3-662-44873-1 978-3-662-44874-8. DOI: 10.1007/978-3-662-44874-8 Accessed: Mar. 19, 2023. [Online]. Available: https://link.springer.com/10.1007/978-3-662-44874-8

References II

[Ger87] B. German, *Glass Identification*, UCI Machine Learning Repository, 1987. DOI: 10.24432/C5WW2P Accessed: Mar. 19, 2026. [Online]. Available: https://archive.ics.uci.edu/dataset/42

[Gol89] D. E. Goldberg, *Genetic Algorithms in Search, Optimization and Machine Learning*, 1st ed. USA: Addison-Wesley Longman Publishing Co., Inc., 1989, 372 pp., ISBN: 978-0-201-15767-3.

[HML13] L. Hernando, A. Mendiburu, and J. A. Lozano, "An evaluation of methods for estimating the number of local optima in combinatorial optimization problems," *Evolutionary Computation*, vol. 21, no. 4, pp. 625–658, Nov. 1, 2013, ISSN: 1063-6560. DOI: 10.1162/EVCO_a_00100 Accessed: Jan. 28, 2024. [Online]. Available: https://dl.acm.org/doi/10.1162/EVCO_a_00100

[IRB12] D. Izzo, M. Ruciński, and F. Biscani, "The Generalized Island Model," in *Parallel Architectures and Bioinspired Algorithms*, ser. Studies in Computational Intelligence, F. Fernández de Vega, J. I. Hidalgo Pérez, and J. Lanchares, Eds., Berlin, Heidelberg: Springer, 2012, pp. 151–169, ISBN: 978-3-642-28789-3. DOI: 10.1007/978-3-642-28789-3_7 Accessed: Apr. 20, 2023. [Online]. Available: https://doi.org/10.1007/978-3-642-28789-3_7

[JE12] W. Javed and N. Elmqvist, "Exploring the design space of composite visualization," in *2012 IEEE Pacific Visualization Symposium*, Feb. 2012, pp. 1–8. DOI: 10.1109/PacificVis.2012.6183556 Accessed: Apr. 7, 2025. [Online]. Available: https://ieeexplore.ieee.org/document/6183556

[Jul17] Julia Computing, "Julia Joins Petaflop Club," HPCwire, Accessed: Nov. 6, 2025. [Online]. Available: https://www.hpcwire.com/off-the-wire/julia-joins-petaflop-club/

References III

[KRK] Kelly Markelle, Rachel Longjohn, and Kolby Nottingham, "UCI Machine Learning Repository," UCI Machine Learning Repository, Accessed: Jan. 29, 2024. [Online]. Available: https://archive.ics.uci.edu/

[KW19] M. J. Kochenderfer and T. A. Wheeler, *Algorithms for Optimization*. Mit Press, Mar. 12, 2019, 520 pp., ISBN: 978-0-262-35140-9. [Online]. Available: https://algorithmsbook.com/optimization/

[KÖ20] M. Koklu and I. A. Özkan, *Dry Bean*, UCI Machine Learning Repository, 2020. DOI: https://doi.org/10.24432/C50S4B [Online]. Available: https://archive.ics.uci.edu/dataset/602/dry+bean+dataset

[Kwo] T. Kwong, *Hands-On Design Patterns and Best Practices with Julia*, ISBN: 978-1-83864-881-7. Accessed: May 1, 2026. [Online]. Available: https://www.oreilly.com/library/view/hands-on-design-patterns/9781838648817/

[Leh06] P. K. Lehre, "Complexity and Geometry in Artificial Development," Norwegian University of Science and Technology, 2006, 242 pp.

[LEDE17] X. Li, M. G. Epitropakis, K. Deb, and A. Engelbrecht, "Seeking Multiple Solutions: An Updated Survey on Niching Methods and Their Applications," *IEEE Transactions on Evolutionary Computation*, vol. 21, no. 4, pp. 518–538, Aug. 2017, ISSN: 1941-0026. DOI: 10.1109/TEVC.2016.2638437 Accessed: Sep. 25, 2023. [Online]. Available: https://ieeexplore.ieee.org/document/7782373

References IV

[Lyo17] R. Lyon, *HTRU2*, UCI Machine Learning Repository, 2017. DOI: https://doi.org/10.24432/C5DK6R [Online]. Available: https://archive.ics.uci.edu/dataset/372/htru2

[Mas24] C. Masson, "Memetic and Multimethod Stochastic Optimization Algorithms and Landscape Analysis: An Empirical Study," M.S. thesis, NTNU, 2024. Accessed: Jan. 27, 2025. [Online]. Available: https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/3154374

[MSM25] C. Masson, X. F. C. Sánchez-Díaz, and O. J. Mengshoel, "Visualizing Pseudo-Boolean Functions: Feature Selection and Regularization for Machine Learning," in *Evolutionary Computation in Combinatorial Optimization*, M. S. Krejca and M. Wagner, Eds., Cham: Springer Nature Switzerland, 2025, pp. 150–166, ISBN: 978-3-031-86849-8. DOI: 10.1007/978-3-031-86849-8_10

[Mat91] F. H. Mathis, "A Generalized Birthday Problem," *SIAM Review*, vol. 33, no. 2, pp. 265–270, Jun. 1991, ISSN: 0036-1445. DOI: 10.1137/1033051 Accessed: Oct. 10, 2023. [Online]. Available: https://epubs.siam.org/doi/10.1137/1033051

[Mes21] Message Passing Interface Forum, *MPI: A message-passing interface standard version 4.0*, manual, Jun. 2021. [Online]. Available: https://www.mpi-forum.org/docs/mpi-4.0/mpi40-report.pdf

[MRC12] S. Moro, P. Rita, and P. Cortez, *Bank Marketing*, UCI Machine Learning Repository, 2012. DOI: https://doi.org/10.24432/C5K306 [Online]. Available: https://archive.ics.uci.edu/dataset/222/bank+marketing

References V

[MMOE19] W. Mostert, K. M. Malan, G. Ochoa, and A. P. Engelbrecht, “Insights into the Feature Selection Problem Using Local Optima Networks,” in *Evolutionary Computation in Combinatorial Optimization*, A. Liefooghe and L. Paquete, Eds., ser. Lecture Notes in Computer Science, Cham: Springer International Publishing, 2019, pp. 147–162, ISBN: 978-3-030-16711-0. DOI: 10.1007/978-3-030-16711-0_10

[Nak96] K. Nakai, *Yeast*, UCI Machine Learning Repository, 1996. DOI: https://doi.org/10.24432/C5KG68 [Online]. Available: https://archive.ics.uci.edu/dataset/110/yeast

[OM24] G. Ochoa and K. Malan, “Landscape Analysis of Optimisation Problems and Algorithms,” in *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, ser. GECCO ’24 Companion, New York, NY, USA: Association for Computing Machinery, Aug. 1, 2024, pp. 984–1004, ISBN: 979-8-4007-0495-6. DOI: 10.1145/3638530.3648421 Accessed: Nov. 24, 2024. [Online]. Available: https://dl.acm.org/doi/10.1145/3638530.3648421

[OTVD08] G. Ochoa, M. Tomassini, S. Vérel, and C. Darabos, “A study of NK landscapes’ basins and local optima networks,” in *Proceedings of the 10th Annual Conference on Genetic and Evolutionary Computation*, ser. GECCO ’08, New York, NY, USA: Association for Computing Machinery, Jul. 12, 2008, pp. 555–562, ISBN: 978-1-60558-130-9. DOI: 10.1145/1389095.1389204 Accessed: Nov. 28, 2024. [Online]. Available: https://dl.acm.org/doi/10.1145/1389095.1389204

References VI

[OVDT14] G. Ochoa, S. Verel, F. Daolio, and M. Tomassini, “Local Optima Networks: A New Model of Combinatorial Fitness Landscapes,” in *Recent Advances in the Theory and Application of Fitness Landscapes*, H. Richter and A. Engelbrecht, Eds., Berlin, Heidelberg: Springer, 2014, pp. 233–262, ISBN: 978-3-642-41888-4. DOI: 10.1007/978-3-642-41888-4_9 Accessed: Nov. 29, 2024. [Online]. Available: https://doi.org/10.1007/978-3-642-41888-4_9

[Qui87] Quinlan Quinlan, *Credit Approval*, UCI Machine Learning Repository, 1987. DOI: 10.24432/C5FS30 Accessed: Mar. 19, 2026. [Online]. Available: https://archive.ics.uci.edu/dataset/27

[RS02] C. M. Reidys and P. F. Stadler, “Combinatorial Landscapes,” *SIAM Review*, vol. 44, no. 1, pp. 3–54, 2002, ISSN: 0036-1445. JSTOR: 4148412. Accessed: Apr. 8, 2024. [Online]. Available: https://www.jstor.org/stable/4148412

[RN20] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach* (Pearson Series in Artificial Intelligence), 4th ed. Pearson, 2020, ISBN: 978-0-13-461099-3. [Online]. Available: https://books.google.no/books?id=koFptAEACAAJ

[SMM24] X. F. C. Sánchez-Díaz, C. Masson, and O. J. Mengshoel, “Regularized Feature Selection Landscapes: An Empirical Study of Multimodality,” in *Parallel Problem Solving from Nature – PPSN XVIII*, M. Affenzeller et al., Eds., Cham: Springer Nature Switzerland, Sep. 7, 2024, pp. 409–426, ISBN: 978-3-031-70055-2. DOI: 10.1007/978-3-031-70055-2_25

References VII

[SM23a] X. F. C. Sánchez-Díaz and O. J. Mengshoel, "EvoLP.jl: A playground for Evolutionary Computation in Julia," in *Proceedings of the 5th Symposium of the Norwegian AI Society*, ser. NAIS, vol. 3431, Bergen, Norway: CEUR Workshop Proceedings, 2023, ISBN: 1613-0073. [Online]. Available: https://ceur-ws.org/Vol-3431/paper7.pdf

[SM23b] X. F. C. Sánchez-Díaz and O. J. Mengshoel, "Evolutionary Computation with Islands: Extending EvoLP.jl for Parallel Computing," *Norsk IKT-konferanse for forskning og utdanning*, no. 1, 1 Nov. 30, 2023, ISSN: 1892-0721. Accessed: Nov. 30, 2023. [Online]. Available: https://www.ntnu.no/ojs/index.php/nikt/article/view/5667

[SM24] X. F. C. Sánchez-Díaz and O. J. Mengshoel, "Estimating the Number of Local Optima in Multimodal Pseudo-Boolean Functions: Validation via Landscapes of Triangles," in *Proceedings of the Genetic and Evolutionary Computation Conference Companion*, ser. GECCO '24 Companion, New York, NY, USA: Association for Computing Machinery, Aug. 1, 2024, pp. 211–214, ISBN: 979-8-4007-0495-6. DOI: 10.1145/3638530.3654156 Accessed: Aug. 7, 2024. [Online]. Available: https://dl.acm.org/doi/10.1145/3638530.3654156

[SM25] X. F. C. Sánchez-Díaz and O. J. Mengshoel, "Visualizing Multimodality in Combinatorial Search Landscapes," arXiv: 2510.06517 [cs], Accessed: Oct. 9, 2025. [Online]. Available: http://arxiv.org/abs/2510.06517 pre-published.

References VIII

[Sán24] X. F. C. Sánchez-Díaz, *Replication Data for: Regularized Feature Selection Landscapes: An Empirical Study of Multimodality*, DataverseNO, Jun. 24, 2024. DOI: 10.18710/DQZKMX Accessed: Mar. 13, 2025. [Online]. Available: https://dataverse.no/dataset.xhtml?persistentId=doi:10.18710/DQZKMX

[Sim13] D. Simon, *Evolutionary Optimization Algorithms*. John Wiley & Sons, Jun. 13, 2013, 776 pp., ISBN: 978-1-118-65950-2. Google Books: gwUwIEPqk30C.

[SJTR22] M. Själander, M. Jahre, G. Tufte, and N. Reissmann, "EPIC: An Energy-Efficient, High-Performance GPGPU Computing Research Infrastructure," arXiv: 1912.05848 [cs], Accessed: Aug. 14, 2023. [Online]. Available: http://arxiv.org/abs/1912.05848 pre-published.

[TO22] S. L. Thomson and G. Ochoa, "On funnel depths and acceptance criteria in stochastic local search," in *Proceedings of the Genetic and Evolutionary Computation Conference*, ser. GECCO '22, New York, NY, USA: Association for Computing Machinery, Jul. 8, 2022, pp. 287–295, ISBN: 978-1-4503-9237-2. DOI: 10.1145/3512290.3528831 Accessed: Jul. 14, 2022. [Online]. Available: https://doi.org/10.1145/3512290.3528831

[Wil05] L. Wilkinson, *The Grammar of Graphics* (Statistics and Computing), 2nd ed. New York: Springer-Verlag, 2005, 691 pp., ISBN: 978-0-387-24544-7. DOI: 10.1007/0-387-28695-0 Accessed: Nov. 13, 2024. [Online]. Available: http://link.springer.com/10.1007/0-387-28695-0

Section 6

Additional Material

Research Direction 1

Research Structure

Research Objective 1

Develop software tools to design, implement, and analyse evolutionary computation solvers.

- RQ1: How can we develop a software package that captures the basic tools we need?
  - Paper I: EvoLP.jl: A Playground for Evolutionary Computation in Julia (NAIS 2023)
- RQ2: How can the software tools be extended to be able to run on HPC Clusters?
  - Paper II: Evolutionary Computation with Islands: Extending EvoLP.jl for Parallel Computing (NIKT 2023)

Research Direction 2

Research Structure

Research Objective 2	Research Objective 3
Study the search landscape of pseudo-Boolean problems.	Visualise pseudo-Boolean search landscapes.

- RQ3: How can we estimate the number of local optima in a PBF?
- RQ4: What can we learn from creating a function with many known optima to use as a case study?
  - Paper III: Estimating the Number of Local Optima in Multimodal Pseudo-Boolean Problems: Validation via Landscapes of Triangles (GECCO 2024)
- RQ5: How do local optima spread across a pseudo-Boolean search space?
- RQ6: How to plot the entire search space of a PBF in two dimensions?
  - Paper IV: Regularized Feature Selection Landscapes: An Empirical Study of Multimodality (PPSN 2024)
  - Paper V: Visualizing Pseudo-Boolean Functions: Feature Selection and Regularization for Machine Learning (EvoCOP 2025)

Paper I: EvoLP.jl: A Playground for Evolutionary Computation

Norwegian AI Society Symposium 2023 – Bergen, Norway

- **Example 1**: a Simple Genetic Algorithm (GA) [Gol89] solving a continuous problem, the Rosenbrock test function.
  - Random initialisation: $\boldsymbol{x} \in X \sim \mathcal{N}(\mu, \sigma^2)$ with $\mu = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$ and $\sigma^2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$
  - Rank-based selection
  - Interpolation crossover $\lambda = 0.5$
  - Gaussian mutation $\sigma = 0.5$
- **Example 2**: a steady state GA [Sim13] solving a combinatorial problem, the 8-queens problem [RN20].
  - Random initialization: $\boldsymbol{x} \in X \sim \mathcal{U}(\mathcal{S}_8)$ where $\mathcal{S}_8$ is the set of permutations of $[8] = \{1, 2, \ldots, 8\}$
  - Random tournament selection, with tournament size $T_s = 5$
  - Order-one crossover
  - Swap mutation

Worked examples are available in the documentation: `https://ntnu-ai-lab.github.io/EvoLP.jl/`

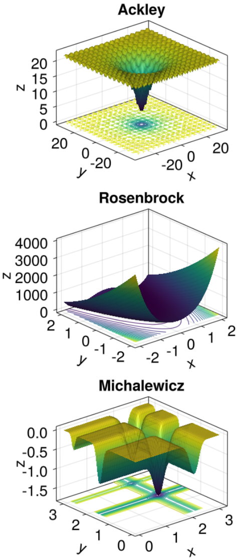
The (1+1)-EA in EvoLP.jl

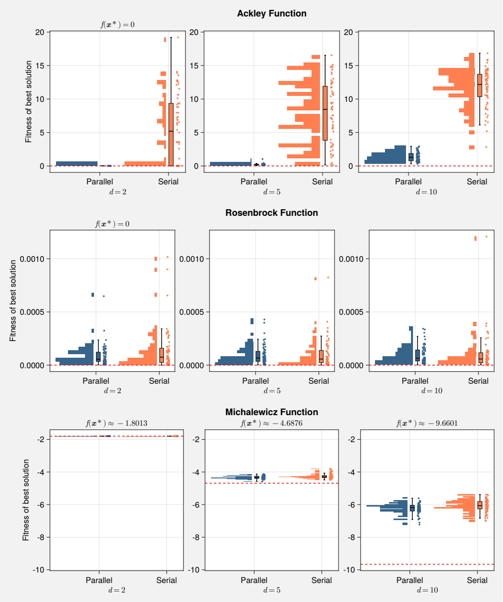
```julia
function oneplusone!(
    logger::Logbook, f::Function, ind::AbstractVector, k_max::Integer, M::Mutator
    )
    fx = Inf  # works only on minimisation problems
    runtime = @elapsed for _ in 1:k_max
        c = mutate(M, ind)
        fx, fc = f(ind), f(c)
        if fc <= fx  # O(2 * k_max)  # minimisation problem
            ind = c
            fx = fc
        end
        compute!(logger, [fx])
    end

    n_evals = 2k_max

    return Result(fx, ind, [ind], k_max, n_evals, runtime)
end
```

Paper II: Test 1





- Serial approach has higher spread
- Parallel approach has better convergence

Paper III: Estimating the Number of Local Optima in Multimodal Pseudo-Boolean Problems: Validation via Landscapes of Triangles

Genetic and Evolutionary Computation Conference 2024 – Melbourne, Australia

Definition

Let $\|\boldsymbol{b}\|$ be the number of 1s in the bitstring $\boldsymbol{b} \in \mathbb{B}^n$. Then, the *triangular positive wave* function, or $\mathrm{Triangle}$ for short, is defined as:

$$\mathrm{Triangle}(\boldsymbol{b}, m, s) = \begin{cases} g(\boldsymbol{b}), & \text{if } \left\lceil \frac{\|\boldsymbol{b}\|}{s} \right\rceil \mod 2 = 1 \\ m\left(\left\lceil \frac{\|\boldsymbol{b}\|}{s} \right\rceil \cdot s - \|\boldsymbol{b}\|\right) & \text{otherwise} \end{cases} \tag{1}$$

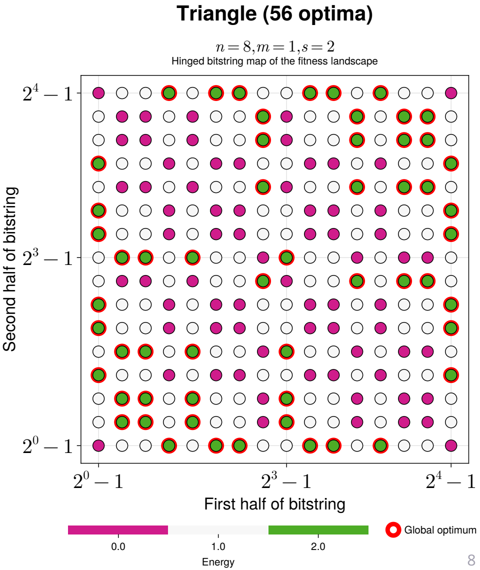
where

$$g(\boldsymbol{b}) = \begin{cases} m \cdot s, & \text{if } \|\boldsymbol{b}\| \mod s = 0 \\ m(\|\boldsymbol{b}\| \mod s) & \text{otherwise.} \end{cases} \tag{2}$$

Triangle($\boldsymbol{b}, m = 1, s = 2$), $\boldsymbol{b} \in \mathbb{B}^8$

- The search space can be partitioned into three levels according to fitness:
  - Level 1: $f(\boldsymbol{b}) = 0$ with 0 ones
  - Level 2: $f(\boldsymbol{b}) = 1$ with 1 one
  - Level 3: $f(\boldsymbol{b}) = 2$ with 2 ones

Easy to find any optima, not as easy to find <u>all</u> of them!



Fitness-Distance Correlation

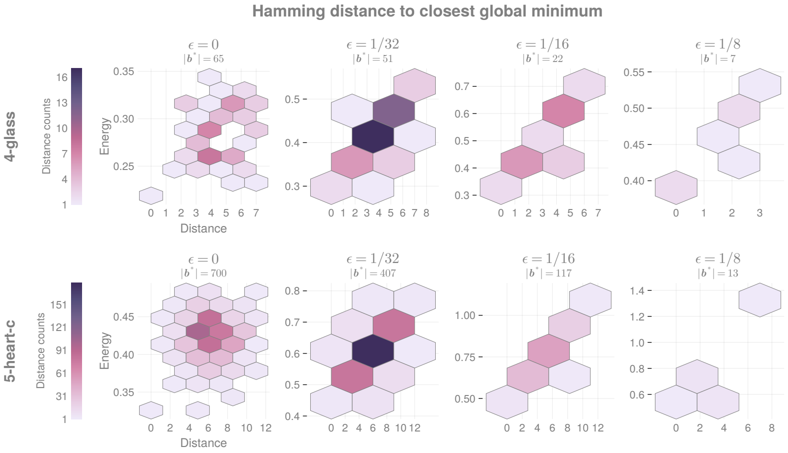
Paper IV — Experiment 2: Visualising Regularisation

Big valley hypothesis [TO22]:

- Funnel structures
- Many local optima concentrated around a global optimum

Regularisation splits the big valley into **rings**

- Informally, we say that basin size is reduced



Preserving Neighbourhood Structure

Paper V — Analysis of Visualisation

Lemma

There is no isometric embedding from $\mathbb{B}^2$ into $\mathbb{R}^2$ that preserves the Hamming distance as the Euclidean distance [Mas24].

Proof.

Let A, B, E and D be the bitstrings $00, 10, 01$ and 11, respectively. Put A anywhere in Euclidean space, and draw its Hamming shell of radius 1 with a circle. Place B (a neighbour of A) on this circle, such that $H(A, B) = 1$. Place D in space such that $H(A, D) = 2$ and $H(B, D) = 1$. However, there is no point in Euclidean space for E other than on top of B. Hence, there is no isometric embedding between the two spaces that preserves the original distance. □

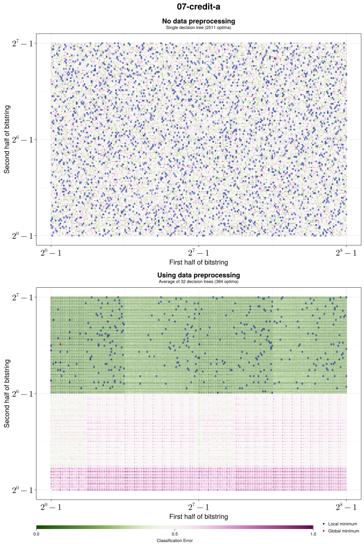
Data preprocessing and treatment

Credit approval [Qui87], with $n = 15$ features and $m = 20000$ rows

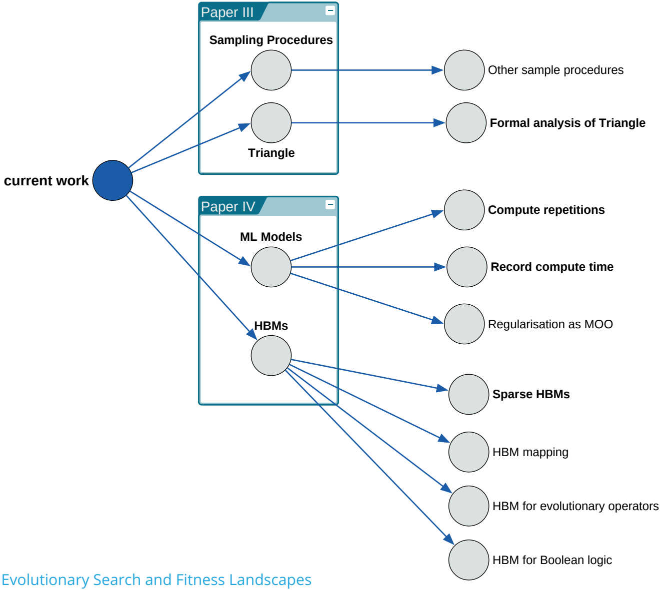
Data preprocessing reduces number of optima and generates *dense* landscapes. Informally:

- Dense landscapes are *easier* to optimise
- Learning occurs due to the presence of *some structure*

Landscapes depend on the data, the algorithm, and its hyperparameters!



Future Work



Paper III
Sampling Procedures
Triangle
current work
Paper IV
ML Models
HBMs
Other sample procedures
Formal analysis of Triangle
Compute repetitions
Record compute time
Regularisation as MOO
Sparse HBMs
HBM mapping
HBM for evolutionary operators
HBM for Boolean logic

Subsection 6

Additional Research Contributions

Additional Research Contributions I

- **GECCO 2023: Comparing Metaheuristic Optimization Algorithms for Ambulance Allocation: An Experimental Simulation Study**
  - Schjølberg, M. E., Bekkevold, N. P., Sánchez-Díaz, X. F. C., & Mengshoel, O. J. (2023). Comparing Metaheuristic Optimization Algorithms for Ambulance Allocation: An Experimental Simulation Study. Proceedings of the Genetic and Evolutionary Computation Conference, 1454-–1463. https://doi.org/10.1145/3583131.3590345
- **GECCO 2023: Controlling Hybrid Evolutionary Algorithms in Subset Selection for Multimodal Optimization**
  - Mengshoel, O. J., Foss, F., & Sánchez-Díaz, X. F. C. (2023). Controlling Hybrid Evolutionary Algorithms in Subset Selection for Multimodal Optimization. Proceedings of the Companion Conference on Genetic and Evolutionary Computation, 507–510. https://doi.org/10.1145/3583133.3590545

Additional Research Contributions II

- **NorwAI Innovate 2024: Investigating the Ethical Dimensions of AI: An Interdisciplinary Approach Combining CBR and “It Could be Otherwise”**
  - Wang, S., Ecclesia, S., Sánchez-Díaz, X. F. C., Søraa, R. A. & Øztürk, P. (2024). Investigating the Ethical Dimensions of AI: An Interdisciplinary Approach Combining CBR and “It Could be Otherwise”. NorwAI Innovate Conference 2024. (Trondheim, Norway. Sep. 2024).

- **NAIS 2025: Multimodality in Combinatorial Search Landscapes**
  - Sánchez-Díaz, X. F. C., & Mengshoel, O. J. (2025). Visualizing Multimodality in Combinatorial Search Landscapes (No. arXiv:2510.06517). arXiv. https://doi.org/10.48550/arXiv.2510.06517

Additional Research Contributions III

- **GECCO 2025: Empirical Studies of Multimodality in Feature Selection: Generalized Crowding for Genetic Algorithms**
  - Masson, C., Mengshoel, O. J., & Sánchez-Díaz, X. F. C. (2025). Empirical Studies of Multimodality in Feature Selection: Generalized Crowding for Genetic Algorithms. Proceedings of the Genetic and Evolutionary Computation Conference Companion, 235–238. https://doi.org/10.1145/3712255.3726535

- **NorwAI Innovate 2025: Understanding Multimodality in Feature Selection Landscapes**
  - Sánchez-Díaz, X. F. C., & Mengshoel, O. J. (2025). Understanding Multimodality in Feature Selection Landscapes. NorwAI Innovate Conference 2025. (Trondheim, Norway. Sep. 2025).

Additional Research Contributions IV

- **GECCO 2026: Benchmarking Evolutionary and Machine Learning Algorithms: The SIREN Challenge of Emergency Medical Services**
  - Kallekleiv, T.A, Sánchez-Díaz, X. F. C., & Mengshoel, O. J. (2026). Benchmarking Evolutionary and Machine Learning Algorithms: The SIREN Challenge of Emergency Medical Services. *In Press*. Proceedings of the Genetic and Evolutionary Computation Conference. (San José, Costa Rica, Jul. 2026)