CSE/ECE 848 Introduction to Evolutionary Computation

Module 3 - Lecture 10 - Part 3

Evolutionary Strategies - Strategy Parameter Handling

Wolfgang Banzhaf, CSE
John R. Koza Chair in Genetic Programming

Strategy Parameters

- The strategy parameters were included on the vector so they, too, could adapt using the same methods for the values. Each value would have associated strategy parameters
- ES was the first to incorporate strategy into the overall approach to solving a problem

Strategy Parameters II

Define an individual as a combination of three vectors:

$$\vec{a} = (\vec{x}, \vec{\sigma}, \vec{\alpha})$$

$\vec{x}$: vector of object variables

$\vec{\sigma}$: vector of step sizes

$\vec{\alpha}$: vector of inclination angles

Strategy Settings, one σ

One σ: one stepsize for the whole population

- $a = ((x_1 \ldots x_n), \sigma)$ -> $a' = ((x'_1 \ldots x'_n), \sigma')$ with
- $\sigma' = \sigma * \exp(\tau_0 * N(0,1))$ and
- $x'_i = x_i + \sigma' * N(0,1)$
- τ_0 is the learning rate

The Learning Rate τ_0

The parameter τ_0 affects the speed of step-size adaptation:

- τ_0 bigger: faster but more imprecise
- τ_0 smaller: slower but more precise
- How to choose τ_0?

According to recommendation of Schwefel:

(n: dimensionality of the solution vector)

$$\tau_0 = \frac{1}{\sqrt{n}}$$

Multiple σ

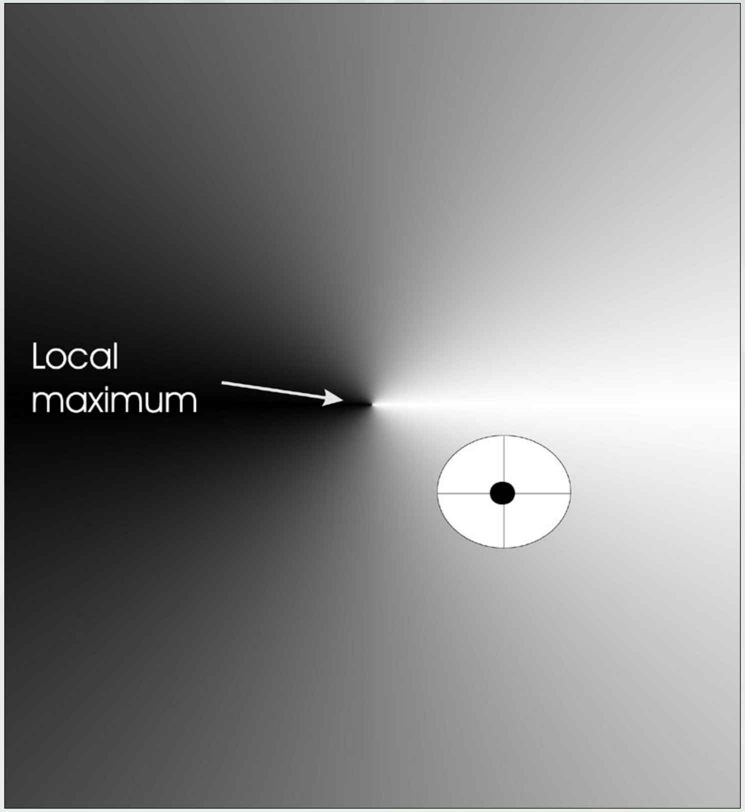
n σ values: standard mutation with individual σ_i for each parameter

- $\sigma_i' = \sigma_i \cdot \exp(\tau' \cdot N(0,1) + \tau \cdot N_i(0,1))$
- τ' is a global learning rate
  - only one realization
- τ is a local learning rate
  - n realizations

Mutation Case 1: Uncorrelated mutation with one σ

- Genomes: $\langle x_1,\ldots,x_n, \sigma \rangle$
- $\sigma' = \sigma \cdot \exp(\tau \cdot N(0,1))$
- $x_i' = x_i + \sigma' \cdot N(0,1)$
- Typically the "learning rate" $\tau \propto 1/ n^{½}$
- With a boundary rule $\sigma' < \varepsilon_0 \Rightarrow \sigma' = \varepsilon_0$

Mutants with Equal Likelihood



Smith/Eiben, Introduction to Evolutionary Computation, Springer

Pros and Cons of one σ

Advantages:

- simple mechanism
- usually fast and precise adaptation

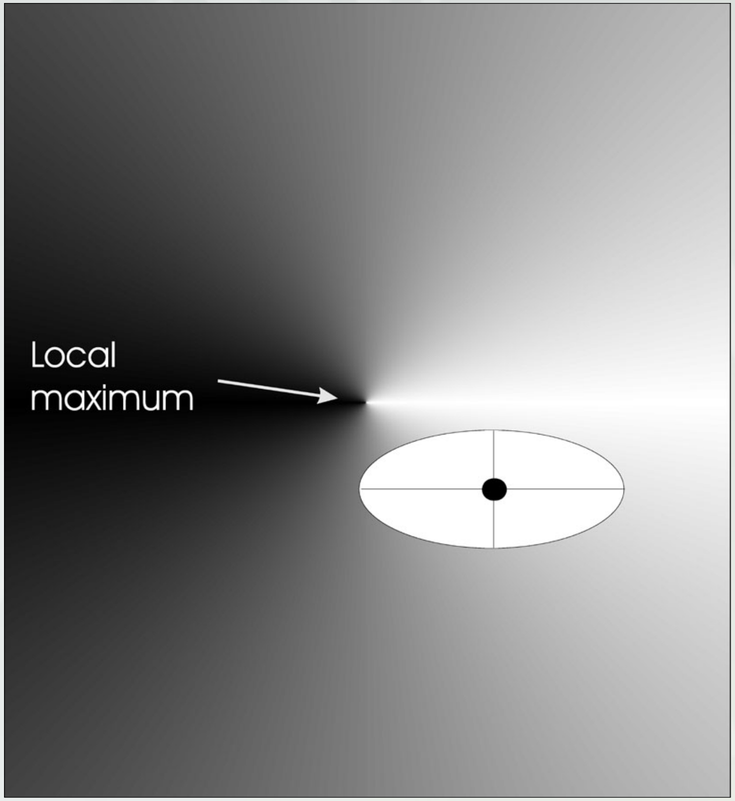
Disadvantages

- bad performances on complicated contours
- bad adaptation on widely differing objective values

Mutation case 2: Uncorrelated mutation with n σ's

- Genomes: $\langle x_1,\ldots,x_n, \sigma_1,\ldots, \sigma_n \rangle$
- $\sigma_i' = \sigma_i \bullet \exp(\tau' \bullet N(0,1) + \tau \bullet N_i(0,1))$
- $x_i' = x_i + \sigma_i' \bullet N_i(0,1)$
- Two learning rate parameters:
  - τ' overall learning rate
  - τ coordinate wise learning rate
- $\tau \propto 1/(2n)^{½}$ and $\tau' \propto 1/(2n^{½})^{½}$
- And $\sigma_i' < \varepsilon_0 \Rightarrow \sigma_i' = \varepsilon_0$

Mutants with Equal Likelihood



Smith/Eiben, Introduction to Evolutionary Computation, Springer

Pros and Cons of Individual σ

Advantages

- individual scaling
- better global convergence

Disadvantages

- slower
- cannot rotate to the coordinate system

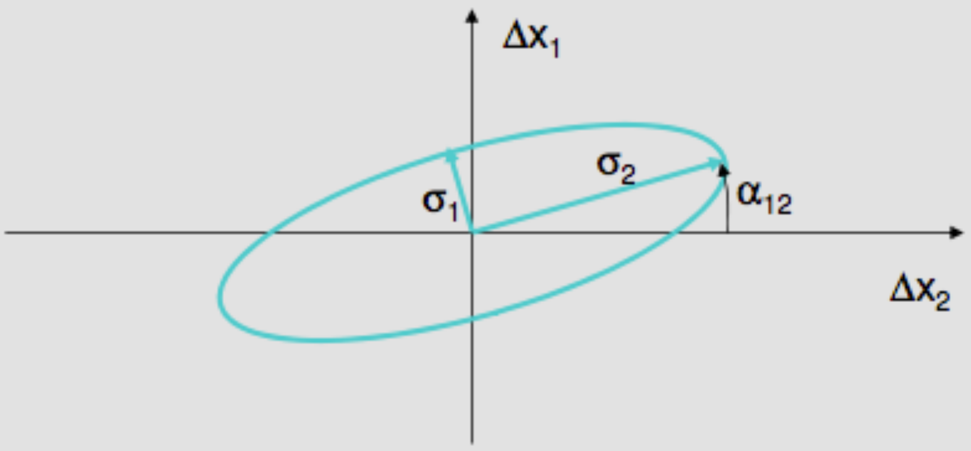
Mutation case 3: Correlated mutations

- Genomes: $\langle\, x_1,\ldots,x_n,\ \sigma_1,\ldots,\ \sigma_n\ ,\alpha_1,\ldots,\ \alpha_k\,\rangle$
- where $k = n \cdot (n-1)/2$
- with a covariance matrix C defined as:
  - $c_{ii} = \sigma_i^2$
  - $c_{ij} = 0$ if i and j are not correlated
  - $c_{ij} = \frac{1}{2} \cdot (\ \sigma_i^2\ -\ \sigma_j^2\ ) \cdot \tan(2\ \alpha_{ij})$ if i and j are correlated
- Note the numbering / indices of the α's

Interpretation of Angles

- Interpretation of rotation angles α_{ij}
- Mapping onto convariances according to

$$c_{ij(i \neq j)} = \frac{1}{2}(\sigma_i^2 - \sigma_j^2)\tan(2\alpha_{ij})$$

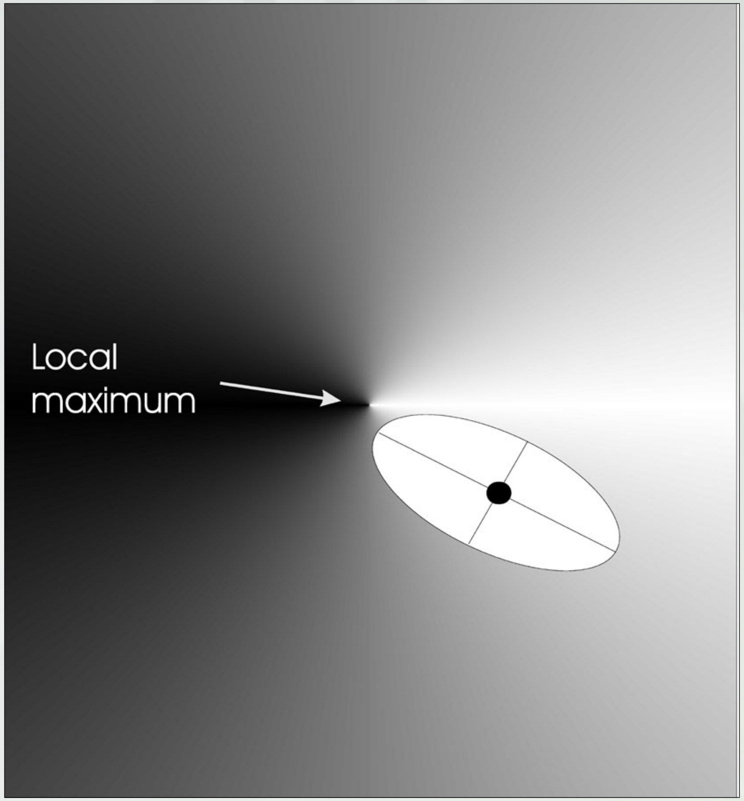


Correlated Mutations II

The mutation mechanism is then:

- $\sigma_i' = \sigma_i \cdot \exp(\tau' \cdot N(0,1) + \tau \cdot N_i(0,1))$
- $\alpha_j' = \alpha_j + \beta \cdot N(0,1)$
- $x' = x + N(0,C')$
  - x stands for the vector $\langle x_1,\ldots,x_n \rangle$
  - C' is the covariance matrix C after mutation of the α values
- $\tau \propto 1/(2n)^{½}$ and $\tau' \propto 1/(2n^{½})^{½}$ and $\beta \approx 5°$
- If $\sigma_i' < \varepsilon_0 \Rightarrow \sigma_i' = \varepsilon_0$ and
- If $|\alpha_j'| > \pi \Rightarrow \alpha_j' = \alpha_j' - 2\pi\,\mathrm{sign}(\alpha_j')$

Mutants with Equal Likelihood

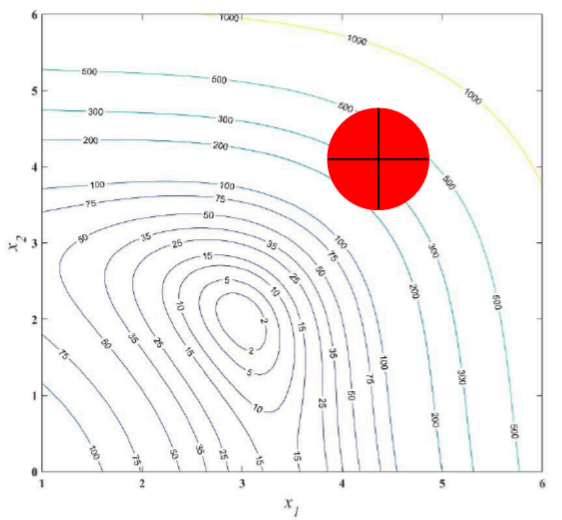


Smith/Eiben, Introduction to Evolutionary Computation, Springer

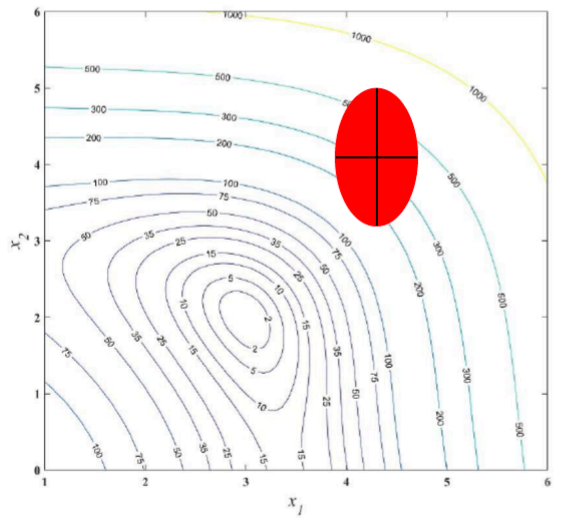
Covariance Matrix Adaptation

- Develops a covariance matrix C between the σ values:
- Modifies the shape of the distribution by modifying the co-variance matrix of the individual σ. In particular, allows for rotation to deal with misaligned function.
- Mutation is based on the developed covariance matrix.

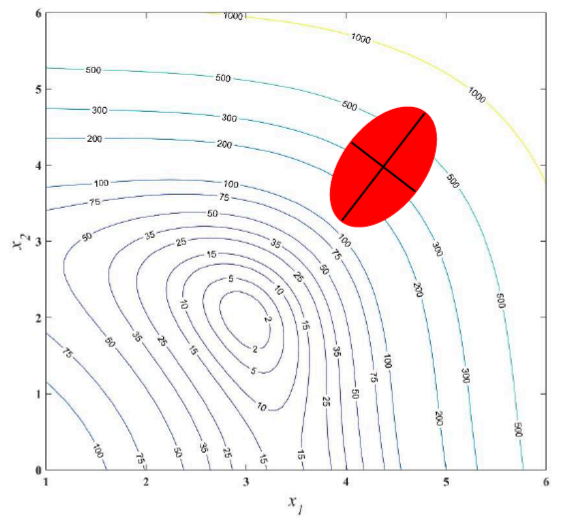
Uncorrelated with one σ



Uncorrelated with one 2 σ



Correlated with one 2 σ



Pros and Cons of Correlated Mutations

Advantages:

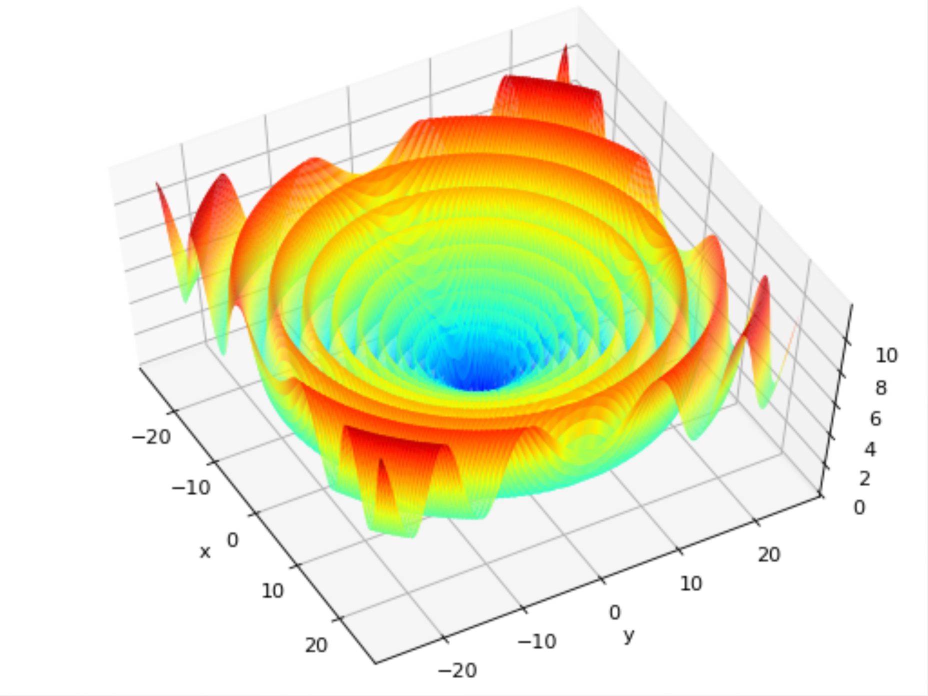
- individual scaling
- rotation
- better convergence

Disadvantages:

- much slower
- mutation effort scales quadratically
- because of speed, self-adaptation slow

Schaffer Test Function

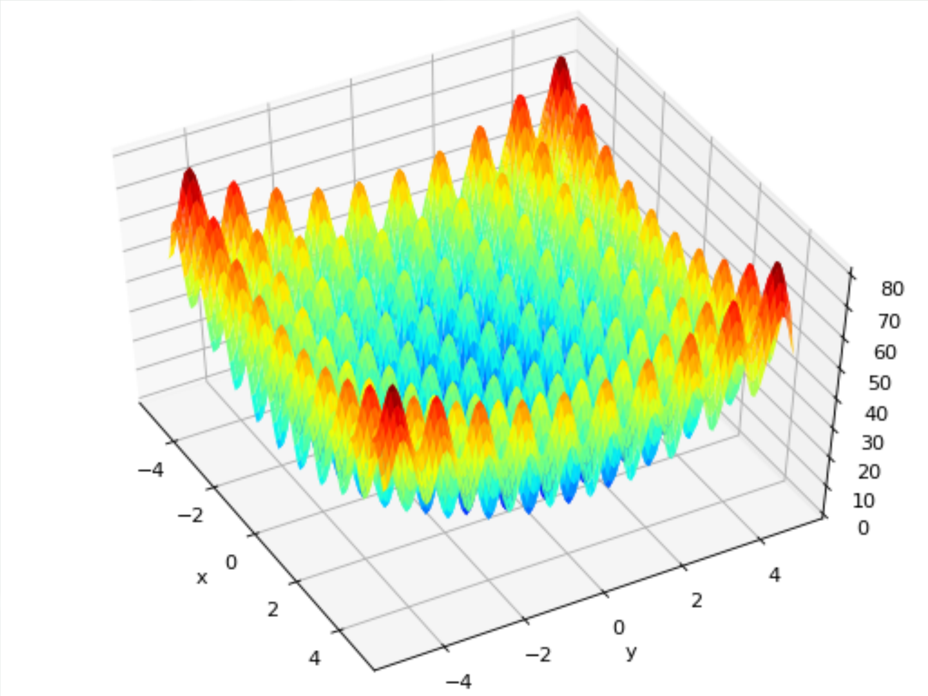
Type	minimization
Range	$x_i \in [-100, 100]$
Global optima	$x_i = 0, \forall i \in \{1 \dots N\}, f(\mathbf{x}) = 0$
Function	$f(\mathbf{x}) = \sum_{i=1}^{N-1} (x_i^2 + x_{i+1}^2)^{0.25} \cdot \left[\sin^2(50 \cdot (x_i^2 + x_{i+1}^2)^{0.10}) + 1.0\right]$



From DEAP Documentation

Rastrigin Test Function

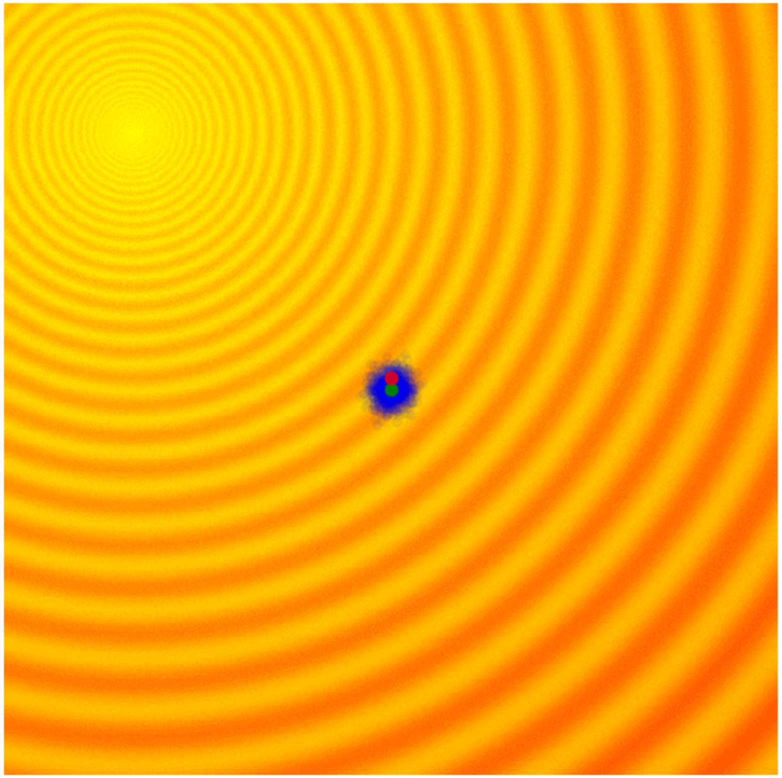
Type	minimization
Range	$x_i \in [-5.12, 5.12]$
Global op-tima	$x_i = 0, \forall i \in \{1 \dots N\}, f(\mathbf{x}) = 0$
Function	$f(\mathbf{x}) = 10N + \sum_{i=1}^{N} x_i^2 - 10\cos(2\pi x_i)$



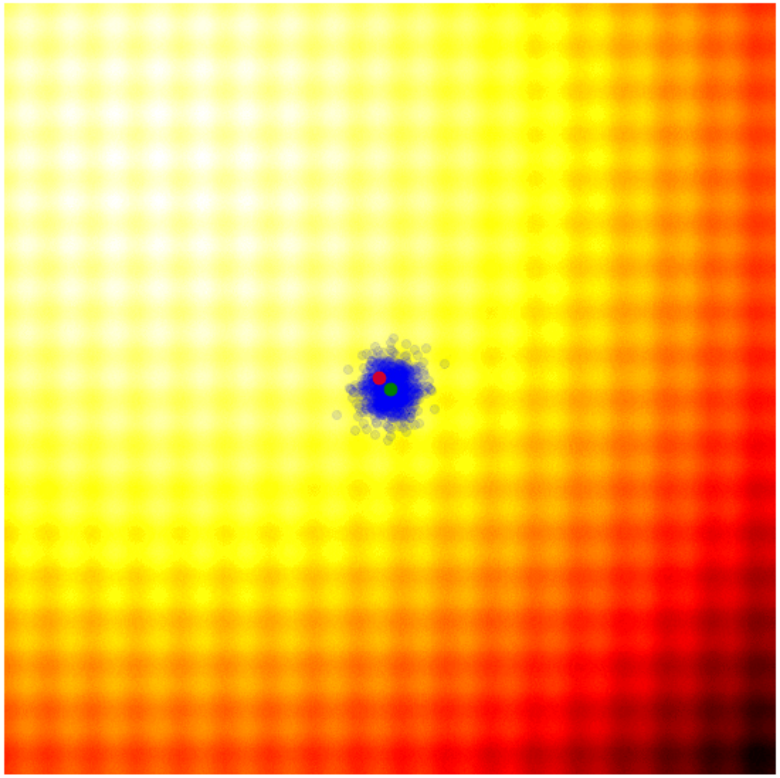
From DEAP Documentation

Simulation of Simple ES (without adaptation), adopted from otoro.net

Schaffer Function



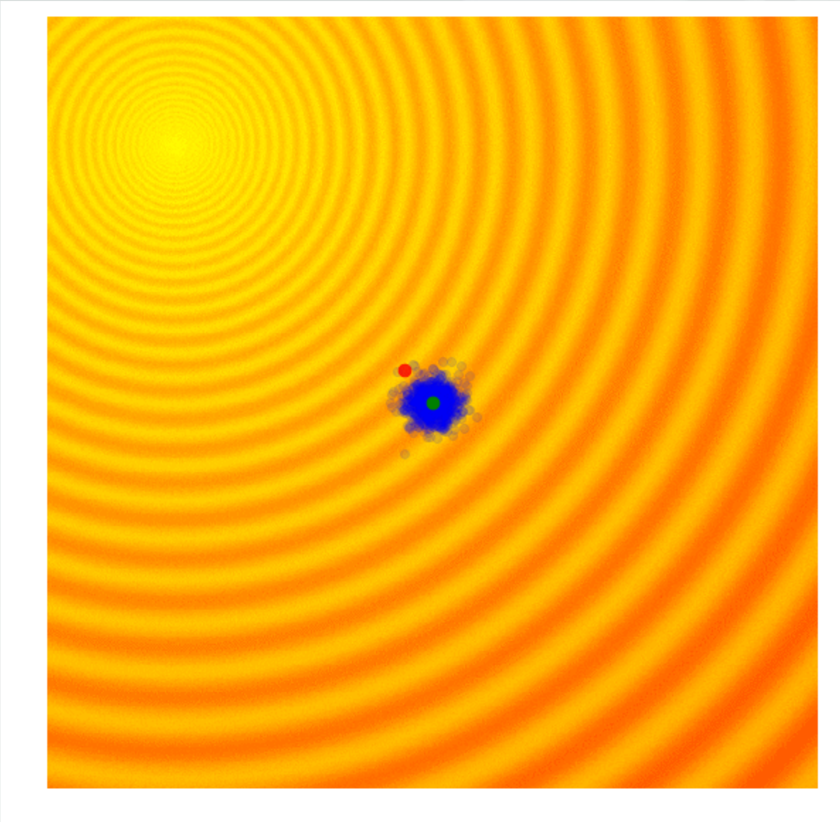
Rastrigin Function



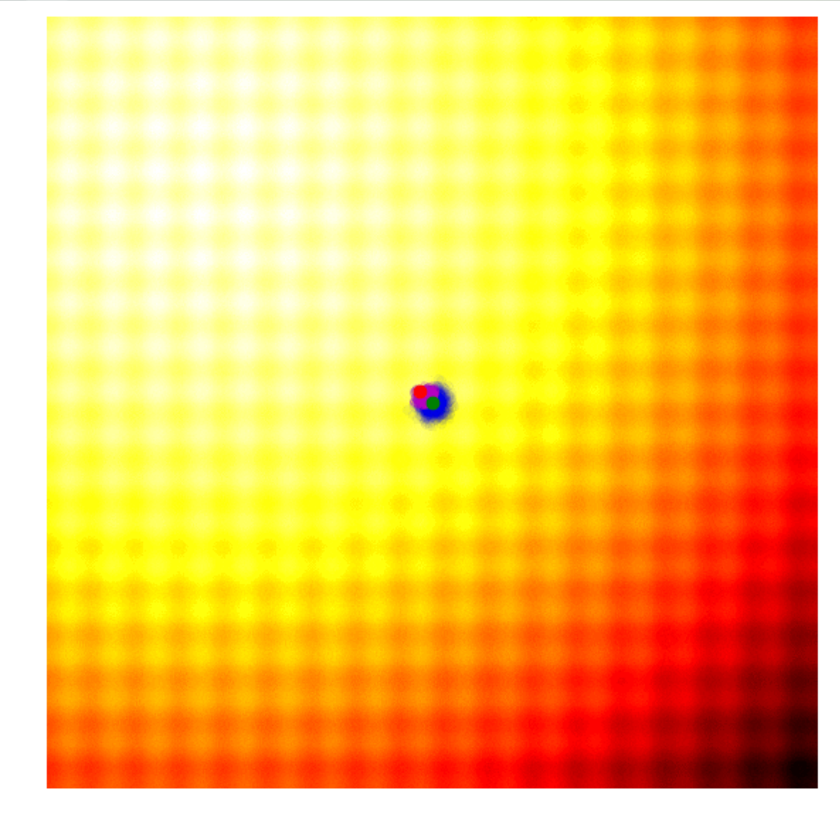
Red dot: Best so far — Green dot: Average of population

Simulation of CMA ES, adopted from otoro.net

Schaffer Function



Rastrigin Function



Red dot: Best so far — Green dot: Average of population