Elementary Landscape Decomposition of the Test Suite Minimization Problem

Francisco Chicano, Javier Ferrer and Enrique Alba

Test Suite Minimization

- **Given:**
  - **A set of test cases $T = \{t_1, t_2, ..., t_n\}$**
  - **A set of program elements to be covered (e.g., branches) $M = \{m_1, m_2, ..., m_k\}$**
  - **A coverage matrix**

T=

	t_1	t_2	t_3	...	t_n
m_1	1	0	1	…	1
m_2	0	0	1	…	0
…	…	…	…	…	…
m_k	1	1	0	…	0

$$T_{ij} = \begin{cases} 1 & \text{if element } m_i \text{ is covered by test } t_j \\ 0 & \text{otherwise} \end{cases}$$

- **Find a subset of tests $X \subseteq T$ maximizing coverage and minimizing the testing cost**
- **Binary representation:**

$$x_i = \begin{cases} 1 & \text{if test } t_i \text{ is selected} \\ 0 & \text{otherwise} \end{cases}$$

$$coverage(x) = \sum_{i=1}^{k} \max_{j=1}^{n}\{T_{ij}x_j\}; \quad ones(x) = \sum_{j=1}^{n} x_j$$

$$f(x) = \sum_{i=1}^{k} \max_{j=1}^{n}\{T_{ij}x_j\} - c \cdot ones(x)$$

Landscape Definition

- **A landscape is a triple *(X,N, f)* where**
  - ***X* is the solution space**
  - ***N* is the neighbourhood operator**
  - ***f* is the objective function**

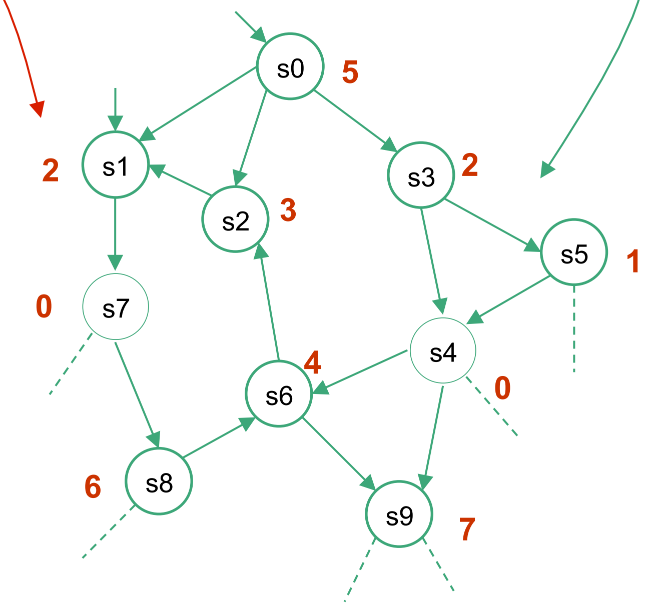
The pair *(X,N)* is called configuration space

- **The neighbourhood operator is a function**

  $N: X \rightarrow P(X)$

- **Solution *y* is neighbour of *x* if $y \in N(x)$**
- **Regular and symmetric neighbourhoods**
  - $d=|N(x)| \quad \forall\, x \in X$
  - $y \in N(x) \Leftrightarrow x \in N(y)$
- **Objective function**

  $f: X \rightarrow R$ (or N, Z, Q)



Elementary Landscapes: Formal Definition

- **An elementary function is an eigenvector of the graph Laplacian (plus constant)**

Adjacency matrix

$$A_{xy} = \begin{cases} 1 & \text{if } y \in N(x) \\ 0 & \text{otherwise} \end{cases}$$

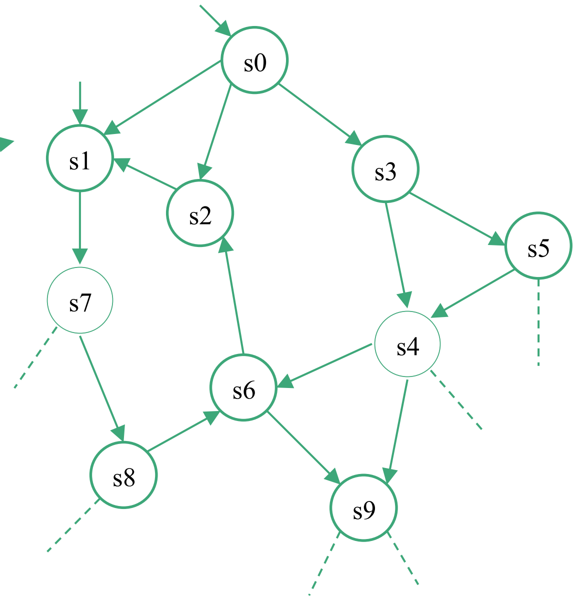
Degree matrix

$$D_{xy} = \begin{cases} |N(x)| & \text{if } x = y \\ 0 & \text{otherwise} \end{cases}$$

- **Graph Laplacian:**

$$\Delta = A - D$$

Depends on the configuration space



- **Elementary function: eigenvector of Δ (plus constant)**

$$(-\Delta) \times \left(\vec{f} - b\right) = \lambda \cdot \left(\vec{f} - b\right)$$

Eigenvalue

Elementary Landscapes: Characterizations

- **An elementary landscape is a landscape for which**

$$\underset{y \in N(x)}{\mathrm{avg}}\{f(y)\} = \alpha f(x) + \beta \quad \forall x \in X$$

Depend on the problem/instance

Linear relationship

where

$$\underset{y \in N(x)}{\mathrm{avg}}\{f(y)\} \stackrel{\text{def}}{=} \frac{1}{d} \sum_{y \in N(x)} f(y)$$

$$\bar{f} = \frac{1}{|X|} \sum_{y \in X} f(y)$$

- **Grover's wave equation**

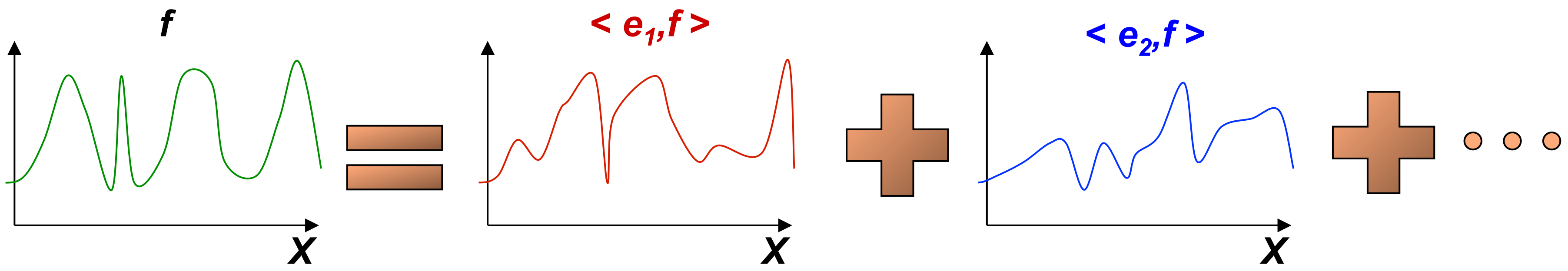
$$\underset{y \in N(x)}{\mathrm{avg}}\{f(y)\} = f(x) + \frac{\lambda}{d}\left(\bar{f} - f(x)\right)$$

$$\alpha = 1 - \frac{\lambda}{d} \qquad \beta = \frac{\lambda}{d}\bar{f}$$

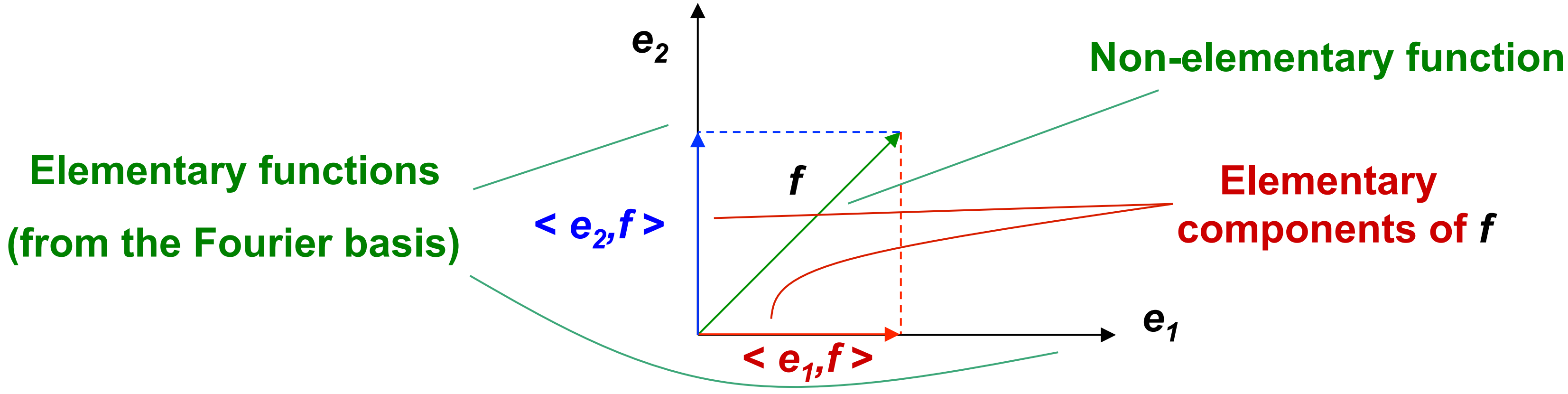
Eigenvalue

Landscape Decomposition

- What if the landscape is **not elementary?**
- Any landscape can be written as the **sum of elementary landscapes**



- There exists a set of **eigenfunctions of Δ** that form a basis of the function space (**Fourier basis**)



Examples

Elementary Landscapes

Problem	Neighbourhood	*d*	*k*
Symmetric TSP	2-opt	$n(n-3)/2$	$n-1$
	swap two cities	$n(n-1)/2$	$2(n-1)$
Graph α-Coloring	recolor 1 vertex	$(\alpha-1)n$	2α
Max Cut	one-change	n	4
Weight Partition	one-change	n	4

Sum of elementary Landscapes

Problem	Neighbourhood	*d*	Components
General TSP	inversions	$n(n-1)/2$	2
	swap two cities	$n(n-1)/2$	2
Subset Sum Problem	one-change	n	2
MAX k-SAT	one-change	n	k
QAP	swap two elements	$n(n-1)/2$	3
Test suite minimization	one-change	n	max $\lvert v_i \rvert$

Binary Search Space

- **The set of solutions *X* is the set of binary strings with length *n***

0	1	0	0	1	0	1	1	1	0

- **Neighborhood used in the proof of our main result: one-change neighborhood**
  - **Two solutions *x* and *y* are neighbors iff *Hamming(x,y)=1***

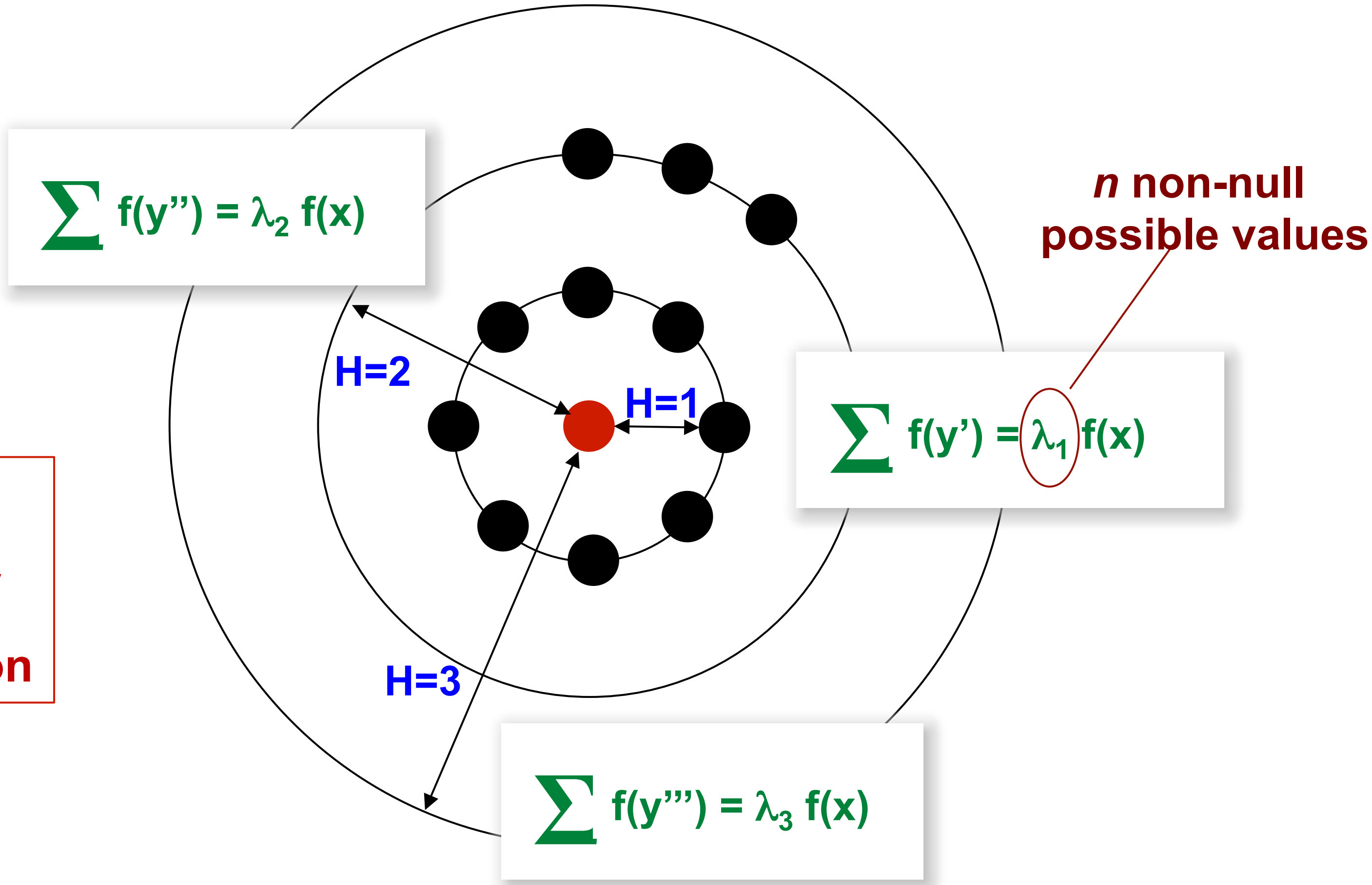
0	1	0	0	1	0	1	1	1	0

1	1	0	0	1	0	1	1	1	0
0	**0**	0	0	1	0	1	1	1	0
0	1	**1**	0	1	0	1	1	1	0
0	1	0	**1**	1	0	1	1	1	0
0	1	0	0	**0**	0	1	1	1	0

0	1	0	0	1	**1**	1	1	1	0
0	1	0	0	1	0	**0**	1	1	0
0	1	0	0	1	0	1	**0**	1	0
0	1	0	0	1	0	1	1	**0**	0
0	1	0	0	1	0	1	1	1	**1**

Spheres around a Solution

- If *f* is elementary, the average of *f* in any sphere and ball of any size around *x* is a linear expression of f(x)!!!



Sutton

Whitley

Langdon

Bit-flip Mutation: Elementary Landscapes

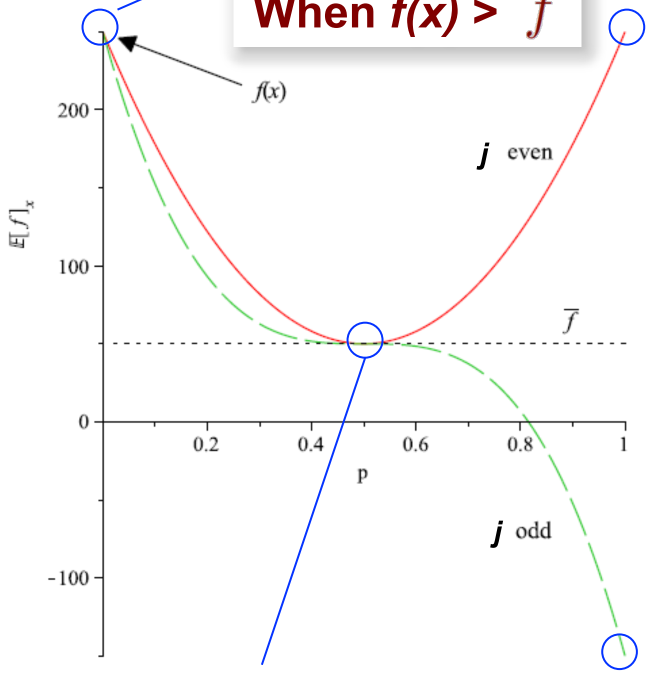
- **Analysis of the expected fitness**

$$\mathbb{E}[f]_x = \bar{f} + (1-2p)^j (f(x) - \bar{f})$$

p=0 → fitness does not change

p=1 → the same fitness

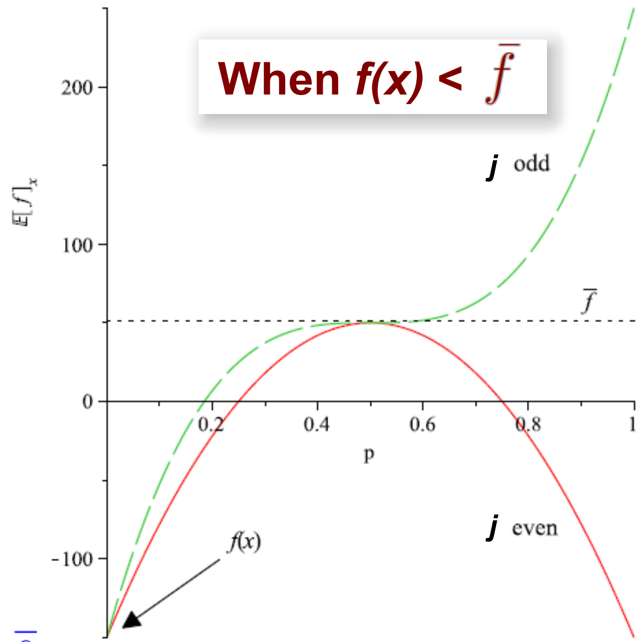
When $f(x) > \bar{f}$



p=1/2 → start from scratch

Sutton
Whitley
Howe
Chicano
Alba

When $f(x) < \bar{f}$



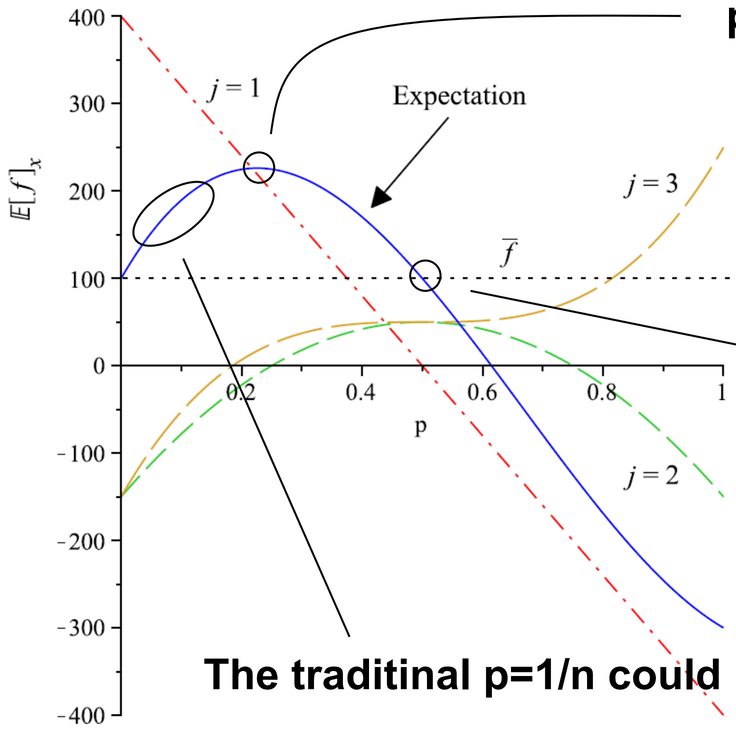
p=1 → flip around $\bar{f}$

Bit-flip Mutation: General Case

- **Analysis of the expected fitness**
- **Example (j=1,2,3):**

$$\mathbb{E}[f]_x = \overline{f} + \sum_{j=1}^{n} (1-2p)^j \left(\Omega_{2j}(x) - \overline{\Omega_{2j}}\right)$$

$$f = \Omega_2 + \Omega_4 + \Omega_6$$



p≈0.23 → maximum expectation

p=1/2 → start from scratch

The traditinal p=1/n could be around here

Elementary Landscape Decomposition of *f*

- **The elementary landscape decomposition of**

$$f(x) = \sum_{i=1}^{k} \max_{j=1}^{n}\{T_{ij}x_j\} - c \cdot ones(x)$$

Computable in *O(nk)*

is

Tests that cover m_i

$$f^{(0)}(x) = \sum_{i=1}^{k}\left(1 - \frac{1}{2^{|V_i|}}\right) - c \cdot \frac{n}{2}$$

← **constant expression**

$$f^{(1)}(x) = -\sum_{i=1}^{k} \frac{1}{2^{|V_i|}} (-1)^{n_1^{(i)}} \mathcal{K}^{|V_i|}_{|V_i|-1, n_1^{(i)}} - c \cdot \left(ones(x) - \frac{n}{2}\right)$$

Krawtchouk matrix

$$f^{(p)}(x) = -\sum_{i=1}^{k} \frac{1}{2^{|V_i|}} (-1)^{n_1^{(i)}} \mathcal{K}^{|V_i|}_{|V_i|-p, n_1^{(i)}} \quad \text{where } 1 < p \leq n$$

Tests in the solution that cover m_i

Elementary Landscape Decomposition of f^2

- **The elementary landscape decomposition of f^2 is**

Computable in $O(nk^2)$

$$\left(f^2\right)^{(0)}(x) = \beta^2 + \frac{c^2}{4}n - \sum_{i=1}^{k} \frac{c|V_i| + 2\beta}{2^{|V_i|}} + \sum_{i,i'=1}^{k} \frac{1}{2^{|V_i \cup V_{i'}|}}$$

$$\beta = k - cn/2$$

Number of tests that cover m_i or $m_{i'}$

$$\left(f^2\right)^{(p)}(x) = -\sum_{i=1}^{k}\left(\frac{(c|V_i| + 2\beta)(-1)^{n_1^{(i)}}}{2^{|V_i|}}\mathcal{K}^{|V_i|}_{|V_i|-p,n_1^{(i)}}\right) \qquad p > 2$$

$$+ \sum_{i,i'=1}^{k}\left(\frac{(-1)^{n_1^{(i\vee i')}}}{2^{|V_i \cup V_{i'}|}}\mathcal{K}^{|V_i \cup V_{i'}|}_{|V_i \cup V_{i'}|-p,n_1^{(i\vee i')}}\right)$$

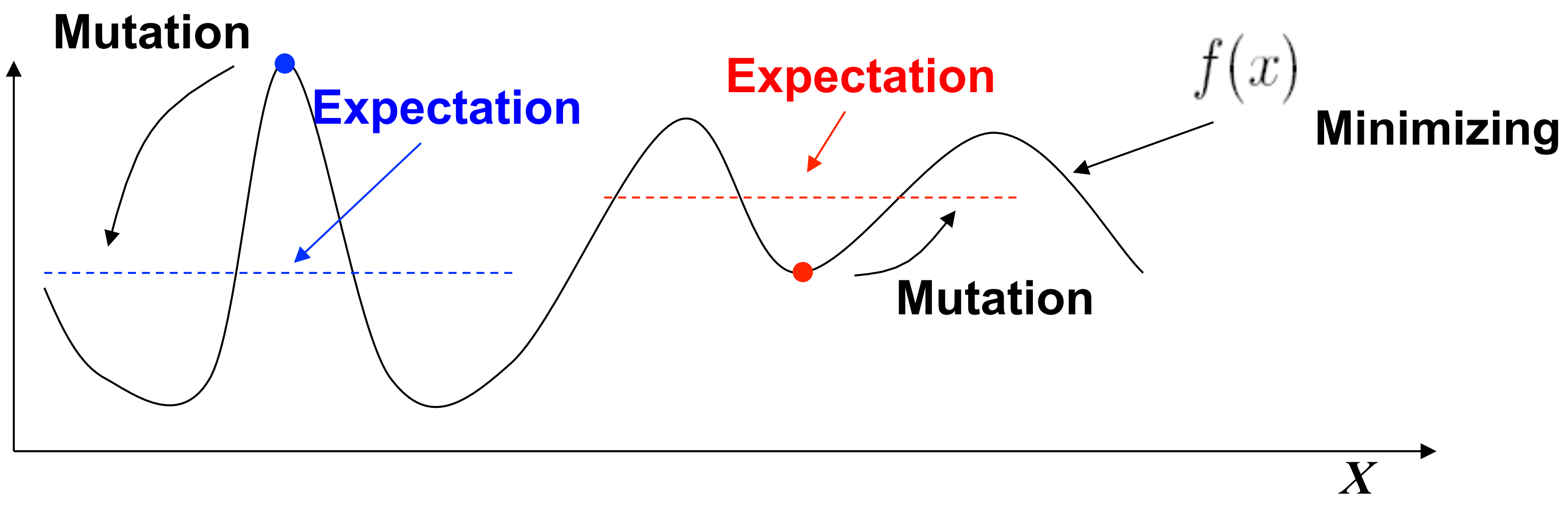
$$- c\sum_{i=1}^{k}\frac{(-1)^{n_1^{(i)}}}{2^{|V_i|}}\mathcal{K}^{|V_i|}_{|V_i|-p+1,n_1^{(i)}}\left(n - 2ones(x) - |V_i| + 2n_1^{(i)}\right)$$

$$\ldots{}_{2,n_1^{(i)}}\Big)$$

Number of tests in the solution that cover m_i or $m_{i'}$

Selection Operator

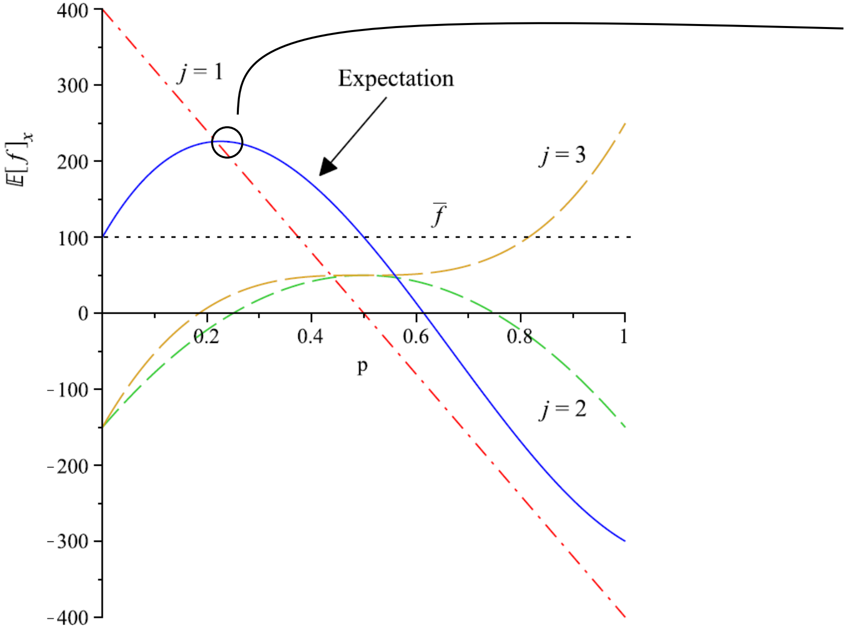
- **Selection operator**



- **We can design a selection operator selecting the individuals according to the expected fitness value after the mutation**

Mutation Operator

- **Mutation operator**
- **Given one individual *x*, we can compute the expectation against *p***



1. **Take the probability *p* for which the expectation is maximum**
2. **Use this probability to mutate the individual**

- **If this operator is used the expected improvement is maximum in one step**

(Sutton, Whitley and Howe in GECCO 2011)

Guarded Local Search

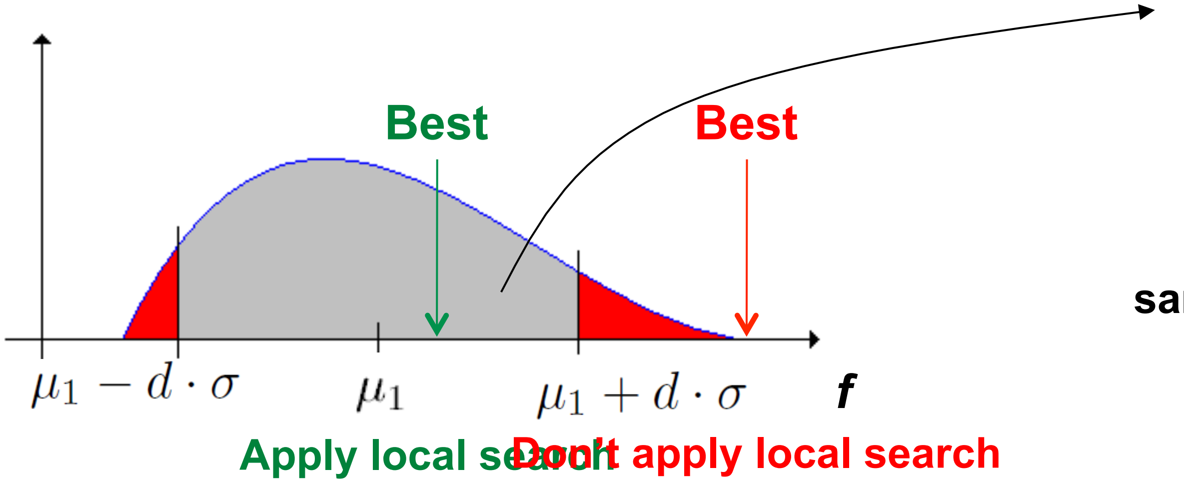
- **With the Elementary Landscape Decomposition (ELD) we can compute:**

$$\mu_c = \underset{y|\mathcal{H}(y,x)=r}{\text{avg}}\{f^c(y)\} = \binom{n}{r}^{-1} \sum_{p=0}^{n} \mathcal{K}_{r,p}^{(n)} \left(f^c\right)^{(p)}(x)$$

- **With the ELD of *f* and *f*² we can compute for any sphere and ball around a solution:**

μ_1 **: the average** $\sigma = \sqrt{\mu_2 - \mu_1^2}$ **: the standard deviation**

- **Distribution of values around the average**



Chebyshev inequality

$$100\left(1 - \frac{1}{d^2}\right)\%$$

At least 75% of the samples are in the interval

$$[\mu_1 - 2\sigma, \mu_1 + 2\sigma]$$

Guarded Local Search: Experimental Setting

- **Steady state genetic algorithm: bit-flip (*p*=0.01), one-point crossover, elitist replacement**
  - **GA (no local search)**
  - **GLS*r* (guarded local search up to radius *r*)**
  - **LS*r* (always local search in a ball of radius *r*)**

- **Instances from the Software-artifact Infrastructure Repository (SIR)**
  - **printtokens**
  - **printtokens2**
  - **schedule**
  - **schedule2**
  - **totinfo**
  - **replace**

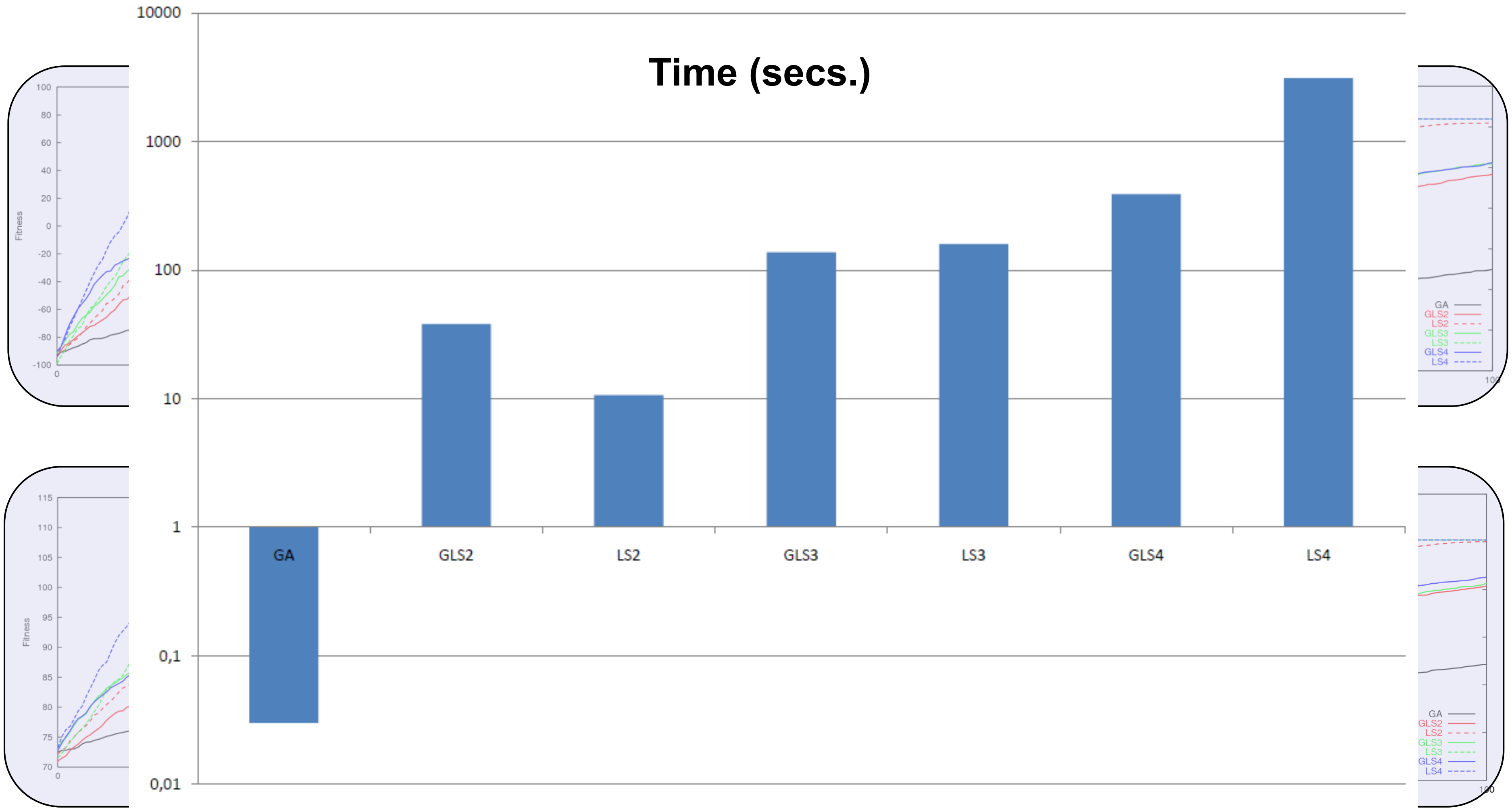
Oracle cost c=1..5

***n*=100 test cases**

***k*=100-200 items to cover**

100 independent runs

Guarded Local Search: Results



Conclusions & Future Work

Conclusions

- Landscape theory provides a promising technique to analyze SBSE problems
- We give the elementary landscape decomposition of the test suite minimization problem
- Using the ELD we can efficiently compute statistics in the neighbourhood of a solution
- We provide a proof-of-concept by proposing a Guarded Local Search operator using the information gained with the ELD

Future Work

- The main drawback of the GLD is runtime: parallelize computation with GPUs
- Expressions for higher order moments (ELD of f^c)
- Remove the current constraint on the oracle cost
- Connection with moments of MAX-SAT

Thanks for your attention !!!