On the Efficiency of Non-Elitist Evolutionary Algorithms in Presence of Local Optima

Anton Eremeev

Novosibirsk State University,
Sobolev Institute of Mathematics SB RAS,
Russia

March 25, 2025

The work is supported by the Mathematical Center in Akademgorodok
The speaker is grateful to D.-C. Dang and P.K. Lehre for their participation in the research at the initial stage.

Outline

Introduction

Local Optima and Evolutionary Algorithms

Definition of $\textsc{SparseLocalOpt}_{\alpha,\varepsilon}$ problem class

The main result

Application to Vertex Cover Problem

Conclusion

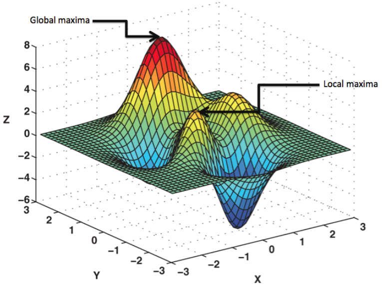
Evolutionary Algorithms in Artificial Intelligence

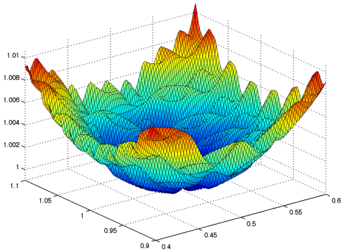
Some well-known applications:

- Feature and/or sample selection
- Neural network architecture design
- Genetic programming for symbolic regression
- Clustering
- Evolution of algorithms for combinatorial optimization (FunSearch[1])
- etc.

[1] B. Romera-Paredes, M. Barekatain, A. Novikov, M. Balog, M. P. Kumar, et al.. Mathematical discoveries from program search with large language models. Nature, 2023

Local and Global Optima (Continuous Search Space)[2]





[2] From http://www.turingfinance.com/fitness-landscape-analysis-for-computational-finance/

Optimization in Space of Binary Strings

Some well-known applications:

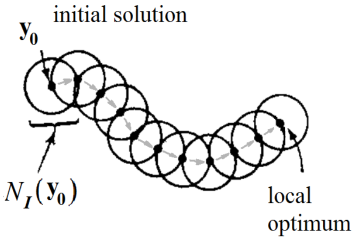
- Feature and/or sample selection
- Neural network architecture design
- Genetic programming for symbolic regression
- Clustering
- Evolution of algorithms for combinatorial optimization (FunSearch[3])
- etc.

[3] B. Romera-Paredes, M. Barekatain, A. Novikov, M. Balog, M. P. Kumar, et al.. Mathematical discoveries from program search with large language models. Nature, 2023

Local Search (Discrete Search Space)

Let $\mathrm{Sol} \subseteq \{0,1\}^n$ be the set of feasible solutions.
$f : \mathrm{Sol} \to \mathbb{R}$ is the objective function (fitness function) to be maximized.

For each $\mathbf{y} \in \mathrm{Sol}$ a neighborhood $\mathcal{N}(\mathbf{y}) \subseteq \mathrm{Sol}$ is defined.
$\mathcal{N} : \mathrm{Sol} \to 2^{\mathrm{Sol}}$ is called a *neighborhood mapping*.



Non-Elitist Genetic Algorithm[4]

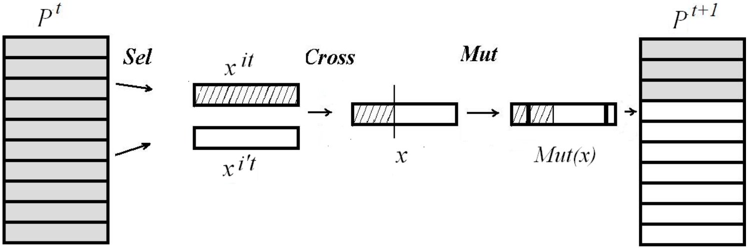
$P^t = (\mathbf{x}^{1t}, \dots, \mathbf{x}^{\lambda t}) \in \mathrm{Sol}^{\lambda}$ denotes the population at iteration t.

Generate the initial population P^0, assign $t := 0$.

While termination condition is not met **do:**

- **For** j **from 1 to** λ **do:**
  - Selection: $\mathbf{p}^1 := \mathbf{x}^{\mathrm{Sel}(P^t),t}$, $\mathbf{p}^2 := \mathbf{x}^{\mathrm{Sel}(P^t),t}$.
  - Crossover: $\mathbf{x} :=$ Cross$(\mathbf{p}^1, \mathbf{p}^2)$.
  - Mutation: $\mathbf{x}^{j,t+1} :=$ Mut$(\mathbf{x}')$.
- **End for.**
- $t := t + 1$.

End while.



[4] Goldberg D.E. Genetic Algorithms in Search, Optimization and Machine Learning. 1989.

Non-Elitist Evolutionary Algorithm (EA) - no crossover

1. Set $t := 0$.
2. Build the initial population P_0 at random.

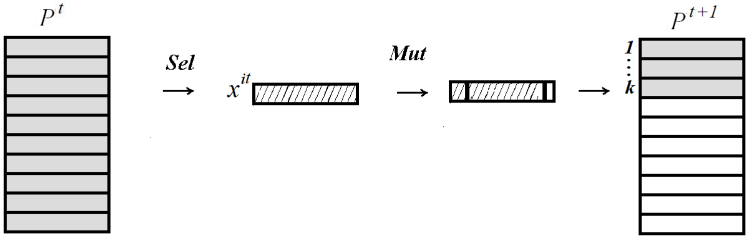
Iteration t:

3. For k from 1 to λ do Steps 3.1–3.2:

3.1. *Selection:* $i := \mathrm{Sel}(P_t)$.

3.2. *Mutation:* $x^{k,t+1} := \mathrm{Mut}(x^{i,t})$.

4. If the terminarion condition is not met, $t := t + 1$ go to Step 3.
5. The resulting genotype $\tilde{x}^t$ is the fittest individual in $P_0, \ldots, P_t$.



EA Operators

Operators $\mathrm{Sel}(P^t)$. Cross$(\mathbf{p}^1, \mathbf{p}^2)$ and Mut$(\mathbf{x})$ are efficiently computable randomized routines.

Examples:

- (μ, λ) **selection:** a parent is selected uniformly at random among μ individuals of highest fitness in P^t. (Survival of the fittest.)
- **Tournament selection:** s individuals are chosen from P_t uniformly at random, and a fittest among them gets selected.
- **Bitwise mutation** consists in changing each bit of a given solution with a fixed mutation probability p_m.
  Usually it is assumed that $p_m = \chi/n$ for some constant parameter $\chi > 0$.

Simple Elitist Algorithm (1+1) EA

Algorithm (1+1)-EA

Generate initial individual $x^{(0)}$, assign $t := 1$.
While termination condition is not met **do:**
    Produce $x' := \mathsf{Mut}(x^{(t-1)})$.
    If $f(x^{(t-1)}) \leq f(x')$ then $x^{(t)} := x'$,
        else $x^{(t)} := x^{(t-1)}$.
    $t := t + 1$.
End while.

$(\mu+\lambda)$ Elitist Black Box Complexity (Doerr, Lengler, 2016)

Optimisation with limited information about the objective function.

$(\mu+\lambda)$-elitist black box algorithm A (Survival of the fittest.)

- For every round t, depending only on the fittest μ observed search points and their fitness values $(x_{(1)}, f(x_{(1)})), \ldots, (x_{(\mu)}, f(x_{(\mu)}))$, alg. A selects λ new search points $x_{\lambda(t+1)+1}, \ldots, x_{\lambda(t+2)}$, and obtains $f(x_{\lambda(t+1)+1}), \ldots, f(x_{\lambda(t+2)})$ from oracle.

The $(\mu+\lambda)$-**elitist black box complexity** of class F is

- $T_F := \min_{A \in \mathcal{A}_{\text{elitist}}} T_{A,F}$,

where $\mathcal{A}_{\text{elitist}}$ is the set of $(\mu+\lambda)$-elitist black box algorithms.

The Set Cover Problem (SCP):

For $S_1, \ldots, S_n \subset U$, find $I \subseteq [n]$ st. $\cup_{i \in I} S_i = U$ and $|I|$ is **minimized**.

Representation of solutions

$x_i = 1 \Leftrightarrow i$ is included into $I(x)$, and $\mathrm{Fitness}(x) = |I(x)| + |U \setminus \cup_{i \in I(x)} S_i|$.

The Set Cover Problem instances

We use the instances of moderate size and sufficient difficulty from the families CYC, CLR, and STEIN, available in the OR-library.[5]

Compared algorithms

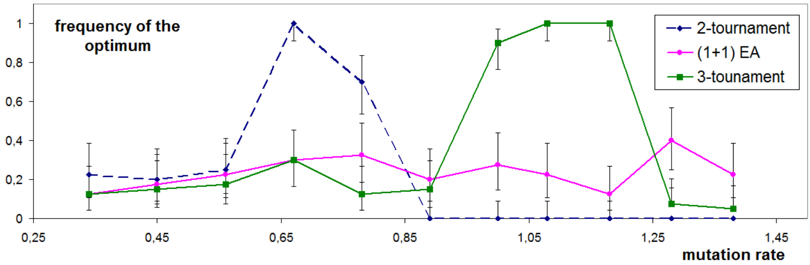
- **Elitist:** $(1+1)$ EA, $(100+200)$ EA, $(66+200)$ EA,
- **Non-elitist:** $(100, 200)$ EA, $(66, 200)$ EA, EAs with 2- and 3-tournament.

EA parameters were chosen as recommended by theory (Lehre, 2010; Dang et al, 2021).

Each algorithm was given the same budget of 2×10^8 fitness evaluations, and each setting on each instance was tested with 40 replications of the run.

[5] SCP files from `http://people.brunel.ac.uk/~mastjjb/jeb/orlib/files/`

Experiments on Instance CYC.6 (success frequency)

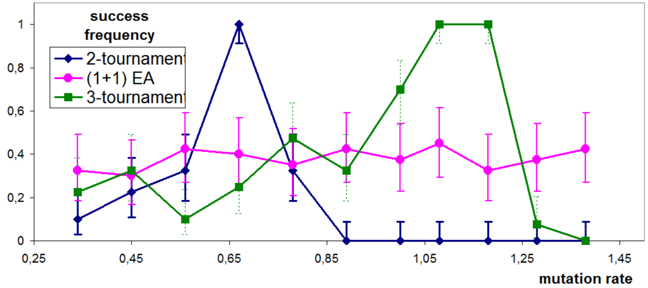


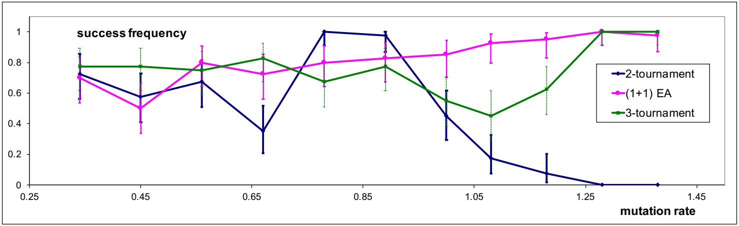
In this figure, we show the frequency of obtaining the optimum and the 95% confidence intervals of the corresponding probability.[6]

The figure also displays the results of the $(1{+}1)$ EA, which was given a similar runtime of $2 \cdot 10^8$ function evaluations.

[6] D.-C. Dang, A.V. Eremeev, P.K. Lehre. Escaping Local Optima with Non-Elitist Evolutionary Algorithms. Proc. of AAAI 2021, pp. 12275-12283.

Success Frequency on Instances CLR.10 and Stein.81





Preliminaries

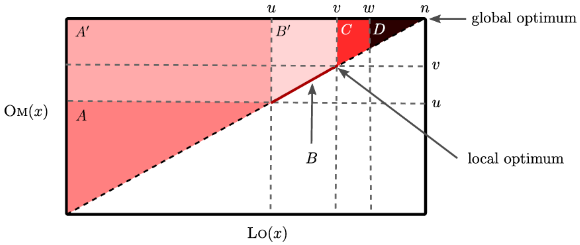
- For any $n \in \mathbb{N}$, define $[n] := \{1, \ldots, n\}$.
- **Standard pseudo-Boolean functions:**

$$\textsc{OneMax}(x) := \mathrm{OM}(x) := \sum_{i=1}^{n} x_i$$

and

$$\textsc{LeadingOnes}(x) := \mathrm{LO}(x) := \sum_{i=1}^{n} \prod_{j}^{i} [x_j = 1].$$

Family of Fitness Functions FUNNEL[7]



$$\text{FUNNEL}(x) := \begin{cases} \text{LO}(x) & \text{if } w < \text{LO}(x) \le n & (D) \\ \text{LO}(x) + u - v & \text{if } v < \text{LO}(x) \le w & (C) \\ \text{LO}(x) + w - v & \text{if } u < \text{LO}(x) \le v \text{ and } \text{LO}(x) = \text{OM}(x) & (B) \\ -n & \text{if } u < \text{LO}(x) \le v \text{ and } \text{LO}(x) < \text{OM}(x) & (B') \\ \text{LO}(x) & \text{if } \text{OM}(x) \le u & (A) \\ -\text{OM}(x) & \text{otherwise} & (A') \end{cases}$$

On the figure, the darker the shade the higher region fitness.

- elitist EAs and $(\mu + \lambda)$ EA are slow (exponential on average), but
- the EA with tournament selection has polynomial optimization time.

---

[7] **Dang D.-C., Eremeev A.V., and Lehre P.K.** Escaping local optima with non-elitist evolutionary algorithms. Proc. of AAAI Conference on Artificial Intelligence (AAAI'2021). 2021.

Function class $\textsc{SparseLocalOpt}_{\alpha,\varepsilon}$ (informal)[8]

$f \in \textsc{SparseLocalOpt}_{\alpha,\varepsilon} \quad \iff$

- "fitness valleys" of f are α-dense
- "deceptive regions" of f are ε-sparse

[8] Dang D.-C., Eremeev A.V., and Lehre P.K. Non-elitist evolutionary algorithms excel in fitness landscapes with sparse deceptive regions and dense valleys. Proc. of the Genetic and Evolutionary Computation Conference (GECCO'2021).

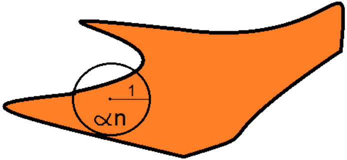
$\textsc{SparseLocalOpt}_{\alpha,\varepsilon}$: Dense and sparse sets

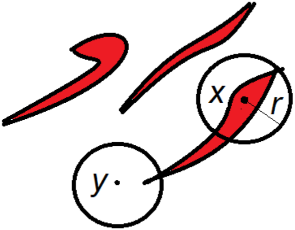
Definition (Hamming sprehe)

$S_r(x) := \{y \in \{0,1\}^n \mid H(x,y) = r\}$. Note that $|S_r(x)| = \binom{n}{r}$.

Definition (α-dense set A w.r.t. a superset D)

For all $x \in A$, $|S_1(x) \cap D| \geq \alpha n$





Definition (ε-sparse set B)

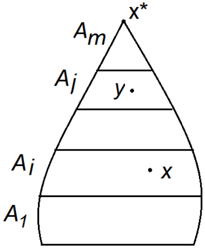
For $\varepsilon \in [0,1]$, a subset $B \subseteq \{0,1\}^n$ is ε-sparse if:

(SP1) $\forall x \in B, \forall r \leq n-1,$
$|S_r(x) \cap B| \leq \varepsilon \cdot \binom{n}{r}$, and

(SP2) $\forall x \in \{0,1\}^n \setminus B, \forall r \leq n-1,$
$|S_r(x) \cap B| = \mathcal{O}\left(\frac{1}{n}\binom{n}{r}\right)$.

Deceptive pairs (deceptive regions and fitness valleys)

Definition



Given a partition $(A_1, \ldots, A_m)$ of $\{0,1\}^n$, a pair (A_i, A_j) is called f-deceptive if

- $i < j$
- $\exists x \in A_i, \exists y \in A_j$ st $f(x) \geq f(y)$

$\textsc{SparseLocalOpt}_{\alpha,\varepsilon}$

Definition

A function $f : \{0,1\}^n \to \mathbb{R}$ belongs to $\textsc{SparseLocalOpt}_{\alpha,\varepsilon}$ if there exists a partition of $\{0,1\}^n$ into $m \in \mathrm{poly}(n)$ subsets $(A_1, \ldots, A_m)$ such that

I. $x \in A_m \iff f(x) = \max_y f(x)$

II. $\forall x \in A_j, \exists y \in A_{j+k},\ k \geq 1$ such that $H(x,y) \leq d,$.

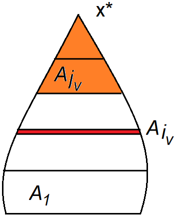
If $(A_{i_1}, A_{j_i}), \ldots, (A_{i_u}, A_{j_u})$ are f-deceptive pairs then

III.

$$\cup_{v=1}^{u} A_{i_v} \text{ is } \varepsilon\text{-sparse, and}$$

IV.

$$\cup_{k=j_v}^{m} A_k \text{ is } \alpha\text{-dense for all } v \in [u].$$



$\alpha' \leq \alpha$ and $\varepsilon \leq \varepsilon' \iff \textsc{SparseLocalOpt}_{\alpha,\varepsilon} \subseteq \textsc{SparseLocalOpt}_{\alpha',\varepsilon'}$.

Example: $\textsc{OneMax} \in \textsc{SparseLocalOpt}_{1,0}$

Summary of Known Theoretical Results (informal)

- For any constant $\delta \in (0,1)$, the **elitist** evolutionary algorithms EAs on $\textsc{SparseLocalOpt}_{1-\delta,\delta}$ in the worst case require expected **exponential time** to reach the global optimum.[9]
- Non-elitist evolutionary algorithms (EAs) with bit-wise mutation have expected **polynomial runtime** on $\textsc{SparseLocalOpt}$, given appropriate values of α, ε and selection and mutation parameters. In particular, this applies to EAs with 2- and 3-tournament selection.[10]
- The function class $\textsc{SparseLocalOpt}$ is hard for all black-box optimisation algorithms when the "sparsity" ε of deceptive regions is larger than the "density" α of fitness valleys.[11]

[9] D.-C. Dang, A.V. Eremeev, P.K. Lehre. Non-elitist evolutionary algorithms excel in fitness landscapes with sparse deceptive regions and dense valleys. Proc. of GECCO 2021.

[10] the same reference

[11] D.-C. Dang, P.K. Lehre. The SLO Hierarchy of pseudo-Boolean Functions and Runtime of Evolutionary Algorithms. Proc. of GECCO 2024.

Contributions of this Talk[12]

Main:
- A new (tighter) polynomial upper bound for the mathematical expectation of the optimization time for non-elitist evolutionary algorithms on problems from $\textsc{SparseLocalOpt}_{\alpha,\varepsilon}$.
- The efficiency of non-elitist evolutionary algorithms is shown on a wider class of problems $\textsc{SparseLocalOptB}_{\alpha,\varepsilon}$.

- **Illustration:** Application of the upper bound to a family of vertex cover problems on star graphs. Advantage of non-elitist evolutionary algorithms, compared to the (1+1)-EA, is demonstrated.

Here we define the class $\textsc{SparseLocalOptB}_{\alpha,\varepsilon}$ similarly to $\textsc{SparseLocalOpt}_{\alpha,\varepsilon}$ with the difference that instead of item IV from definition of $\textsc{SparseLocalOpt}_{\alpha,\varepsilon}$, for the class $\textsc{SparseLocalOptB}_{\alpha,\varepsilon}$ we assume:

IV'. A_{j_v} is an α-dense set *with respect to* $\cup_{k=j_v}^{m} A_k$ for each $v \in [u]$.

[12] A. Eremeev. On the efficiency of non-elitist evolutionary algorithms in the case of sparsity of the level sets inconsistent with respect to the objective function. To appear in Proceedings of the Steklov Institute of Mathematics.

The Main Result

For any $0 \leq \psi \leq \gamma \leq 1$ we denote by $\beta(\psi, \gamma)$ the probability of selecting an individual with rank from $\lceil \psi\lambda \rceil$ to $\lceil \gamma\lambda \rceil$ in the population P.

Theorem

Suppose that the non-elitist EA employs the standard mutation operator with mutation probability $p_{\mathrm{m}} = \chi/n$ and an f-monotone selection operator described by a function $\beta(\psi, \gamma)$, satisfying the following inequalities:

(SM0) $\beta(0, \gamma) \leq \frac{\gamma}{\frac{\varepsilon}{1-\varepsilon} + \left(1 - \frac{\chi}{n}\right)^n}$ *for all* $\gamma \in [\psi_0, 1]$,

(SM2a) $\beta(0, \gamma) \geq \frac{\gamma(1+\delta)}{\left(1 - \frac{\chi}{n}\right)^n}$ *for all* $\gamma \in (0, \gamma_0]$,

(SM2b') $\beta(\psi_0 + \gamma_0, \psi_0 + \gamma_0 + \gamma) \geq \frac{\gamma(1+\delta)}{\left(1 - \frac{\chi}{n}\right)^n (1+\alpha\chi)}$ *for all* $\gamma \in (0, \gamma_0]$

for some constants $\psi_0, \gamma_0, \alpha \in (0, 1)$, $\psi_0 \geq \gamma_0$. *In addition to that, let the functions* $\lambda = \lambda(n)$, $1/\varepsilon = 1/\varepsilon(n), 1/\delta = 1/\delta(n)$ *and* $1/\chi = 1/\chi(n)$ *be polynomially bounded and the following inequality holds for a sufficiently large constant* k

(SM3) $k \ln(n)/\varepsilon^2 \leq \lambda$.

Then the expected optimization time of this EA on the class $\textsc{SparseLocalOptB}_{\alpha,\varepsilon}$ *is polynomially bounded and equals* $O(m(\lambda \ln(\lambda) + (n/\chi)^d)/\delta + 1/\varepsilon)$, *where* d *is the constant from condition II.*

Corollaries form the Main Result

Corollary

The non-elitist EA with tournament selection for tournament size $k = 3$, polynomially bounded population size $\lambda \geq c\ln(n)$, where c is a sufficiently large constant, and standard mutation, where $p_m = 1/\chi,\ \chi = \ln(3) - 10^{-7}$, has polynomially bounded expected optimization time on problems from the class $\textsc{SparseLocalOptB}_{\alpha,\varepsilon}$, where $\alpha = 1/4$, $\varepsilon = 3/10^5$.

Corollary

A similar result holds for $k = 2$.

Minimum Weight Vertex Cover Problem (VCP)

Let $G = (V, E)$ be a graph where the vertices are numbered, i.e. $V = [n]$, and weighted by a function $w \colon V \to \mathbb{R}^+$. Then any bitstring $x \in \{0,1\}^n$ encodes a subset $V(x) \subseteq V$, i.e. $x_i = 1$ if vertex i belongs to $V(x)$ and $x_i = 0$ otherwise.

A subset $C \subseteq V$ is called a vertex cover if each edge from E is incident to at least one vertex from C (such edges are called covered).

VCP: find a vertex cover of minimum weight $\sum_{i=1}^{n} w_i x_i$.

Solving VCP is equivalent to the minimisation of:

$$f_{vcp}(x) := |E'(x)| \cdot S + \sum_{i=1}^{n} w(v_i) x_i,$$

where $E'(x) := \{e = u_i v_j \in E \ : \ x_i = x_j = 0\}$ is the set of uncovered edges and $S := \sum_{i=1}^{n} w(v_i)$.

Minimum Weight Vertex Cover Problem

We consider a family of instances of VCP on $K_{1,n-1}$, i.e. assume that G is a bibartite graph with $\ell = n-1$ vertices $v_1, \ldots, v_\ell$ in ''left-hand" part V_L, and one vertex v_n in the "right-hand" part V_R.

Given a parameter $\alpha \in (0,1)$, we consider the VCP instances where $w(v_n) \leq (1-\alpha)\ell$ and $w(v_1) = \cdots = w(v_\ell) = 1$.

The global optimum is the singleton set $\{v_n\}$ with weight $(1-\alpha)\ell$.

The local optimum is V_L with weight ℓ.

Randomized Local Search is Inefficient

Randomized Local Search operates as follows.

We use x^t to denote the current solution at iteration t, and initially x^0 is picked uniformly from $\{0, 1\}^n$. An iteration t consists in building an offspring y by flipping exactly one randomly chosen bit in x^t.
If $f(y)$ is at least as food as $f(x^t)$ then x^t is replaced by y,
otherwise $x^{t+1} = x^t$.

This is a simple elitist algorithm.

Remark: the expected optimisation time of the Randomized Local Search (RLS) algorithm (using the neighbourhood of Hamming distance 1) is infinite on VCP problems presented above.

Non-Elitist EAs are Efficient

Proposition. Let an instance of the VCP be given by a bipartite graph $K_{1,n-1}$, where one of the parts consists of a single vertex v_n and the other part is formed by the remaining vertices, which are assigned unit weights; in addition, $w(v_{n+1}) \leq (1-\alpha)(n-1)$ for some $0 < \alpha < 1$.
Then the fitness function f_{vcp} belongs to $\text{SparseLocalOptB}_{\alpha,1/n}$.

Corollary

If

- *population size* $c\ln(n) \leq \lambda \in \text{poly}(n)$ *for a sufficiently large* c*,*
- *mutation parameter* $\chi = 1.09812$*,*
- $0 < \alpha \leq 1/4$ *and* n *is sufficiently large*

then Non-elitist EA with 3-tournament selection has polynomially bounded expected runtime on the fitness function f_{vcp} *for the above mentioned family of VCP problems.*

Conclusion

- Characterisation of fitness landscapes is further developed in terms of
  - ε-sparsity of deceptive regions
  - α-density of fitness valleys
- Elitist evolutionary algorithms (survival of the fittest) fail on problems with "deceptive" local optima.
- Non-elitist evolutionary algorithms can overcome the local optima, surrounded by sparse "deceptive regions" and dense "valleys".
- Theory is supported by experiments with set cover problems.

Open Problems

- Bridging the gap between the theory and practice
- Extension of theoretical results from the search space $\{0,1\}^n$ to $\mathbb{Z}^n$, the set of permutations etc.
- Include the crossover operator (genetic algorithms on local optima)
- Fast evaluation of parameters (features) of a given instance, such as
  - the number of local optima
  - the distance between local optima
  - sparsity ε
  - density α
  - etc ...

  in order to forecast performance and tune the EA parameters.