Universitat Autònoma de Barcelona

A Statistical Framework for Terminating Evolutionary Algorithms at their Steady State

A dissertation submitted by **David Roche Valles** at Universitat Autònoma de Barcelona to fulfil the degree of **Doctor en Informàtica**.

Bellaterra, June 2015

Co-director:	**Dra. Debora Gil Resina**
	Universitat Autònoma de Barcelona
	Dep. Ciències de la Computació & Computer Vision Center
Co-director:	**Dr. Jesús Giraldo Arjonilla**
	Universitat Autònoma de Barcelona
	Institut de Neurociències & Unitat de Bioestadística
PhD candidate:	**David Roche Valles**
	Universitat Autònoma de Barcelona
	Dep. Ciències de la Computació & Institut de Neurociències

Universitat Autònoma de Barcelona

This document was typeset by the author using LaTeX 2$_\varepsilon$.

The research described in this book was carried out at the Dept. Ciències de la Computació, Universitat Autònoma de Barcelona and Institut de Neurociències.

Acknowledgment

I would like to thank my advisors, Professor Jesús Giraldo and Professor Debora Gil. Their contributions to this Thesis have been absolutely definitive and were not only for research guidance but also for personal aspects. I consider them as friends. This work has lasted a long time since I stated my scientific career. My advisors have been close to me for all time supporting me in several aspects. I also would like to thank my family for supporting me during these years. My wife Amparo, my four children, Mar, Kolbe, Claudia and Heidi, my parents and my brothers. All of them have been essential for the successful consecution of this work. I want to remember my friends that in different ways have contributed at least at personal help. I also thank to Dep. Ciències de la Computació & Institut de Neurociències for their support.

This work was supported by the Spanish projects TIN2012-33116, TIN-2009-13618, and CSD2007-00018.

Abstract

As any iterative technique, it is a necessary condition a stop criterion for terminating Evolutionary Algorithms (EA). In the case of optimization methods, the algorithm should stop at the time it has reached a steady state so it can not improve results anymore. Assessing the reliability of termination conditions for evolutionary algorithms is of prime importance. A wrong or weak stop criterion can negatively affect both the computational effort and the final result.

In this Thesis, we introduce a statistical framework for assessing whether a termination condition is able to stop Evolutionary Algorithms (EA) at its steady state. In one hand a numeric approximation to steady states to detect the point in which EA population has lost its diversity has been presented for EA termination. This approximation has been applied to different EA paradigms based on diversity and a selection of functions covering the properties most relevant for EA convergence. Experiments show that our condition works regardless of the search space dimension and function landscape and Differential Evolution (DE) arises as the best paradigm. On the other hand, we use a regression model in order to determine the requirements ensuring that a measure derived from EA evolving population is related to the distance to the optimum in x-space. Our theoretical framework is analyzed across several benchmark test functions and two standard termination criteria based on function improvement in f-space and EA population x-space distribution for the Differential Evolution (DE) paradigm. Results validate our statistical framework as a powerful tool for determining the capability of a measure for terminating EA and select the x-space distribution as the best-suited for accurately stopping DE in real-world applications.

Contents

List of Figures

Chapter 1

Introduction

There are a large variety of applications that require finding optimal parameters for the numerical methods involved in their implementation. These optimal parameters can be usually defined as the minimum of a cost function or energy that mathematically models a trade off between several quality standards that the application has to meet. In some cases these standards reduce to finding those parameters settings that ensure an optimal algorithm performance in terms of (maximum) accuracy and the cost function can be expressed as a least squares approximation problem. This is the best case, since least squares approximation yields convex functions which are known to have a unique solution that can be found using local optimization techniques as gradient descent. Local optimization techniques find the optimum of a function by iteratively evolving an initial point along the direction of maximum function decay, which is usually given by the function gradient. In case that the function has a continuous derivative, gradient descent algorithms are fast and its convergence to the closest local minimum is guaranteed. It follows that they are one of the most widely used high-speed optimization methods.

Unfortunately, the rate of success of local methods much depends on the function profile, which should be as convex as possible. In case of multimodal functions presenting multiple local minima, gradient descent success strongly depends on the proximity of the initial seed to the target optimum. Seed points distant from the optimum may lead the local method to premature convergence far from the target minimum. In case the energy is sensitive to experimental noise, premature convergence may occurs even for closest seeds. Another limitation of local methods, is the need to deal with functions that have a differentiable formulation so that the gradient can be computed. This condition discards problems requiring simultaneous minimization of several energy functions, black-box parameter optimization problems and also complex cost functions sensitive to noise. Examples of applications requiring optimization of functions bad posed for local methods can be found in areas as different as image processing and pharmacology.

In the image processing side, a common task in medical diagnosis decision sup-

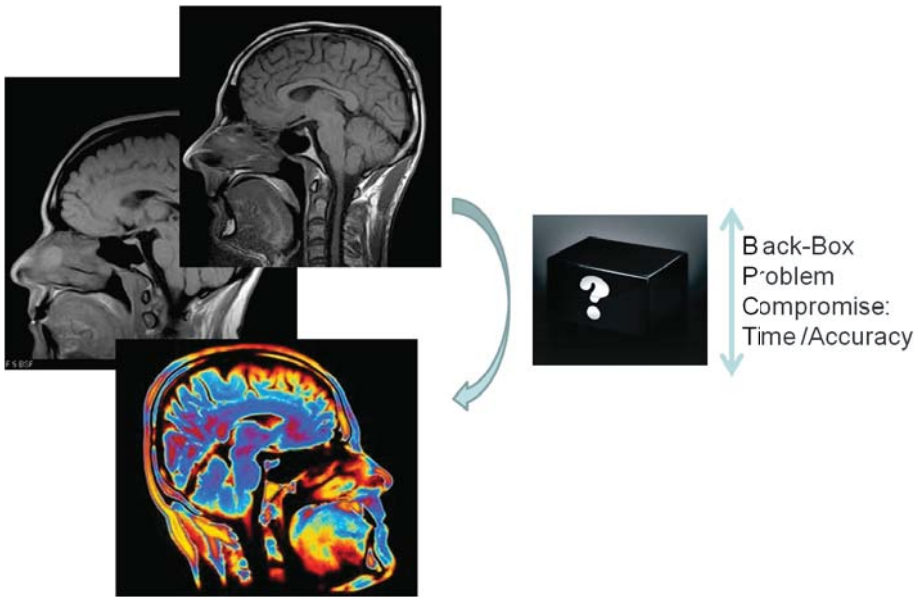
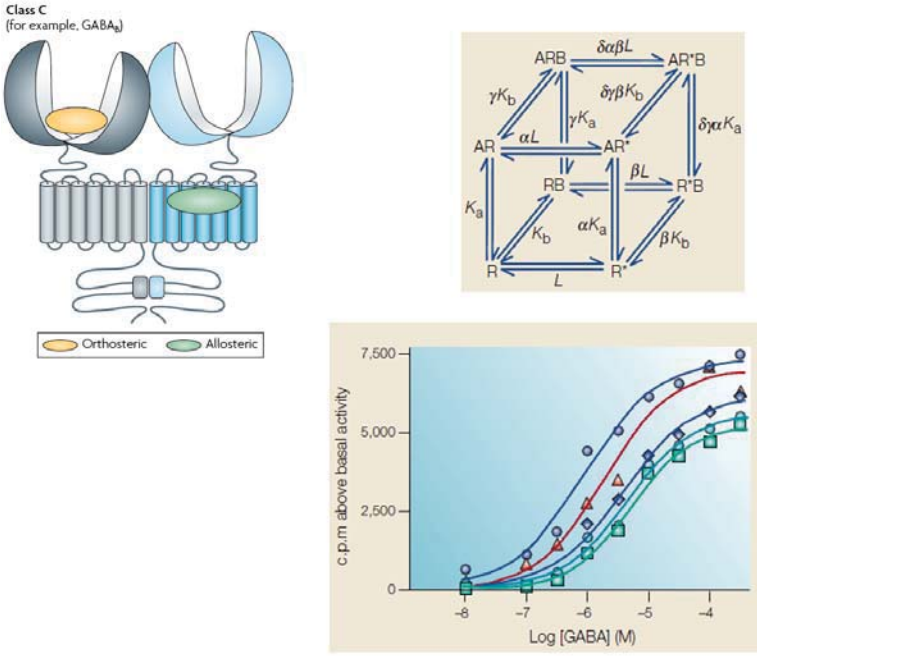


Figure 1.1: Multiobjective Black-Box problem: Efficient registration of medical multimodal scans

port applications is the multimodal registration of data acquired with different devices. Volume registration consists in finding the spatial (either in 2D for images or in 3D for volumes) transformation that best matches the anatomy of a given subject scanned using two different medical image devices as illustrated in figure 1.1. This figure 1.1 shows 2 brain scans acquired using a standard Magnetic Resonance , MR (gray intensity images) and a functional MR (colored images). MR scans are used to capture anatomy, while Functional MR are used to detect areas of unusually high metabolic activity prone to be tumors. A combination of both modalities increases Cancer diagnostic yield and it is a common procedure. The spatial transformation that matches 2 different scans is given as the minimum of an energy functional that splits into a data term matching the transformed and target medical scans and a regularity term determining some properties (usually smoothness) of the transformation. The definition of this kind of energy functionals combining data and regularity terms arises in other image processing tasks (like motion estimation and image segmentation) and has two critical issues. First, the choice of the matching metric defining the data term and second, setting the parameters that define the trade-off between data and regularity constraints. The best choice from an application point of view should reach a maximum accuracy in the matching transformation with the minimum computational cost. Such a compromise between accuracy and computational cost can not be easily formulated in a compact way suitable for local optimization methods. On one hand, accuracy and time metrics and scales are not comparable and thus, they should not be combined in a single cost function. Instead they should be rather

minimized simultaneously in a multi-objective approach. On the other hand, the cost function depends upon the output of a registration algorithm whose definition also has to be optimized. It follows that the sensible search, off a method achieving a compromise between accuracy and efficiency is. This is an example of black-box non-analytic optimization in a multi-objective scheme.



$$E = \frac{E_m \chi^n (KM + \varepsilon_A M[A] + \varepsilon_B K[B] + \alpha\varepsilon_A\varepsilon_B\delta[A][B])^n}{(KM + M[A] + K[B] + \alpha[A][B])^n + \chi^n (KM + \varepsilon_A M[A] + \varepsilon_B K[B] + \alpha\varepsilon_A\varepsilon_B\delta[A][B])^n}$$

Figure 1.2: Life-cycle of the parameter estimation in pharmacological modelization problem.

In the pharmacological side, a major research area is finding the parameters of a given pharmacological model explaining the response of a pharmacological system to a given dose of a chemical compound (figure 1.2). This adjustment is usually done by fitting a model to a experimental data with unknown parameter that should explain the experimental data.This adjustment is done by minimizing an energy error that is multimodal. Low multimodality (figure 1.3 on the left) could be handled by gradient descent using a multiple seed strategy. However, in pharmacological experiments, a main concern is the sensitivity of the energy function to noise in experimental data which propagates to the energy response introducing a saw-shape pattern (figure 1.3 on the right). This makes multiple seed gradient descent strategy falls into local

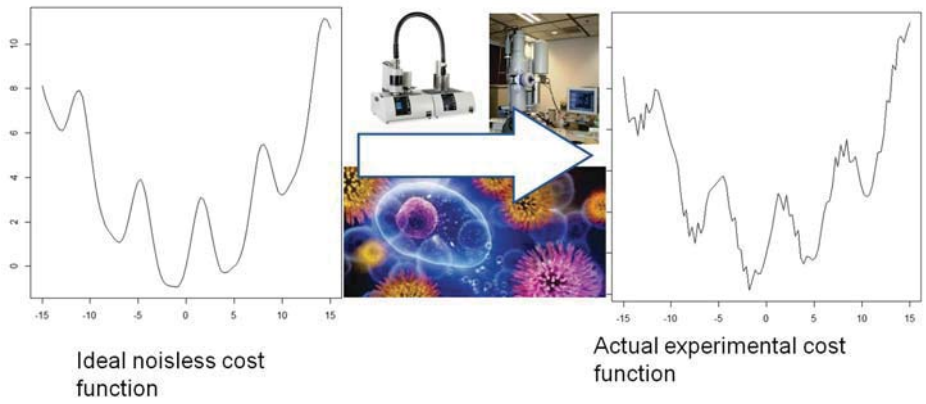


Figure 1.3: Impact of Experimental noise in the estimation of pharmacological models parameters.

minima introduced by noise that might be far away from the true global ones. In such noise multimodal cases, Evolutionary Algorithms (EAs) can contribute to find optimal initial seeds for further refinement using gradient descent methods.

The above problems could be solved by exhaustive search of the parameter space, which is computationally unfeasible. This issue has motivated the development of more effective global exploration techniques taking into account the values of the cost function. These global techniques are usually inspired in biological mechanism and are named as Evolutionary Algorithms (EAs).

EAs are a class of stochastic optimization methods that simulate the process of natural evolution [6]. Unlike gradient descent methods that evolve a single initial value each time, EAs maintain a population of possible solutions that evolve according to bio-inspired simulated rules. This population-based strategy implements a global approach to multiple minima optimization problems. A gradely definition of Evolutionary Algorithms taking into account the connection between the essentials and the applications can be stated as follows:

1. **Essential Definition** An Evolutionary Algorithm (EA) is an iterative population-based technique to find (and explore) as well as possible in a given search space the most fit set of points according to a cost function.

2. **Problem Solver** When a problem is presented in practically all situations, an exploration of a set of possible solutions is needed and for each possible solution its goodness is valuable. This set of possible solutions, when is defined, is the search space and we are in the first definition.

3. **Optimization Technique**. From another point of view, as mentioned above, Evolutionary Algorithms can be viewed as a optimization technique if a functions is given.

In the literature several possible definitions of Evolutionary Algorithms can be found [57]. Concepts related to EA could be: population-based optimization technique, computer intelligence, soft computing, machine learning technique, heuristic algorithms, and so on. Concepts and terminology are imprecise and context-dependent. A possible complete and compact definition could be the following. Evolutionary Algorithms (EAs) are a stochastic, population-based and an iterative process to explore a search space defined by a set of points to find a good zone determined by a function of its goodness simulating the process of natural evolution. The numeric implementation of EAs maintain a population of possible solutions that iteratively evolve according to rules of selection and other natural evolution inspired operators, such as recombination and mutation. One of the main reasons for the wide used range of EAs is its simplicity for handling complex problems by only determining the following points: codification of the space of possible solutions, rules for conducting the evolution of the initial population, the cost function to be optimized and a criterion to stop evolution. A main advantage over other optimization techniques is that EA admit search spaces of discrete, continuous and qualitative elements.

Codification for representing the search space of possible solutions. A codification is a mapping from the space of possible solutions (whose bio-inspired name is phenotype) to a space of individuals in the EA process (genotype). A main advantage over other search (optimization) techniques is that, in general terms, EA can deal with practically all kinds of candidate solutions if a good codification is found. Among possible representations we can find real-valued vectors, permutations, bit representation, trees and a lot of different and creative possibilities. In real-valued optimization cases this mapping is the identity function because possible solutions as well as representation into the EA process are real-valued vectors plus, in some advanced cases, some added parameters if EA paradigm needs them to control the search process as in the case of Covariance Matrix Adaptation Evolutionary Strategies [45].

Rules for conducting the evolution of the initial population. Such rules are implemented by bio-inspired operators (selection, recombination and mutation) that conduct the evolution of the possible solutions to better zones of the search space. These operators have a strong dependency on the codification of the search space because they need to be applied over the represented individuals. This simple mechanism allow easy parallelization of the algorithm.

Cost function to be optimized. This cost function assigns the goodness of each solution and is the main criterion for selecting the individuals that survive across EA generations. EAs don't need information about derivatives like gradient-descent methods. In fact, there are no restrictions in the cost function, which could be anything that can be evaluated on the search space elements. This is a main advantage over other optimization technique and allows EA to be suitable for optimizing complex functions such as multimodal, multi-objective, constrained, and specially non-analytical functions, including black-box functions. By their ability to optimize several functions, EAs have been successfully applied to a wide range of real life problems, such as parameter estimation [42], text recognition [47] or image processing [11], among

others.

A criterion to stop evolution. As any iterative technique, a stop criterion for terminating EA numeric implementation is mandatory. In the case of optimization methods, the algorithm should stop at the time it has reached a steady state so it can not improve results anymore.

1.1 Goal and Contributions

The goal of this Thesis is assessing the quality of existing stopping conditions for terminating EA at its steady state. The diagram shown in figure 1.4 sketches the main points that should be checked in order to assess the quality of a stopping condition in this context. We contribute in the following aspects (also sketched in figure 1.4) to each quality stage.

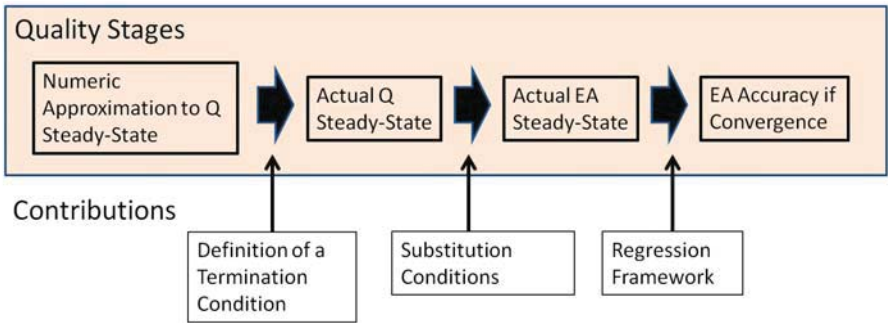


Figure 1.4: Flowchart of the main points that should be checked in order to assess the quality of a stopping condition for terminating EA at its steady state.

1. Definition of a Termination Condition

1. **EA population steady state in terms of Cauchy convergence of sequences of real numbers** EA convergence is a complex and unless partial unsolved problem. In general, EA optimality is difficult to assess. Under our point of view, the algorithm should stop at the time it has reached a steady state so it can not improve results anymore. In these scenario, a stop criterion should ensure EA has reached a steady state. Defining steady states is not straightforward for EA paradigms. That definition implies taking into account the stochastic aspect of the EA process and the diversity of the population. We provide a formulation of EA steady state inspired in the concept of Cauchy sequences. The state of the EA process across the iterations is kept in a random variable quantity that captures the diversity of the evolution. This quantity is posed in terms of real number sequences and its stabilization (EA population steady state) in terms of Cauchy convergence.

2. **Numeric approximation to stabilization of sequences of real numbers** We give a numerical formula to determine the moment a quantity has reached its steady state. The formula (called range formula) is given in terms of the quantity range of variability across iterations and approximates the continuous definition of Cauchy sequences. That formula depends on two parameters that can be adjusted according to the accuracy of the approximation to Cauchy steady state. We provide a statistical way in terms of a hypothesis test based on proportions to set the parameter optimal values ensuring a given accuracy of the approximation. In this manner, we ensure that such accuracy will be kept for new cases within the class of functions and applications chose to do the inference test.

2. Evaluation Mechanism In this contribution we propose a regression framework for determining whether a measure derived from EA evolving population statistically behaves like its accuracy.

1. **Substitution Conditions** We provide three conditions that evaluate in several aspects (from more theoretical to more empirical) the capability of a quantity Q to act as a measure of EA accuracy. First, a regression model is used to relate Q values to EA accuracy given by the distance to the function optimum. Second, the analysis of the regression coefficients provides a first theoretical condition for substitution (strong condition). Then, such condition is relaxed to a weak condition given in terms of Cauchy sequences. Finally, we also propose an empirical substitution criterium that can be computed using the approximation formula.

2. **Regression framework for prediction of the actual EA Accuracy form the alternative Quantity**. Based on regression framework modelling our alternative quantity (Q) over the distance to the (known) function minimum in x-space (Ref) we can specify a confidence interval for the response variable (Ref) from the alternative quantity Q values.

Aside, we also contribute to the selection of the best suited quantity for terminating EA in black-box problems in the following aspects:

3. Evaluation of current criteria We have applied our evaluation framework to representative set of EA paradigms and alternative quantities reported in the literature. Our experiments on a benchmark set of functions [24] covering the main landscapes influencing EA convergence [37] select a known paradigm and quantity as the best suited for solving black-box problems.

4. Application to Pharmacology. The selected EA paradigm and quantity have been applied to a pharmacological model. The problem is fitting a model to a experimental data with unknown parameter that should explain the data. Finally, the solution is conformed hybridizing our termination condition to stop the EA algorithm with a non-linear regression step for precisely find the best solution. Several publications prove the utility of the framework to solve the pharmacological parameter estimation problem.

The rest of the Thesis is organized as follows. In Chapter 2 we briefly review concepts that we use through the thesis as statistical tools like linear regression or hypothesis testing, existing evolutionary algorithms and stop criteria. Chapter 3 introduces the first contribution of this thesis. The Chapter presents termination conditions based on steady states covering a formulation of steady states inspired in the concept of Cauchy sequences and how to detect when this phenomenon occurs in an alternative measure that correlates the true EA accuracy with a empirical and statistical two parameter framework. Two experiments are carried out for distinct EA paradigms and cost functions. The first experiment validates our framework and the second one explores the suitable parameters and measures concluding that distribution-based measures and DE paradigm are the best choice for termination condition. This sort of measures are used in the rest of the Thesis. Chapter 4 describes a evaluation mechanism based on linear regression that offers a statistical method to evaluate when an EA process measure behaves like the true distance to the optimum and then, this distance can be substituted by the EA diversity measure. After the contributions presentation, in Chapter 5 our framework is applied in a pharmacological parameter setting problem. In this problem, a 9-parameter dose-response model is adjusted to a experimental data. Descend gradient techniques fail in the optimization task and our approximation shows a good solution with significant, valuable and pharmacological-sense parameters. This particular application is part of a collaborative project with a pharmacological group and published. Finally, in the last Chapter conclusions and future work are presented.

Chapter 2

Technical Review

As we said in Chapter 1 during this work we use a set of tools needed to define our framework and to get results in each of our contributions. Firstly, a summary of EA's definition and principles are presented. In the same manner, a revision of the stop criteria for EA's are visited. Finally, the common statistical tools Linear Regression Analysis and Hypothesis Testing are listed.

2.1 Existing Evolutionary Algorithms

Classical Evolutionary Algorithms and more recent EAs could be a first classification for the existing Evolutionary Algorithms. Evolutionary Computation [6] is a search technique algorithm based on the principles of biological evolution that may be used to search for optimal solutions to a problem.In particular, the principle of the "survival of the fittest" proposed by Charles Darwin has especially captured for this technique. As mentioned above, Evolutionary Algorithms simulate evolution. Darwin inspired used to problem solving originated in 1950s and several references could be found [57]. For about 15 years different approaches were conducted separately until from the nineties were considered as the same field called Evolutionary Algorithms. These different approaches were Genetic Algorithms by John Holland (1975), Evolutionary Strategies by Ingo Rechenberg and Hans Paul Schwefel, Evolutionary Programming by Lawrence Fogel and finally about 1990 Genetic Programming.

In a search technique algorithm , a number of possible solutions are available and the task is to find the best solution possible in a fixed amount of time (optimization process). For a search space with only a small number of possible solutions, all the solutions can be examined with exhaustive search. In other cases is possible to use traditionally techniques (e.g., gradient descent) but derivatives are necessary.

Evolutionary search is usually better than random search and does not suffer from hill-climbing artifacts like gradient-based search. The key aspect of evolutionary search is that it is population-based and simulates the evolution of individual structures via processes of selection and reproduction

Evolutionary algorithms keep a population of structures that evolve according to rules of selection and other operators, such as recombination and mutation. Each

Scheme.png Scheme.png Scheme.png

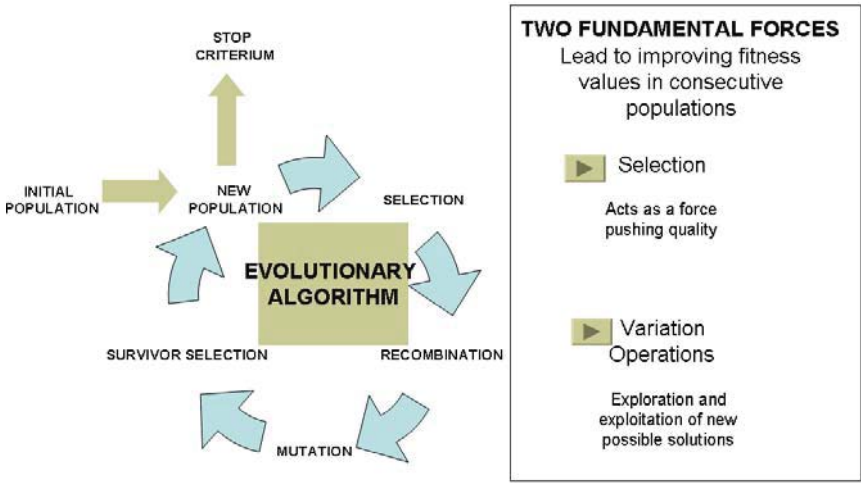


Figure 2.1: Iterative Mechanisms of a generic Evolutionary Algorithm.

individual in the population receives a measure of its fitness in the environment. Selection focuses attention on high fitness individuals and determines which individuals are chosen for producing offsprings. Recombination and mutation perturb those individuals, providing general heuristics for exploration. Finally a survival step decides who survives (among parents and offsprings) to form the new population. This process iterate until stop criteria occurs. Figure 2.1 outlines the scheme of a standard Evolutionary Algorithm (EA).

1. Representation. The first step in defining an EA is to link the "real world" to the "EA world", that is, to set up a bridge between the original problem context and the problem solving space. This first design step is commonly called representation. The most standard one is an array of bits or a real vectors representation. Array of other types and complex structures can be used.

2. Fitness function. A fitness function is a particular type of objective function that quantifies the optimality of a solution and, thus, introduces a criterion for selection of individuals. The fitness function should not only indicate how good the individual is, but also it should correspond to how close the individual is to the optimal one.

3. Population. The role of the population is to hold possible solutions. Given a representation, defining a population can be as simple as specifying how many individuals are (population size). In some sophisticated EA's a population has an additional spatial structure, with a distance measure or a neighborhood relation.

4. Selection operator. The role of selection is to allow the better individuals to become producers of the next generation (offspring creation). Selection is responsible for pushing quality improvements. In EA, selection can be probabilistic or deterministic. In both cases, high quality individuals are likely to produce new individuals. Each EA paradigm performs in a different way the selection operator.

5. Variation operators. The role of variation operators is to create new individuals from old ones(generate new candidate solutions). In general terms, a good compromise between exploration and exploitation of the search space is needed. This task is covered by these operators. The most common are Recombination and Mutation. Recombination exploits the area defined by the initial population, while mutation guarantees that other areas are explored.

   Recombination (or crossover) merges information from two parents into one or more offspring by means of a binary variation operator. By combining two individuals with different but desirable features, we can produce an offspring that combines both of those features. There are a lot of recombination operators depending on EA paradigm. Some paradigms apply recombination probabilistically, that is, with an existing chance of not being performed. After recombination every offspring undergoes mutation. Offspring variables are mutated by small perturbations (size of the mutation step) with low probability (mutation rate). The representation of the variables (and not the population size) determines the mutation algorithm parameters. In some cases, the probability of mutating a variable is inversely proportional to the number of variables (i. e. representation space dimension), n. A mutation rate of $1/n$ (only one variable per individual is mutated) produces good results for a wide variety of test functions [6]. Although the size of the mutation step is usually difficult to choose, small steps are preferred.

6. Survival selection. The role of survivor selection is to determine which individuals among parents and offspring will be allowed in the next generation so that the population size is constant. This decision is usually deterministic based on their fitness.

7. Initialization. Initialization is kept simple in most EA applications: the initial population is seeded by randomly generated individuals.

8. Termination. Commonly used options for termination condition are: 1.Fixed number of iterations 2. Fixed number of fitness evaluations. 3.The population diversity drops under a given threshold(convergence).

The most significant differences from more traditional search and optimization methods are determined in the definition of the Selection and Mutation stages. Selection deals with the type of cost function, while variation deals with the exploration and diversity of the search space.

Related to the optimization process EAs could be classified in two main groups:

1. **Single-objective Strategies** In this case, the goal of the optimization process is to find the optimum (minimum or maximum) of a unique function f (constrained or unconstrained). The original algorithms of EC as Genetic Algorithms, Evolutionary Strategies and Evolutionary Programming were designed for this purpose.

2. **Multi-objective Strategies** When tackling real-world problems, particulary in some complex fields, the desired optimal configuration to be optimized may not be expressed in terms of a single objective. In this case we are interested in optimizing more than one function simultaneously. Regardless EA paradigms, practically, all of them have a multi-objective adapted version. EAs have recently become more widely used for their ability to work well with multi-objective problems.

In this work we will focus on evolutionary strategies for single-objective optimization. In particular, the evolutionary strategies we have chosen are Differential Evolution (DE), Particle Swarm Optimization(PSO) and Covariance Matrix Adaptation Evolutionary Strategy (CMAES).

1. **Differential Evolution (DE)** is a real parameter encoding evolutionary algorithm for global optimization over continuous spaces [60, 13]. In this paper, we use the 3-parameter DE scheme [60] for solving DE. For a real search space of dimension D, the population is randomly initialized with ND vectors (for ND the first algorithm parameter). Each vector v in the population is evolved by mutation and recombination operators. Given a mutation rate $F \in [0, 2]$ (second parameter of the algorithm), the mutation operator produces a new vector vm by adding a vector difference of two randomly chosen population vectors $v1$ and $v2$ to another randomly chosen vector $v3$:

$$vm = v1 + F(v2 - v3)$$

For the recombination step, a new vector vf is created from the mutation vector by means of a combination rate CR (third parameter of the algorithm) as follows:

$$vf_i = \begin{cases} vm_i & \text{if} \quad r_i < CR \text{ or } i = k \\ v_i & \text{otherwise} \end{cases}$$

for vf_i the i-th component of vf and $r_i \in [0, 1]$ a random number and k a random number uniformly distributed in $[1, D]$. Finally a selection operator is applied. The vector vf and the initial vector v are compared and the vector that better fits the objective function is selected and remains in the next population. This process is iteratively repeated until a stopping criterion is reached.

For all experiments in the next Chapters we have used the 3-parameter DE/rand/1/bin scheme reported in [60]. For a real search space of dimension D, the population is randomly initialized with ND vectors. Each vector in the population is evolved by mutation and recombination operators. The mutation rate is given by a parameter $F \in [0, 2]$ and the combination rate by $CR \in [0, 1]$. Following the literature [13], we have chosen the following values for DE parameters: F=0.9, CR=0.5.

2. **Particle Swan Optimization (PSO)** Particle Swarm Optimization is also bio-inspired and based in collective intelligence of some natural systems like groups of animals. In a basic Particle Swarm Optimization scheme for continuous space the population is randomly initialized with N individuals (vectors x_i, $i \in [1, N]$). The fundamental difference between PSO and other EAs is that other EAS do not model the dynamics of the movement through the search space. PSO leads with this aspect with the concept of velocity and neighborhood. Each individual has some inertia and tries to maintain its velocity (v_i), $i \in [1, N]$. Velocity indicates how will the next movement be and it is another parameter of each individual. As position vector x_i, $i \in [1, N]$ the velocity is random selected in the initial population or set to zero. The neighborhood of each individual H_i, $i \in [1, N]$ is defined by the number of the nearest individuals that are considered, σ, and serves to define the velocity parameter update. For each individual and for each iteration, the velocity is updated and the individual moves with the rule: $x_i \leftarrow x_i + v_i$. The velocity update is built with the scalar products between the real vector position x_i and the best-so-far position b_i in one hand and, in the other hand, the real vector position x_i and the best neighbor position h_i with the rule $v_i \leftarrow v_i + \phi_1 \circ (b_i - x_i) + \phi_2 \circ (h_i - x_i)$, $i \in [1, N]$, where ϕ_1 and ϕ_2 are the maximum influence values. The absolute values of the velocities v_i are maintained between 0 and v_max and ϕ_1 and ϕ_2 between 0 and $\phi_{1,max}$ and $\phi_{1,max}$ respectively.

   - Initialize random population position vectors x_i, velocities v_i and $b_i = x_i$ $i \in [1, N]$ for N the population size.
   - Define neighborhood size $\sigma < N$, maximum influence values $\phi_{1,max}, \phi_{2,max}$ and maximum velocity v_{max}
   - WHILE (not stopping criterion)
     FOR each x_i, $i \in [1, N]$
     - $H_i \leftarrow \sigma$ nearest neighbors of x_i
     - $h_i \leftarrow argmin_x(f(x) : x \in H_i)$
     - Generate a random vector $\phi_1(k) = (\phi_1(1), \ldots, \phi_1(D))$ where $\phi_1(k) \sim U[0, \phi_{1,max}]$
     - Generate a random vector $\phi_2(k) = (\phi_2(1), \ldots, \phi_2(D))$ where $\phi_2(k) \sim U[0, \phi_{2,max}]$
     - $v_i \leftarrow v_i + \phi_1 \circ (b_i - x_i) + \phi_2 \circ (h_i - x_i)$
     - if $v_i > v_{max}$ $v_i \leftarrow v_i v_{max}/v_i$
     - $x_i \leftarrow x_i + v_i$
     - $b_i \leftarrow argmin_x(f(x_i), f(b_i))$
       END FOR

     END WHILE

   ForPSO we have chosen the default parameters used in [7]. The population size was 40. The σ parameter is calculated by a percentage of informants for each particle and defaults to $1 - (1 - 1/40)^3$, v_{max} is not fixed which disables clamping of the velocity.

3. **Covariance Matrix Adaptation Evolution Strategy (CMA-ES)**. In its beginnings, ES (Evolution Strategy) [44] was a single search technique that randomly select one single individual (vector) and only with mutation explores the search spaces adding a random normal vector with mean zero and standard deviation σ to the initial vector and selecting the best of the two vectors for each iteration. Firstly, the added vector was performed with the same standard deviation σ for all vector component ($x_1 \leftarrow x_1 + r$ where the components of r are computed as $r_i \sim N(0, \sigma)$). Posteriorly, each component had its own σ with the logically improvement ($r_i \sim N(0, \sigma_i)$, $x_1 \leftarrow x_0 + N(0, \Sigma)r$ where Σ is a nxn diagonal matrix with diagonal $\sigma_i i \in [1, n]$). This scheme is notated by (1+1)-ES. An adaptive scheme was proposed in 1973 by Rechenberg who analyzes the success rates and perform the 1/5 rule for the adaptation of the deviation standard parameters through the iterations [44]. In the nineteens, the first generalization of the (1+1) ES was the $(\mu + 1)$-ES where μ (user-defined parameter) parents were used in each generation to generate a single new individual. Posteriorly, $(\mu + \lambda)$-ES and (μ, λ)-ES were performed [8] and a self-adaptive evolutionary strategy variations appear [6]. A success one of this variations was CMA-ES (Covariance Matrix Adaptation Evolution Strategy). The pseudo-code for a simple CMA-ES for dimension N is:

   - Initialize τ and τ_c in R
   - Initialize $C \leftarrow I = DxD$ identity matrix and μ the number of parents
   - Generate randomly $(x_k, \sigma_k) \in R^D xR$ where $k \in [1, \mu]$
   - WHILE (not stopping criterion)
     - $\sigma^{'} \leftarrow \sum_k \frac{\sigma_k}{\mu}$, $x^{'} \leftarrow \sum_k \frac{x_k}{\mu}$ where $\sigma^{'} \in R$ and $x^{'} \in R^D$
     - FOR (each k)
       - $r \leftarrow N(0, 1)$ random normal scalar. and $R \leftarrow N_n(0, 1)$ n dimensional normal random vector
       - $\sigma_k \leftarrow \sigma^{'} e^{r\tau}, s_k \leftarrow \sqrt{C}R, z_k \leftarrow \sigma_k z_k, x_k \leftarrow x^{'} + z_k$ END FOR
     - $S^{'} \leftarrow \sum_\lambda \frac{s_k s_k^T}{\lambda}$
     - $C \leftarrow (1 - \frac{1}{\tau_c})C + \frac{S^{'}}{\tau_c}$

   where τ and τ_c are time constant learning parameters that determine the adaptation speed of the σ_k values and covariance matrix C respectively. D is the search space dimension,

For CMA-ES we have chosen the default parameters used in [62].

2.1.1 Existing Stop Criteria for Single Objective Problems

There is a rich literature on online termination of EA. The simplest and most extended one [55, 43, 61] consists in reaching a number of iterations or function evaluations. This stopping criterion is not useful by itself since the number of iterations that guarantee a steady state significantly varies across problems [55]. This has motivated the

definition of alternative criteria based on either a measure derived from EA evolving population [24, 67, 11, 39, 31, 72, 43, 61, 64, 65] or the internal parameters of a particular EA algorithm [23, 20, 14, 1]. In the case of convergence to the minimum, some of the former criteria reduce the number of iterations do not improving results anymore.

However, there are no systematic conditions ensuring that such alternative criteria terminate EA at the steady state. Also, even in the case that EA has converged to the optimum, the relation between the alternative measures values and EA solution distance to the optimum (accuracy) remains unknown.

The definition of a general termination criterion addresses two issues. First, defining a measure, Q, reflecting the state of the optimization at EA current iteration. Second, setting the conditions that the values of the former measure taken across EA iterations, namely $(Q_k)_{k\in\mathbb{N}}$ for k the k-th iteration and $\mathbb{N}$ the natural numbers, should satisfy in order to terminate the algorithm.

Concerning the termination condition, two different types are considered. The first condition [24, 67] terminates EA if the measure reflecting the state of the optimization is below a given threshold $Q < \epsilon_g$. In order to avoid early termination, it is usual to require that Q is below the threshold for a (given) number of generation changes n_g, $Q_k < \epsilon_g$ for $k = k_0, \ldots, k_0 + n_g$.

The second condition terminates EA if the measure Q reaches a steady state. A steady state is an asymptotic concept that is achieved the moment that the variation of Q is under a given threshold. Several implementations for computing steady states are found in the literature. Some authors [11, 39, 31, 72] consider that Q has reached its steady state if it does not vary more than a fixed threshold between consecutive iterations for n_g generations changes: $\forall k \in \{k_0, k_0+1, .., k_0+n_g\} \quad |Q_k - Q_{k+1}| \leq \epsilon_g$. However, since it only bounds the difference between consecutive iterations, it does not guarantee that Q has reached the steady state. Recent works make use of descriptive statistics. In particular, a threshold over the standard deviation for all Q_k values is used in [51, 66]. This might require a large number of EA iterations for achieving the steady condition given that EA first stages do not contribute to the asymptotic behavior. Other approaches use more complex statistics, like regression analysis [21] or hypothesis tests [35], and, thus, add computational cost at each iteration.

Aside from measures derived for specific EA paradigms, the measures reported in the literature [31, 55] are based on either the objective function in f-space [11, 39] (improvement-based) or the distribution of the evolving population in x-space [72, 55] (distribution-based). In both cases measures are computed from a given percentage of the (best) evolving population. Improvement-based criteria may lead to early termination (possibly far from the optimum) due to the stochastic nature of EA [72]. Distribution-based quantities are better suited for terminating EA for some test functions. They terminate EA ensuring a given accuracy, as far as suitable parameters for the termination condition (threshold and number of generations) are chosen [72]. Given that these parameters are chosen according to the profile of the objective test function, the above result does not generalize to real-world (unknown) functions [72]. In this context, some recent works [39] adapt termination parameters to the current evaluation of the fitness function. However the simplest (and most extended [55, 43, 61, 64, 67]) way of ensuring that the algorithm terminates is by

setting a maximum number of iterations. Since the number of iterations ensuring convergence significantly varies across problems [55], this maximum must be large enough to ensure EA has reached steady state and is prone to lead to useless iterations [72, 65].

Noting by $(Ind_j(k))_{j=1}^n$ the set of the n best (ordered in decreasing fitness) individuals at the k-th iteration, the measure Q is defined at each iteration as:

- **Improvement-based (Fitval)** In this case Q is defined by the average of the objective function values for the best individuals:

$$Q_k = \frac{\sum_j f(Ind_j(k))}{n} \tag{2.1}$$

- **Distribution-based.** These quantities measure the sparseness of a given percentage of the (best) evolving population. Two measures are mainly considered:

  Maximum Distance (MxD). It is given by the maximum distance of the population to the best individual:

$$Q_k = max_{j>1} d(Ind_1(k) - Ind_j(k)) \tag{2.2}$$

  for $d(\cdot, \cdot)$ the Euclidean distance in the population parametric space.

  Population Variability (Std). It is the maximum of the standard deviations (computed using the population best individuals) of each dimension of the search space. If such dimension is N, so that $Ind_j(k) = (Ind_j^1(k), \ldots, Ind_j^N(k))$, then it is defined as:

$$Q_k = max_{i\in(1,\ldots,N)} \left(\frac{\sum_j (Ind_j^i(k) - \mu_i)^2}{n-1} \right) \tag{2.3}$$

  for $\mu_i = \frac{\sum_j Ind_j^i(k)}{n}$ the average i-th coordinate.

2.2 Inference Statistics

2.2.1 Regression Model

Given a sampling of two random variables (x and y), the linear regression of y (response variable) over x (explicative variable) is formulated as:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \tag{2.4}$$

for x_i a set of fixed explicative variable values (the actual sampling of x), y_i the response random variable and ε_i a random error satisfying:

2.2.1.1 Model Assumptions

1. Linearity: $E(\varepsilon_i) = 0$

2. Homoscedasticity: $VAR(\varepsilon_i) = \sigma^2, \forall i$
3. Uncorrelation: $COV(\varepsilon_i \varepsilon_j) = 0, , \forall i, j$
4. Gaussianity: $\varepsilon_i \sim N(0, \sigma^2)$, for $N(0, \sigma^2)$ a normal distribution.

The parameters of the regression model (2.4) are the regression coefficients $\beta = (\beta_0, \beta_1)$ and the error variance σ^2. The regression coefficients describe the way the two variables relate, while the variance indicates the accuracy of the model and, thus, measures to what extent x can predict y.

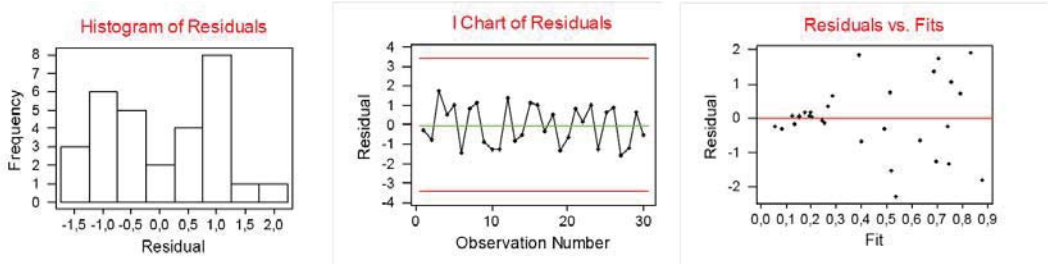


Figure 2.2: Residual diagnosis plots

For a sample of length N, the regression coefficients estimations, $\widehat{\beta} = (\widehat{\beta}_0, \widehat{\beta}_1))$, are estimated by Least Squares Estimation (LSE) as:

$$\widehat{\beta} = (X^T X)^{-1} X^T Y \tag{2.5}$$

for $X = \begin{pmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_N \end{pmatrix}$, $Y = (y_1, \ldots, y_N)$ and ^{T} denoting the transpose of a matrix. The differences between estimated responses, $\widehat{y}_i = \widehat{\beta}_0 + \widehat{\beta}_1 x_i$, and observed responses y_i:

$$e_i = y_i - \widehat{y}_i$$

are called residuals. Their square sum provides an estimation of the error variance:

$$S_R^2 = \widehat{\sigma}^2 = \frac{\sum e_i^2}{n-2} \tag{2.6}$$

So, the estimation of the error standard deviation (standard error) is expressed by:

$$S_R = \sqrt{\widehat{\sigma}^2 = \frac{\sum e_i^2}{n-2}} \tag{2.7}$$

The four model conditions endow desirable properties to the LSE of the regression coefficients [3]. By the Gauss-Markov theorem under the first three assumptions, the LSE are best linear unbiased estimators and assure that predictions made by

least squares fitted equations are good. By adding the fourth assumption (error gaussianity), the LSE has minimum variance among all unbiased estimators (not just linear) and allows the use of parametric tests, such as the Student's t-test for testing hypothesis on parameter values. The central limit theorem (asymptotically) guarantees this last property for large samples. Therefore, given that we have as much samples as EA iterations, in our case, the gaussianity is not a critical issue.

The standarized residuals:

$$en_i = (e_i - \mu(e_i))/std(e_i) \tag{2.8}$$

for μ the average and std the standard deviation, are used to verify the model assumptions. The plot of en_i over $\widehat{y}_i$ is called the versus fit plot and reflects linearity (in the measure that it is centered at zero) and homoscedasticity (uniform deviation from zero). The plot of en vs the sorted explicative variable is called the versus order plot and serves to detect any correlation pattern. Finally, the histogram of the standarized residuals reflects Gaussianity [41].

The estimation of the standard error of the regression slope estimator ($\widehat{\beta}_1$) can be expressed by:

$$S_{\widehat{\beta}_1} = \frac{S_R}{\sqrt{SS_x}}$$

where $SS_x = \sum (x_i - \bar{x})^2$

The problem of heterocedasticity is usually solved by taking logarithms in both variables [2]. In this case, our model formulation is as follows:

$$log(y_i) = \beta_0 + \beta_1 log(x_i) + \varepsilon_i \tag{2.9}$$

By taking exponentials, the regression model in the original scale is polynomial with multiplicative errors:

$$y = e^{\beta_0} x^{\beta_1} e^{\varepsilon} \tag{2.10}$$

Previous to any kind of inference, it is mandatory to verify that the estimated parameters make sense. That is, whether it really exists a linear relation between x and y. By the Gauss-Markov theorem, such linear relation can be statistically checked using the following T-test [41]

$$\begin{array}{ll} H_0: & \beta_1 = 0 \\ H_1: & \beta_1 \neq 0 \end{array} \tag{2.11}$$

where the statistic used is:

$$T = \frac{\widehat{\beta_1}}{S_{\widehat{\beta_1}}} \tag{2.12}$$

and a $p - value$ close to zero (below α) ensures the validity of the linear model with a confidence $(1 - \alpha)100\%$.

Another important issue is assessing the predictive value of the model. That is, to what extend the explicative variable can predict the values of the response variable.

This is quantified by the percentage of the response variability that is explained by the explicative variable using the R^2 rate:

$$R^2 = 1 - \frac{\sum(y_i - \widehat{y}_i)^2}{\sum(y_i - \bar{y})^2} = 1 - \frac{S_R^2(N-2)}{\widehat{\sigma}_y^2(N-1)} \in [0,1] \tag{2.13}$$

for $\widehat{\sigma}_y^2$ the sample variance of the response variable y. The higher R^2 is, the more reliable the prediction is.

2.2.2 Hypothesis Testing

An hypothesis in statistics is an assumption about a population parameter or a population parameters relation. Hypothesis testing begins with a formulation of a null hypothesis (H_0) that it is supposed hypothetically true and the alternative hypothesis (H_a) that is the opposite of the null hypothetical. In fact, the alternative hypothesis is the hypothesis that the researcher sets out to prove.
The general procedure continues with seeking a function U assuming that (H_0) is true that contains the objective parameter and a statistics with known distribution. Posteriorly, a sample is selected and the value of U with the data from the sample selected is calculated. Finally, the statistically-based decision is taken considering a significance level α, the probability to reject H_0 in the case that H_0 is true. Because the distribution of U is known, the decision is taken based on a critical zones defined by critical values determined with the α level. If the statistics calculated with the sample fall in the critical zone the null hypothesis H_0 is rejected and the alternative hypothesis can be hold with certain risk determined by α. In 2.3 three possible situations are drawn.

In the case of a proportion population parameter the formulation of the one-tailed test (upper bound) hypothesis is:

$$\begin{array}{ll} H_0: & q \leq q_0 \\ H_1: & q > q_0 \end{array} \tag{2.14}$$

The statistic for the sampling proportion follows a normal distribution $N(0,1)$ and is given by:

$$Z = (\widehat{q} - q_0)/\sqrt{q_0(1-q_0)/n_{EA}} \backsim N(0,1) \tag{2.15}$$

for $\widehat{q}$ the sampling proportion and n_{EA} the sample size. The null hypothesis H_0 is rejected if the statistic Z_Q has a p-value below α case in which Z falls in the reject zone because is greater than the critical value.

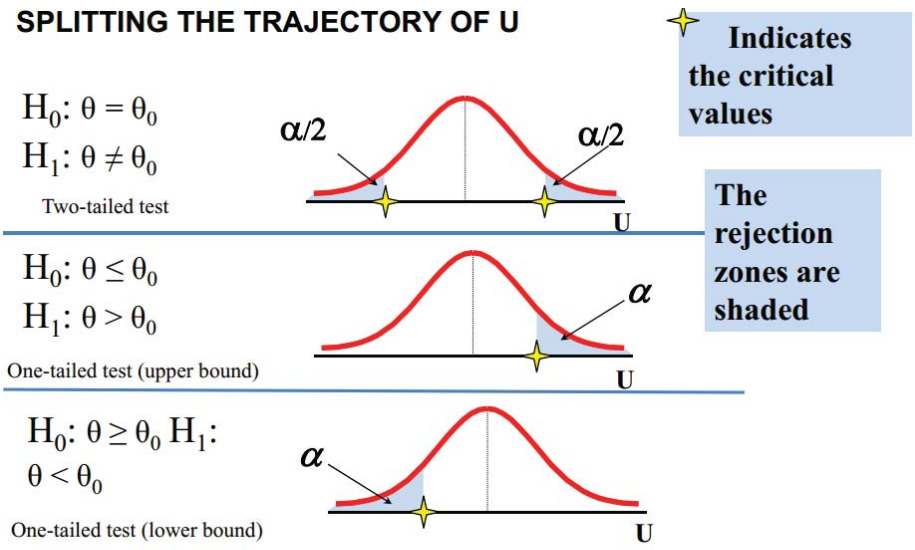


Figure 2.3: Splitting the trajectory of U

Chapter 3

Statistical Tools for Steady State Termination Conditions

Defining steady states is not straightforward for all EA paradigms and, in fact, they can not be properly defined if the mutation rate (like genetic algorithms [53]) is always positive across EA iterations. For these cases, EA can only be stopped if there is a theoretical analysis of its convergence to the optimum. However, whether EA has reached the optimum can only be answered for very specific algorithms and optimization problems [34, 69, 54, 6]. In general, EA optimality is difficult to assess and, for many paradigms, is impossible when no gradient information is available. Given that this could be the usual case for EA optimization problems, a stop criterion should ensure EA has reached a steady state [65].

First, in this Chapter, we provide a formulation of EA steady state inspired in the concept of Cauchy sequences. Second, we present a formula for determining the moment a quantity has converged to its steady state. The formula is given in terms of convergence rates and only depends on two parameters. Finally, we provide a statistical way to set the values for these parameters to ensure accuracy of the formula.

3.1 Formulation of EA Steady State in Cauchy terms

In a theoretical framework, analysis of EA convergence properties addresses the conditions that ensure reaching the set of optima individuals, regardless of the initial population. Given the stochastic mechanisms of EA paradigms, such general properties must be given in probabilistic terms [53, 6, 5]. In practice, each run of EA iterative scheme depends on a single given initial population and consists of a sequence of states of such population. In this context and assuming that the search space is R^N, exploring EA population steady state can be posed in terms of convergence of real number sequences.

We will consider that EA has reached a steady state if the values of EA best individuals (given by a percentage of EA population) jointly converge, as sequences of real numbers, to the same values that give the coordinates of a given point of the search x-space, $P \in \mathbb{R}^N$. This definition of steady state assumes that EA population diversity decreases across evolution and that we deal with a single objective optimization. This is the case for a large range of EA paradigms (like covariance matrix adaptation [4], differential evolution [60] or particle Swarm optimization [32] to mention just a few).

Let n_{best} be the number of EA best individuals and, for each individual $j = 1, \ldots, n_{best}$, note by $(Ind_j(k))_k \in \mathbb{R}^N$ their sequence of values across EA iterations, which we index by k. Then, we say that EA population has converged to its steady state if $\exists P \in \mathbb{R}^N$ such that:

$$Ind_j(k) \xrightarrow[k\to\infty]{} P \quad , \forall j = 1, \ldots, n_{best} \tag{3.1}$$

In real world applications, the point P is unknown, and, thus, the above definition is not feasible. In real vector spaces of finite dimension, convergence to unknown points is naturally handled by using the concept of Cauchy sequence. A sequence of real numbers, $(\tilde{q}_k)_{k\in\mathbb{N}}$, is Cauchy if the following condition holds:

$$\forall \epsilon, \quad \exists k_0 \quad , \text{such that } \forall k_1, k_2 \geq k_0. \quad \left\| \tilde{q}_{k_1} - \tilde{q}_{k_2} \right\| \leq \epsilon \tag{3.2}$$

for $\|.\|$ denoting a norm in $\mathbb{R}^N$.By general theory of real analysis, a sequence converges if and only if it is Cauchy [52]. This equivalence is broadly used in numerical analysis to terminate iterative algorithms having as parametric space the real numbers [12]. It is worth noticing that the above Cauchy condition measures the amount of change of the sequence of real numbers. This fits into the intuitive idea that an algorithm should terminate if it can not improve results anymore.

In the case of EA, given that all individuals should simultaneously converge to the same point, the condition (3.2) can be stated in Cauchy terms as follows:

$$\forall i, j \qquad ||Ind_i(k) - Ind_j(k)|| \xrightarrow[k\to\infty]{} 0 \tag{3.3}$$

for $i, j \in 1, \ldots, n_{best} and (Ind_j(k))_{j=1}^{n_{best}}$ the set of the n_{best} best individuals at the k-th iteration.

Let us see that the above condition relates to the accuracy of EA solution. The accuracy of EA solutions is given by the distance of the best individuals to the (unknown) function minimum, $Ref_j(k) := d(Ind_j(k), Opt)$, $j = 1, \ldots, n_{best}$. If EA reaches a steady state, we have that $Ref_j(k)$ simultaneously converge to the unknown (and possibly positive) distance $d(P, Opt)$:

$$\text{EA Reaches a Steady-State} \Leftrightarrow Ref_j(k) \xrightarrow[k\to\infty]{} d(P, Opt), \ \forall j = 1, \ldots, n_{best} \tag{3.4}$$

Consequently, its maximum value for the n_{best} individuals, $Ref_k := max_j(Ref_j(k))$, provides the following equivalent formulation:

$$\text{EA Reaches a Steady-State} \Leftrightarrow Ref_k \xrightarrow[k\to\infty]{} d(P, Opt) \tag{3.5}$$

In a similar manner, any quantity Q computed from EA evolving population is simultaneously Cauchy across the n_{best} individuals and, thus, its maximum value $\max_j(Q_j(k))$ is a sequence of real numbers that it is Cauchy and convergent to Q evaluated at P.

Posing EA steady state in terms of Cauchy sequences allows the definition of a numeric termination condition which parameters can be adjusted according to the accuracy of such numeric formula for detecting steady states.

3.2 Stopping condition based on EA steady states

In our approach, the natural termination condition is the Cauchy convergence of the sequence of real numbers, $(\tilde{q}_k)_k$, given by the values of the measure for all iterations. To detect the stabilization of a measure $\tilde{q}$ some authors [11, 39, 31, 72] consider that $\tilde{q}$ has reached its steady state if it does not vary more than a fixed threshold between consecutive iterations for n_st generations changes:

$$\forall k \in \{k_0, k_0 + 1, .., k_0 + n_st\} \quad |\tilde{q}_k - \tilde{q}_{k+1}| \leq \epsilon_g. \tag{3.6}$$

However, since it only bounds the difference between consecutive iterations, it does not guarantee that $\tilde{q}$ has actually reached the steady state. This could be the case, for instance, if $\tilde{q}$ had a linear profile across iterations of slope less or equal than ϵ_g, i.e., $\tilde{q}_k = \epsilon k$ for $\epsilon \leq \epsilon_g$. In such case $|\tilde{q}_k - \tilde{q}_{k+1}| = \epsilon \leq \epsilon_g \forall k$, but the sequence is by no means Cauchy since it is not convergent. $\tilde{q}_k \xrightarrow[k\to\infty]{} \infty$. Figure 3.1(a) illustrates the limitations of the criterion (3.6) for detecting the moment a quantity has reached a steady state in case of a constant decreasing behavior less than ϵ_g.

Our approach connecting with the idea of Cauchy sequences is implemented by a range formulae by checking that the variation of $(\tilde{q}_k)_{k\in\mathbb{N}}$ keeps below a threshold, ϵ_{st}, for a given number of n_{st} generation changes (variations):

$$\forall i, j \in \{k, k + 1, .., k + n_{st}\} \quad |\tilde{q}_i - \tilde{q}_j| \leq \epsilon_{st} \tag{3.7}$$

This can be formulated in terms of the range of $\tilde{q}$ in the generational interval $\{k, \ldots, k + n_{st}\}$ as:

$$Rng_{\{k,\ldots,k+n_{st}\}}(\tilde{q}) = Rng_{\tilde{q}} = \max_{i\in\{k,\ldots,k+n_{st}\}} \tilde{q}_i - \min_{i\in\{k,\ldots,k+n_{st}\}} \tilde{q}_i \leq \epsilon_{st} \tag{3.8}$$

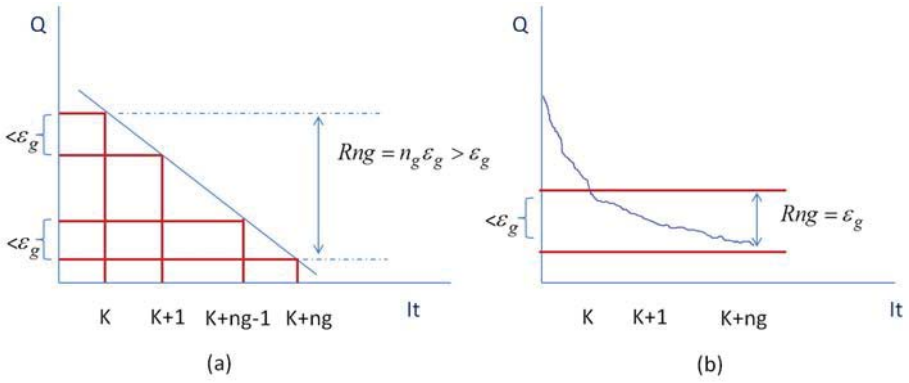


Figure 3.1: Possible numeric criteria to detect wether a quantity has reached a steady state: Stabilization formula and the proposed Range formula

The sequence is assumed to have reached its steady at the first generation satisfying (3.8):

$$K^{\tilde{q}}_{Ter} := \min_{k}(Rng_{\{k,\ldots,k+n_{st}\}}(\tilde{q}) < \epsilon_{st}) \tag{3.9}$$

Figure 3.1(a) illustrates the higher capability of the range criteria (3.8) for detecting convergent asymptotic behavior compared to the point criteria (3.6). In the case of a linear sequence, the range $Rng_{\tilde{q}} = n_{st}\epsilon$, and thus, (3.8) does not hold.

In case of EA termination $\tilde{q}$ is defined by either the gold-standard Ref or of the alternative measure Q defined in Chapter 2 taken across EA iterations that are indexed by k so that either $\tilde{q} = (Ref_k)_k$ or $\tilde{q} = (Q_k)_k$.

We would like to note that, in the next section, we will check in case that EA converges to the optimum, the relation between the parameter ϵ_{st} in formula (3.8) and the algorithm accuracy, provided that formula (3.8) is an accurate approximation to $\tilde{q}$ steady state.

The number of generations, n_{st}, and the maximum variation range across them, ϵ_{st} determine how good for detecting that $\tilde{q}$ has reached a steady state the range bound (3.8) is. The parameter ϵ_{st} represents the accuracy of the approximation, while n_{st} sets the temporal generational window we expect such accuracy to hold. We note that in stochastic process like EA, the number of generations such that $\tilde{q}$ has a low variation should be large enough to prevent premature convergence at EA early stages. We also note that $K^{\tilde{q}}_{Ter}$ varies due to EA stochastic nature when taken for different runs. In this context, the parameters chosen to compute (3.9) should also ensure that EA termination keeps stable across EA runs.

The link between formula (3.9) parameters and deviation for detecting true steady state can be used to set their values using the following statistical analysis [32]. Let

$K_{Ter}^{\tilde{q}}$ be the minimum iteration achieving the bound (3.8). By definition of a steady state, $\tilde{q}$ has reached it if the range inequality given by (3.8) is below ϵ_{st} for all k greater than $K_{Ter}^{\tilde{q}}$.

$$Rng_k(\tilde{q}) \leq \epsilon_{st} \forall k \geq K_{Ter}^{\tilde{q}} \tag{3.10}$$

For each n_{st} and ϵ_{st}, this condition can be expressed by the following function:

$$X_{\tilde{q}} := \begin{cases} 1 & \text{if } Rng(k)(\tilde{q}) \leq \epsilon_{st} \quad \forall k \geq K_{Ter}^{\tilde{q}} \\ 0 & \text{otherwise} \end{cases} \tag{3.11}$$

The function $X_{\tilde{q}}$ takes values in $\{0, 1\}$ and taken across n_{EA} independent EA runs is a discrete random variable that follows a Bernoulli distribution with parameter $q = P(X_{\tilde{q}} = 1)$. This probability q is estimated by the sample proportion given by:

$$\widehat{q} = \sum X_{\tilde{q}}/n_{EA} \tag{3.12}$$

for n_{EA} the number of EA runs considered.

In this context, $\tilde{q}$ has actually reached its steady state if the probability $P(X_{\tilde{q}} == 1) = q$ is close to 1. This can be statistically checked using the proportion test:

$$\begin{array}{ll} H_0: & q \leq q_0 \\ H_1: & q > q_0 \end{array} \tag{3.13}$$

for q_0 representing the minimum proportion of times that an EA run will be stabilized using the range formula (3.9). This q_0 provides a lower bound for q with a given confidence level $(1 - \alpha)$. The statistic for the sampling proportion $\widehat{q}$ given by (3.12) follows a normal distribution $N(0, 1)$ and is given by:

$$Z_{\tilde{q}} = (\widehat{q} - q_0)/\sqrt{q_0(1 - q_0)/n_{EA}} \backsim N(0, 1) \tag{3.14}$$

The null hypothesis H_0 is rejected if the statistic $Z_{\tilde{q}}$ has a p-value below α which implies that the number of EA runs achieves stable behavior given by condition (3.10) is above $(q_0)100\%$ of the cases. Therefore, for each ϵ_{st} we consider that the number of generations n_{st} ensuring that $\tilde{q}$ has reached a steady state with a confidence α is given by the minimum integer such that H_0 is rejected. On the other side, (3.13) can be used to explore weather a given quantity is suitable for terminating a particular EA paradigm. We consider that Q can effectively terminate EA if there are a set of combinations (ϵ_{st},n_{st}) that can determine Q steady state for different function types regardless of the search space dimension. This can be checked by considering a sampling of the parameter space (ϵ_{st},n_{st}) and applying the proportion test for a representative set of functions covering the landscapes most influential on EA convergence ([24, 25]). Scalability across search space dimensionality is checked by considering the same test functions for increasing dimension. Those pairs (ϵ_{st},n_{st})with the most stable rejection rate across function landscape and dimension are the best suited for

terminating EA using Q. In case that no pair (ϵ_{st},n_{st}) achieves a stable rejection across dimensions for any function type, the quantity Q is not a good candidate for EA termination.

3.3 Experimental Set-up

We have applied our framework to the benchmark used in [24] to test 31 state-of-the-art evolutionary algorithms. Details about function definition might be found in [25]. The functions are clustered according to their overall properties in five groups [25]: separable, low (good) conditioning, high (bad) conditioning, multi-modal with strong global structure and with weak global structure. Besides, the functions cover the main properties (multimodality, global structure and scalability) reported in a recent study [37, 9] to have a high influence in the performance of EAs. For this experiment we have chosen 5 functions (one from each clustered group mentioned above) representative of the benchmark used in [24]: Sphere, Rosenbrock, Ellipsoidal, Rastrigin and Schwefel functions. Figure 3.2 shows these functions for dimension 2.

The goal of these experiments is to asses our numerical termination condition in two aspects:

1. **Termination of EA paradigms relying on diversity.** First of all, we have validated the presented framework for terminating EA paradigms relying on diversity. The three EA strategies representative of diversity based paradigms described in Chapter 2 have been analyzed in order to check if they admit a termination using one of the measures of population diversity given in chapter 2. We recall that such EA paradigms based on diversity chosen for this experiment are: Differential Evolution (DE) [60, 13], Particle Swarm Optimization (PSO) [29, 32] and Covariance Matrix Adapting Evolutionary Strategy (CMA-ES) [4, 40].

   The 3 EA paradigms are DE, PSO and CMA-ES and were ran using the parameters and computational algorithms described in Chapter 2

   For all paradigms population size has been set to 20*D dimension of the search space) and EA optimization was computed for 3000 iterations. Dependency of termination parameters with respect the search space dimension has been explored by considering the definition of the functions for dimensions 2, 4 and 10. The quantity measuring population diversity is the distribution of EA population in x-space given by the maximum Euclidean distance, MxD given by:

   $$MxD = max_{j\in(1,\ldots,n_{best})} d(Ind_j, Ind_1) \tag{3.15}$$

   for $(Ind_j)_{j=1}^{n_{best}}$ the set of the n_{best} best individuals and Opt the function optimum in x-space. This quantity is set to be $\tilde{q}$ for the computation of (3.10) and the proportion test. For each paradigm and test function, the range formula (3.10) has been computed for $\epsilon = 10^{-1}$ and $n_{st} = 10, 50, 100, 200, 500$. For each function and dimension the function $X_{\tilde{q}}$ in (3.11) computed for $n_{EA} = 30$

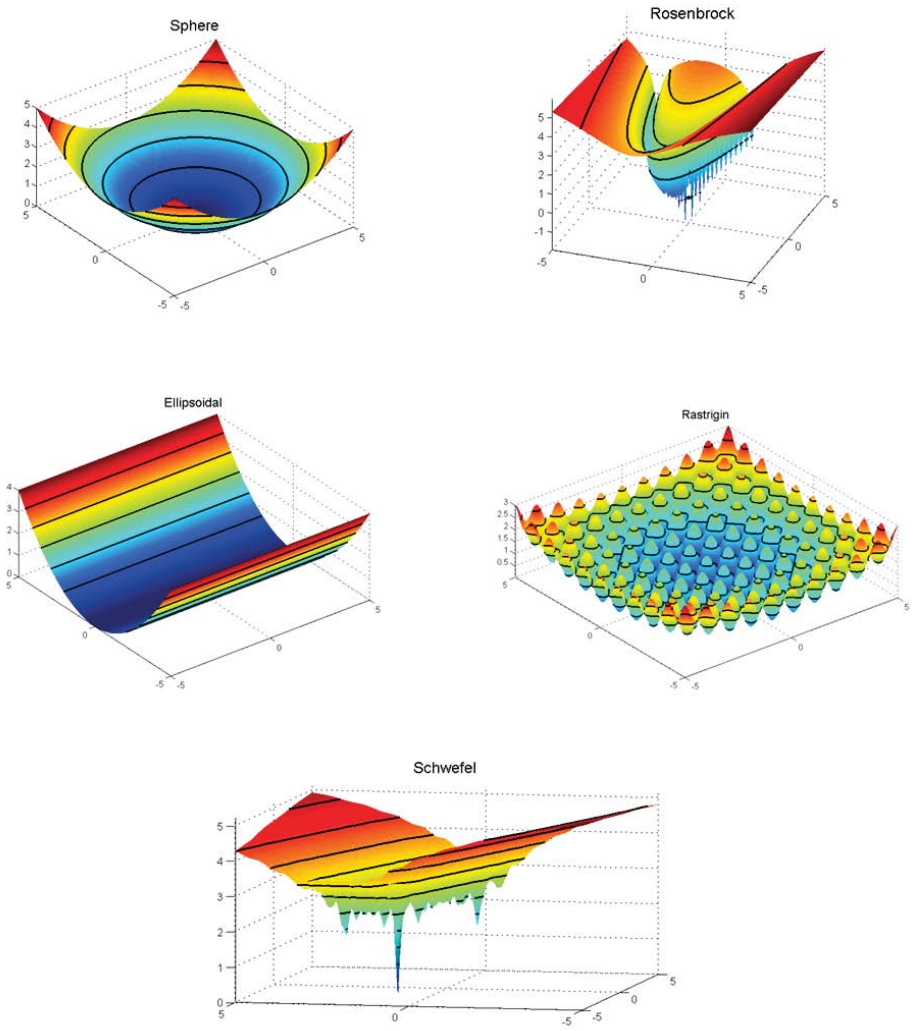


Figure 3.2: Representative set of five functions covering different landscapes influencing EA convergence

independent EA runs defines the sample for the random variable used in the proportion test. The proportion test has been computed for $q_0 = 0.95$ and at confidence $\alpha = 0.05$. The EA paradigm presenting a most suitable rejection profile across increasing dimension is the one selected for the subsequent experiments.

2. **Optimal Parameters for the Stopping Condition.** Once the EA paradigm

most suitable to be terminated using diversity has been determined, the optimal parameters for the range formula should be set. For a given accuracy in steady states, ϵ_{st}, in this experiment we will determine the minimum number of generations n_{st} in the range formula required to actually have a good approximation to the steady state. Such number of generations should be the same for all quantities used to terminate EA. The quantities stabilized using the range formula and the proportion test are the gold-standard $\tilde{q} = Ref$ and the two alternative quantity MxD (2.2)and $Fitval$ value (2.1).

For this experiment, we have used the proportion test to determine the minimum number of generations in the range formula (3.9) required to actually have a good approximation to Q steady state. We have applied the proportion test with $\alpha = 0.05$ and $q_0 = .95$ for $n_{st} \in \{10, 50, 100, 200, 500\}$ and $\epsilon_{st} \in \{10^{-1}, 10^{-2}\}$. For each paradigm and x-space dimension, the parameters of formula (3.8) best suited for its termination are the minimum number of generations that reject the proportion test (3.13).

The EA paradigm selected in the first experiments with the range formula parameters determined in the second one have been used in the remaining experiments presented in the next Chapter.

3.4 Results

3.4.1 Termination of EA paradigms relying on diversity

Tables 3.1, 3.2, 3.3 report the results of the proportion test for DE, PSO and CMA, respectively. Each table reports results across the selected number of generations (columns) and x-space dimension (rows). We report the hypothesis test result (1 for null hypothesis rejection, 0 for not rejection), the test p-value and the sampling proportion.

For the 2D case a minimum number of generations could be achieved for all paradigms. For dimension 4, our range formula fails to detect PSO steady state and for dimension 10 the proportion of detected steady states is below 0.9 for all n_{st} and paradigms. An analysis of the profiles of population diversity across iterations shows that failures arise from a number of iterations insufficient for reaching steady states of some particular functions. These cases are Rosenbrock, Schwefel for DE, Schwefel for PSO and Rosenbrock for CMA. These functions have a landscape showing a poor EA convergence rate at least for the chosen parameters for each paradigm and would require more iterations. Figure 3.3 shows the profiles of MxD for one of the failing cases (solid line) in comparison to the always convergent Sphere (dot line). We show an example for each of the EA paradigms, Schwefel for DE and PSO and Rosenbrock for CMA.

In order to faithfully asses the capability of MxD for detecting EA steady state, functions not reaching EA steady state should be excluded. We have repeated the

Table 3.1: Optimal Number of Termination Generations for DE

	$p-value$			Sampling Proportion q			Null Hypothesis Rejection		
	dim2	dim 4	dim 10	dim2	dim 4	dim 10	dim 2	dim 4	dim 10
10	1.0	1.0	1.0	0.55	0.23	0	0	0	0
50	2.2e-05	0.9	1.0	1	0.77	0.32	1	0	0
100	2.2e-05	0.13	1.0	1	0.89	0.50	1	0	0
200	2.2e-05	2.2e-04	0.8	1	0.97	0.53	1	1	0
500	2.2e-05	2.2e-03	1.0	1	0.94	0.33	1	1	0

Table 3.2: Optimal Number of Termination Generations for PSO

	$p-value$			Sampling Proportion q			Null Hypothesis Rejection		
	dim2	dim 4	dim 10	dim2	dim 4	dim 10	dim 2	dim 4	dim 10
10	1.0	1.0	0.3	0.75	0.59	0.52	0	0	0
50	2.2e-04	0.9	0.2	0.97	0.77	0.87	1	0	0
100	2.2e-05	0.9	0.2	1	0.77	0.88	1	0	0
200	2.2e-05	0.9	0.8	1	0.77	0.88	1	0	0
500	2.2e-05	1	0.7	1	0.74	0.83	1	0	0

Table 3.3: Optimal Number of Termination Generations for CMA-ES

	$p-value$			Sampling Proportion q			Null Hypothesis Rejection		
	dim2	dim 4	dim 10	dim2	dim 4	dim 10	dim 2	dim 4	dim 10
10	0.01	0.8	1	0.95	0.82	0.34	1	0	0
50	2.2e-05	2.2e-05	1.0	1	1	0.65	1	1	0
100	2.2e-05	2.2e-05	1.0	1	1	0.70	1	1	0
200	2.2e-05	2.2e-05	0.5	1	1	0.86	1	1	0
500	2.2e-05	2.2e-05	0.2	1	1	0.88	1	1	0

experiments without the failing cases and in order to explore the impact of the accuracy ϵ_{st}, we have computed the range formula for $\epsilon_{st} = 10^{-1}$, and $\epsilon_{st} = 10^{-2}$. Results for the proportion test as reported as before in Tables 3.4,3.5 and 3.6. For low dimensions (up to 4) the number of generations required for termination remains unchanged across ϵ_{st} for the 3 methods considered. For higher dimensionality, PSO is the only method that has the same behavior regardless of ϵ_{st}, while DE and CMA present a variability in the number of generations that should be further investigated .It is worth noticing that PSO termination is also stable across dimensions and, for low dimensionality, DE as well. This is not the case for CMA, for which termination generation seems to increase across dimensionality.

There are several interesting conclusions that can be derived from our experiments.

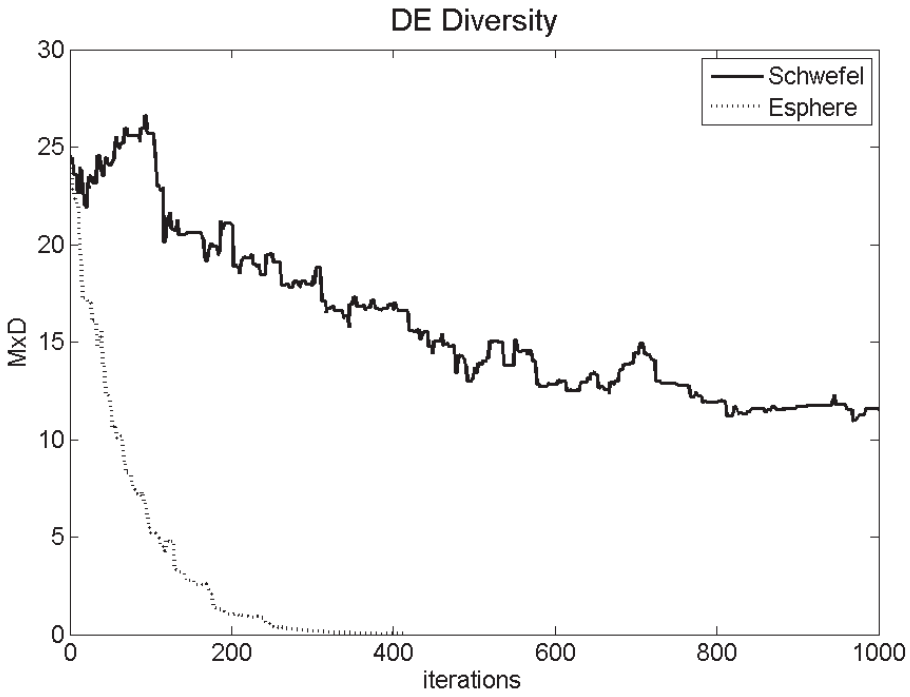


Figure 3.3: Loss of population diversity across iterations for DE

Table 3.4: Impact of ϵ_{st} in DE Termination

	$p-value$			Sampling Proportion q			Null Hypothesis Rejection		
$\epsilon_{st}=10^{-1}$	**dim2**	**dim 4**	**dim 10**	**dim2**	**dim 4**	**dim 10**	**dim 2**	**dim 4**	**dim 10**
10	0.98	1.0	1.0	0.83	0.47	0	0	0	0
50	7.8e-04	0.0025	1.0	1	0.99	0.63	1	1	0
100	7.8e-04	0.0025	0.36	1	0.99	0.91	1	1	0
200	7.8e-04	0.0025	0.039	1	0.99	0.96	1	1	1
500	7.8e-04	7.8e-04	1	1	1	0.66	1	1	0
$\epsilon_{st}=10^{-2}$	**dim2**	**dim 4**	**dim 10**	**dim2**	**dim 4**	**dim 10**	**dim 2**	**dim 4**	**dim 10**
10	0.14	0.52	1.0	0.93	0.61	0.0	0	0	0
50	7.8e-04	0.007	0.5	1	0.98	0.59	1	1	0
100	7.8e-04	0.0025	0.75	1	0.99	0.82	1	1	0
200	7.8e-04	0.0025	0.07	1	0.99	0.94	1	1	0
500	7.8e-04	7.8e-04	0.25	1	1	0.33	1	1	0

For the PSO paradigm, convergence to steady states is apparently unchanged across the dimension of the search space. For low dimensions, this also holds for DE. In such

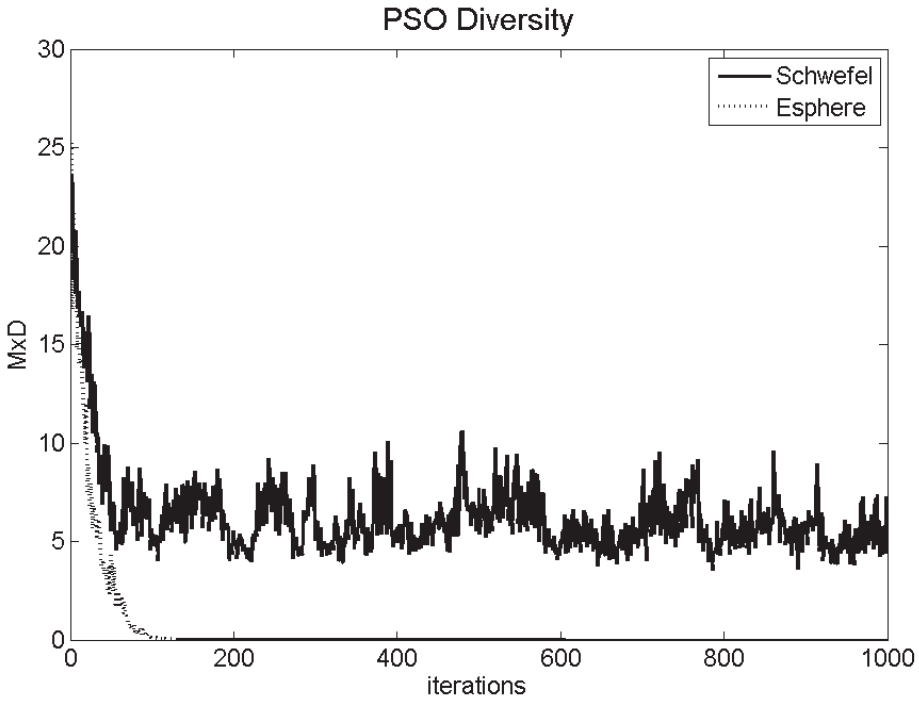


Figure 3.4: Loss of population diversity across iterations for PSO

Table 3.5: Impact of ϵ_{st} in PSO Termination

	$p-value$			Sampling Proportion q			Null Hypothesis Rejection		
$\epsilon_{st} = 10^{-1}$	**dim2**	**dim 4**	**dim 10**	**dim2**	**dim 4**	**dim 10**	**dim 2**	**dim 4**	**dim 10**
10	0.63	0.99	0.54	0.89	0.80	0.63	0	0	0
50	0.0025	0.0025	0.0025	0.99	0.99	0.99	1	1	1
100	7.8e-04	0.0025	7.8e-04	1	0.99	1	1	1	1
200	7.8e-04	0.0025	7.8e-04	1	0.99	1	1	1	1
500	7.8e-04	0.039	7.8e-04	1	0.96	1	1	1	1
$\epsilon_{st} = 10^{-2}$	**dim2**	**dim 4**	**dim 10**	**dim2**	**dim 4**	**dim 10**	**dim 2**	**dim 4**	**dim 10**
10	0.92	1.0	0.57	0.85	0.77	0.66	0	0	0
50	0.0025	0.0025	0.0025	0.99	0.99	0.99	1	1	1
100	7.8e-04	0.0025	7.8e-04	1	0.99	1	1	1	1
200	7.8e-04	0.0025	7.8e-04	1	0.99	1	1	1	1
500	7.8e-04	0.039	7.8e-04	1	0.97	1	1	1	1

cases, the number of generations required to terminate EA can be kept relatively low

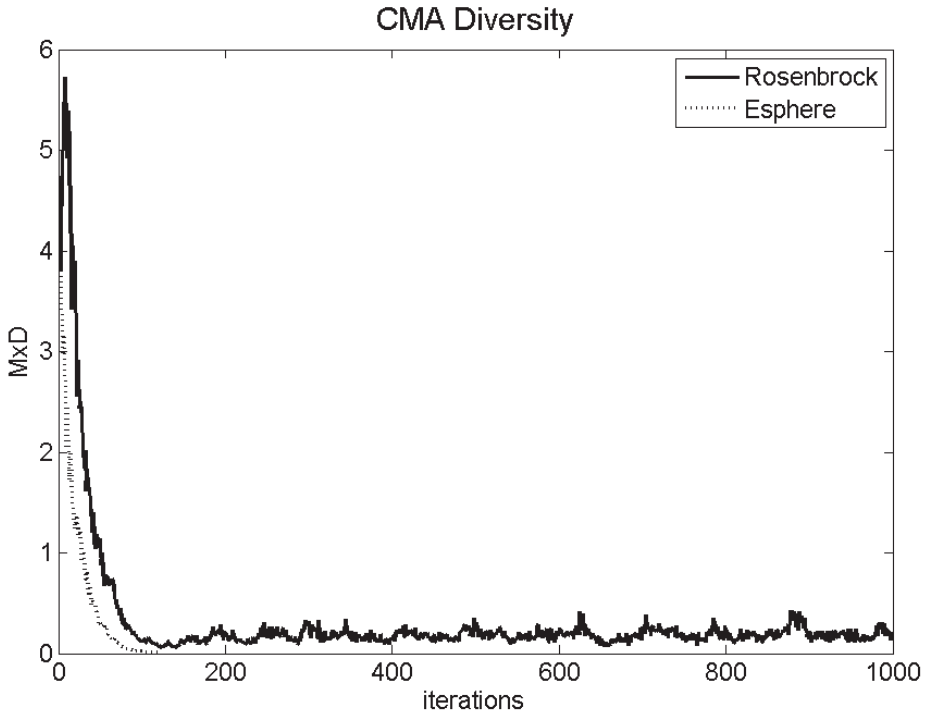


Figure 3.5: Loss of population diversity across iterations for CMA

Table 3.6: Impact of ϵ_{st} in CMA-ES Termination

	$p-value$			Sampling Proportion q			Null Hypothesis Rejection		
$\epsilon_{st} = 10^{-1}$	**dim2**	**dim 4**	**dim 10**	**dim2**	**dim 4**	**dim 10**	**dim 2**	**dim 4**	**dim 10**
10	0.0075	0.11	1.0	0.96	0.93	0.5	1	0	0
50	1.3e-04	1.3e-04	0.98	1	1	0.84	1	1	0
100	1.3e-04	1.3e-04	0.27	1	1	0.92	1	1	0
200	1.3e-04	1.3e-04	4.1e-04	1	1	0.99	1	1	1
500	1.3e-04	1.3e-04	1.3e-04	1	1	1	1	1	1
$\epsilon_{st} = 10^{-2}$	**dim2**	**dim 4**	**dim 10**	**dim2**	**dim 4**	**dim 10**	**dim 2**	**dim 4**	**dim 10**
10	4.1e-04	0.38	0.45	0.99	0.90	0.53	1	0	0
50	1.3e-04	1.3e-04	0.27	1	1	0.91	1	1	0
100	1.3e-04	1.3e-04	1.3e-04	1	1	0.99	1	1	1
200	1.3e-04	1.3e-04	1.3e-04	1	1	1	1	1	1
500	1.3e-04	1.3e-04	1.3e-04	1	1	1	1	1	1

and, thus, termination based on population diversity is computationally efficient. For the CMA paradigm, loss of diversity drops as the dimension increases and, thus, it requires a higher number of generations for ensuring termination at the steady state.

The dependency of parameters across paradigms illustrates that the choice of the quantity used for EA termination is linked to EA internal mechanisms and, thus, its selection is at the very core of the methods used as termination criteria. In this context, loss of population diversity seems to be a good candidate for PSO and DE, but might nor be the most appropriate one for CMA. Although, in case of DE diversity seems more suitable for low dimensions, by its higher simplicity and easier parallelization [71], we have chosen DE for the remaining experiments.

3.4.2 Optimal Parameters for the Stopping Condition

Tables 3.7, 3.8, 3.9 report the results of the proportion test given in Section 3.2 for setting the termination parameters (number of generations n_{st} and threshold ϵ_{st}) in the range formula (3.8). Each table reports results across the generations sampling $n_{st} \in \{$ 10,50,100,200,500,750, 1000 $\}$ for a given $\epsilon_{st} \in \{10^{-3},10^{-6},10^{-12}\}$. We report the null hypothesis rejection state (1 for rejection) as well as the test p-value. We recall that rejecting the test implies that the configuration (ϵ_{st}, n_{st}) provides an accurate approximation to each quantity steady state and, thus, it is a good parameter setting for EA termination.

Table 3.7: Optimal Number of Termination Generations for $\epsilon_{st} = 10^{-3}$

	Null Hypothesis Rejection							$p-value$						
	10	50	100	200	500	750	1000	10	50	100	200	500	750	1000
MxD	0	0	0	0	1	1	1	1.0	1.0	1.0	0.9	1e-9	5e-15	5e-22
Ref	0	0	0	0	1	1	1	1.0	1.0	1.0	0.8	2e-10	1e-18	1e-22
Fitval	0	0	0	0	1	1	1	1.0	1.0	1.0	0.8	7e-13	2e-21	3e-22

Table 3.8: Optimal Number of Termination Generations for $\epsilon_{st} = 10^{-6}$

	Null Hypothesis Rejection							$p-value$						
	10	50	100	200	500	750	1000	10	50	100	200	500	750	1000
MxD	0	0	0	0	1	1	1	1.0	1.0	1.0	1.0	1e-3	5e-8	1e-10
Ref	0	0	0	0	1	1	1	1.0	1.0	1.0	1.0	3e-3	6e-12	7e-15
Fitval	0	0	0	0	1	1	1	1.0	1.0	1.0	0.9	4e-10	5e-19	2e-20

As expected, the minimum number of generations required to provide a good approximation to Q steady state increases as ϵ_{st} decreases. Also, the behavior varies across quantities, with MxD requiring, for a given ϵ_{st}, a higher number of iterations

Table 3.9: Optimal Number of Termination Generations for $\epsilon_{st} = 10^{-12}$

	Null Hypothesis Rejection							$p - value$						
	10	50	100	200	500	750	1000	10	50	100	200	500	750	1000
MxD	0	0	0	0	0	0	0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
Ref	0	0	0	0	0	1	1	1.0	1.0	1.0	1.0	1.0	1e-3	5e-5
Fitval	0	0	0	0	1	1	1	1.0	1.0	1.0	0.9	2e-10	1e-18	1e-20

in order to accurately detect its steady state. The set of parameters that statistically provide a good approximation to steady states for all quantities and test functions are $(\epsilon_{st}, n_{st}) \in \{$ $(10^{-3}, 500)$, $(10^{-3}, 750)$, $(10^{-3}, 1000)$, $(10^{-6}, 500)$, $(10^{-6}, 750)$, $(10^{-6}, 1000)$ $\}$. Given that it is desirable a ϵ_{st} as small as possible with a minimum number of iterations, the parameter setting for formula (3.8) that will be used in the following experiments is $\epsilon_{st} = 10^{-6}$, $n_{st} = 500$.

The parameters that will be used in the following experiments to be presented in Chapter 4 are $\epsilon_{st} = 10^{-6}$, $n_{st} = 500$ applied to MxD, Ref, Fitval and DE paradigm.

Chapter 4

Evaluation Mechanism based on Regression Analysis

A main challenge in Evolutionary Algorithms (EAs) is determining a termination condition ensuring stabilization close to the optimum in real-world applications. A termination condition is suitable if it compares to the accuracy of the EA solution. The distance to the (known) function minimum in x-space (which we will note by Ref) is the gold-standard reference convergence criterion, given that is directly associated to the algorithm accuracy. This criterion can only be computed if the optimum of the test function is known and, thus, is useless in real-world problems. It follows that real-world problems require the definition of alternative measures derived from EA current state. In order to avoid iterations that do not improve results any more, such alternative measure should be related to the steady state of EA evolving population.

A measure Q derived from EA evolving population is a good candidate for terminating EA if it behaves like the gold-standard Ref given by EA accuracy. In this context, we propose posing the termination problem in statistical inference terms. From the perspective of statistical inference, the termination problem consists in designing a quantity (depending only on the EA output) that correlates to the accuracy of the solution, so that they can be swapped. In this Chapter, we present several statistical conditions using linear regression models to evaluate to what extent an alternative quantity can substitute the gold standard Ref for EA termination in Cauchy terms established in Chapter 3.

First, we introduce three conditions (strong,weak and empirical) to evaluate whether Ref can be swap by the alternative quantity. The first equivalence condition evaluates the deviation from identity regression line relating Ref and Q , since for the identity case both quantities have exactly the same convergence rate. Based on the idea of Cauchy convergence stated in Chapter 3, the strong condition in the regression coefficients is relaxed to an inequality that bounds Ref Cauchy convergence by Q Cauchy convergence. Finally, we also present an empirical substitution condition in terms of the range formula presented in the previous Chapter 3. A second contribution of

this Chapter is the derivation of a prediction formula to actually compute Q values ensuring a given EA accuracy in case of convergence to the optimum.

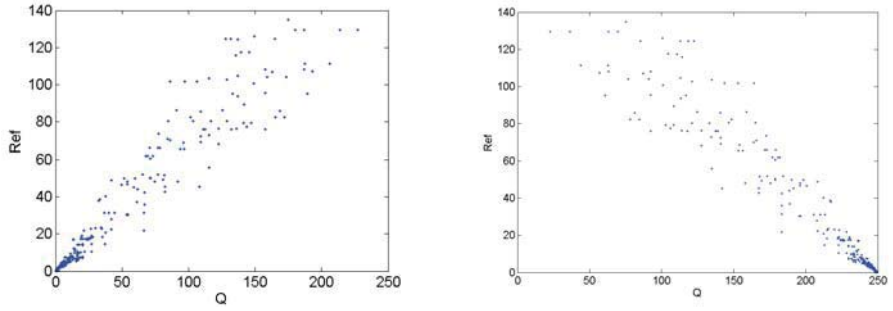
Finally, we use the inference model to compare several types of distribution-based quantities reported in the literature [72]. Our experiments indicate that the maximum distance to the best individual is the best choice in terms of computational efficiency and capability of predicting EA accuracy for the DE paradigm.

4.1 Substitution Conditions

Our final goal is to control (predict) the values taken by Ref from the values taken by the alternative measure Q. In inference statistics, this can be achieved by relating both quantities using a regression model. We remit the reader to the Chapter 2 for details on inference regression models or to [3]. Given that, in our case, the inference is over Ref, our model is:

$$Ref = \beta_0 + \beta_1 Q + \varepsilon \tag{4.1}$$

It is worth noticing that as EA approaches its steady state, population sparseness decreases, while it increases for EA first iterations. This introduces heteroscedasticity (i.e. a non-uniform variance of random errors) and also affects the Gaussianity assumption [49]. The monotonic behavior of the standard deviation of the random error in the regression model, depends on the relation between Ref and Q. If $Q \to 0$ as EA approaches its steady state, since early iterations will also have a large Q, there is an increase of σ as a function of Q (as figure 4.1(a) illustrates). Meanwhile, if $Q \to \infty$ as EA approaches its steady state, then there is an decrease of σ as a function of Q (as figure 4.1(b) also illustrates).



(a) (b)

Figure 4.1: Types of heterocedasticity in regression models due to a decrease in population sparseness as EA approaches its steady state: increasing σ, (a), and decreasing σ, (b).

Heteroscedasticity is mostly associated to a multiplicative errors and, thus, is usually solved [2] by taking natural logarithms in both variables, so that the regression model is given by:

$$log(Ref) = \beta_0 + \beta_1 log(Q) + \varepsilon \tag{4.2}$$

For $\epsilon \approx N(0, \sigma)$, we note that, by taking exponentials, the regression model in the original scale is polynomial with multiplicative errors:

$$Ref = e^{\beta_0} Q^{\beta_1} e^{\varepsilon} \tag{4.3}$$

Since the regression error ε follows a centered normal distribution, the average model in logarithmic scale is $log(Ref) = \beta_0 + \beta_1 log(Q)$. This provides the following intuitive interpretation [49] of the regression coefficients by taking exponentials. In average, if $\beta_1 \approx 1$ and $\beta_0 < 0$, then Q might ensure an upper bound for Ref:

$$Ref = e^{\beta_0} Q^{\beta_1} \approx e^{\beta_0} Q < Q \tag{4.4}$$

We note that if the slope $\beta_1 \approx 1$, then both quantities stabilize at the same time and, thus,they are equivalent under the stabilization termination condition. This requirement can be statistically checked using the following unilateral T-tests:

$$\begin{array}{lll} TP_1: & H_0: \beta_1 - 1 \geq \epsilon_1 & , \quad H_1: \beta_1 - 1 < \epsilon_1 \\ TP_2: & H_0: \beta_1 - 1 \leq -\epsilon_2 & , \quad H_1: \beta_1 - 1 > -\epsilon_2 \end{array} \tag{4.5}$$

Posing the hypothesis over the slope parameter, we get our **Strong Equivalence Condition**:

Definition 1 (Strong Equivalence Condition) *A measure Q is equivalent to Ref for EA termination if there exists a small number $\epsilon^* \sim 0$ that simultaneous rejects the following two hypothesis test:*

$$\begin{array}{lll} TP_1: & H_0: \beta_1 \geq 1 + \epsilon^* & , \quad H_1: \beta_1 < 1 + \epsilon^* \\ TP_2: & H_0: \beta_1 \leq 1 - \epsilon^* & , \quad H_1: \beta_1 > 1 - \epsilon^* \end{array} \tag{4.6}$$

for β_1 the slope coefficient of the regression model in logarithmic scale given by (4.2).

The hypothesis tests are rejected if their p values are close enough to zero. In case they are bellow a given α, we have that $|\beta_1 - 1| \leq \epsilon^*$ with a confidence $(1 - \alpha)100\%$.

The minimum ϵ, namely ϵ^*, ensuring rejection of, both, TP_1 indicates the accuracy of the assumption that $\beta_1 \sim 1$ and, thus, measures to what extend Q and Ref are equivalent for EA stabilization. In particular, the higher ϵ^* we have, the least stabilization equivalence. The minimum ϵ^* ensuring rejection of the pair of hypothesis TP_1 and TP_2 with a given confidence (1-α) can be computed from the one side statistic for TP_1 and TP_2 as follows:

Proposition 1 *Let $\widehat{\beta_1}$ the estimated slope of the regression model (4.2) and $\widehat{s_{\beta_1}}$ its standard error. Then, the minimum number ϵ^* that ensures rejection of the double hypothesis test is given by:*

$$\epsilon^* = max((1 - \widehat{\beta_1}) + t_\alpha^{N-2} s_{\widehat{\beta_1}}, (\widehat{\beta_1} - 1) - t_\alpha^{N-2} s_{\widehat{\beta_1}}) \tag{4.7}$$

For t_α^{N-2} the value of a T-student distribution with N-2 (N being the sample size) degrees of freedom having a cumulative probability equal to α.

Proof.

Assuming that H_0 is true for TP_1, we have that the statistic:

$$t^* = \frac{\widehat{\beta_1} - (1 + \epsilon)}{s_{\widehat{\beta_1}}} \tag{4.8}$$

is a random variable that follows a T-student distribution with $N - 2$ degrees of freedom, $\mathcal{T}^{N-2}$, being N the sample size. Therefore, given a risk α, the null hypothesis has to be rejected if the probability that $\mathcal{T}^{N-2}$ is under t^* is less than α. In terms of the distribution function of $\mathcal{T}^{N-2}$, namely $F(t^*)$, this is formulated as:

$$F(t^*) = P(\mathcal{T}^{N-2} < t^*) = P\left(\mathcal{T}^{N-2} < \frac{\widehat{\beta_1} - (1 + \epsilon)}{s_{\widehat{\beta_1}}}\right) \leq \alpha$$

By taking the inverse of the distribution function, we obtain the optimal value for t^* that leaves α cumulative probability as:

$$t^* = F^{-1}(\alpha) = t_\alpha^{N-2}$$

And, by substituting the value of t^* given by (4.8), we have that

$$\frac{\widehat{\beta_1} - (1 + \epsilon)}{s_{\widehat{\beta_1}}} = t_\alpha^{N-2}$$

Therefore, the minimum ϵ rejecting TP_1 is given by:

$$\epsilon = \epsilon_1 = (\widehat{\beta_1} - 1) - t_\alpha^{N-2} s_{\widehat{\beta_1}} \tag{4.9}$$

Proceeding in a similar way for TP_2, but considering left tile probabilities we have that:

$$F(t^*) = P(t^{N-2} > t^*) = P\left(t_{n-2} > \frac{\widehat{\beta_1} - (1 + \epsilon)}{s_{\widehat{\beta_1}}}\right) \leq \alpha$$

Thus, the minimum ϵ rejecting TP_1 is given by:

$$\epsilon = \epsilon_2 = (1 - \widehat{\beta_1}) + t_\alpha^{N-2} s_{\widehat{\beta_1}} \tag{4.10}$$

The maximum value between (4.9) and (4.10) defines the minimum common number, ϵ^* ensuring rejection of both hypothesis tests.

□

The hypothesis tests given by (4.6) are the most suitable theoretical way to check that a given quantity Q can substitute the reference one, Ref. However, they require simultaneous interpretation of the two hypothesis p-values and, possibly, their confidence intervals. This might be an inconvenient from a practical point of view. An alternative for assessing to what extend a quantity behaves like Ref for EA termination, could be to derive a swapability condition from the Cauchy steady state convergence introduced in the Chapter 3.

We recall that EA has reached a steady state if its n_{best} individuals simultaneously converge to a given point P

$$Ind_j(k) \xrightarrow[k\to\infty]{} P \quad , \forall j = 1, \ldots, n_{best} \tag{4.11}$$

Consequently their distance to the optimum

$$Ref_k := max(d(Ind_j(k), Opt), j = 1, \ldots, n_{best}) \tag{4.12}$$

is a sequence of reals numbers that also converges to $d(P, Opt)$. In an ideal setting of EA convergence to the optimum $Ref_k \to 0$ and Q would be suitable for substituting Ref if Q_K also converged to zero and $Ref_K \leq Q_K$. Since in the general case $d(P, Opt) > 0$ we should reformulate the bound on Ref in more flexible terms.

By general theory of real analysis, a sequence converges if and only if it is Cauchy [52]. A sequence of real numbers, $(Q_k)_{k\in\mathbb{N}}$, is Cauchy if the following condition holds:

$$\forall\epsilon, \quad \exists k_0 \quad , \text{such that } \forall k_1, k_2 \geq k_0. \quad \left\|Q_{k_1} - Q_{k_2}\right\| \leq \epsilon \tag{4.13}$$

It follows that Q is suitable for EA termination if it is Cauchy (thus, convergent) and its Cuachy condition provides an upper bound for Ref Cauchy condition.

Definition 2 (Weak Equivalence Condition) *A measure Q is suitable for EA termination if and only if $\forall\epsilon$, $\exists k$ such that $\forall k_1, k_2 > k$ we have the following equivalence of Cauchy sequences:*

$$||Q_{k_1} - Q_{k_2}|| \leq \epsilon \Rightarrow ||Ref_{k_1} - Ref_{k_2}|| \leq \epsilon' < \epsilon \tag{4.14}$$

Definition 2 is given in terms of Cauchy sequences, which deal with rates of convergence to the limit point of a sequence. Convergence rates are independent of the scale and range of the sequence values. In particular, Definition 2 states that Ref convergence rate is higher than Q convergence rate. It is worth noticing that this

can hold regardless of Ref and Q scales and limit points, which by no means need to be the same. Consider, for instance, $Q_1 = e^{-x}$ and $Q_2 = Q_1/100 - 100$. It should be clear that $Q_2 < Q_1$ and that they are neither in the same scale nor converge to the same point. However, $Q2$ has a higher convergence rate than $Q1$, so that, for any sequence $x_n \to \infty$, $Q_1(n) = Q_1(x_n)$ satisfies the Cauchy-measure condition with respect to $Q_2(n) = Q_2(x_n)$:

$$\begin{array}{l} \text{if } |Q_1(k_1) - Q_1(k_2)| = |e^{x_{k1}} - e^{x_{k2}}| \leq \epsilon \text{ , then} \\ |Q_2(k_1) - Q_2(k_2)| = |e^{x_{k1}} - e^{x_{k2}}|/100 \leq \epsilon' = \epsilon/100 < \epsilon \end{array}$$

The regression model is useful for checking theoretical condition like the one given in Definition 2 that can not be easily checked analytically [68]. An algebraic condition over β_0 and β_1 ensuring that Definition 2 is satisfied in statistical inference terms is given by the following Proposition:

Proposition 2 *Let P be the limit point of an EA algorithm reaching a steady state and β_1, β_0 the regression coefficients of the following logarithmic regression model:*

$$log(\widetilde{Ref}) = \beta_0 + \beta_1 log(\widetilde{Q}) + \varepsilon \tag{4.15}$$

for $\widetilde{Ref} = Ref - d(P, Opt)$, $\widetilde{Q} = Q - Q(P)$. If $\beta_1 > 0$ and the following inequality for the regression coefficients holds:

$$\beta_0 + 3\sigma \leq 0 \tag{4.16}$$

then $\widetilde{Q}$ satisfies the Cauchy-measure Condition.

Proof.

The Cauchy-condition given by definition 2 is satisfied if $\forall\epsilon$, $\exists k_0$ such that $\forall k_1, k_2 \leq k_0$, we have that:

$$|Q_{k_1} - Q_{k_2}| < \epsilon \Rightarrow |Ref_{k_1} - Ref_{k_2}| < \epsilon$$

By the regression model (4.34), we have that:

$$|Ref_{k_1} - Ref_{k_2}| = e^{\beta_0}|(Q_k^{\beta_1} e^{\varepsilon_{k_1}} - Q_{k_2}^{\beta_1} e^{\varepsilon_{k_2}})|$$

for ε_{k_1}, ε_{k_2} following a gaussian distribution $N(0, \sigma^2)$. Given that 99.7% of the values of a gaussian are in the interval given by three standard deviations, $[-3\sigma, 3\sigma]$, we have that:

$$
\begin{aligned}
&|Ref_{k_1} - Ref_{k_2}| \leq e^{\beta_0}|Q_{k_1}^{\beta_1}e^{3\sigma} - Q_{k_2}^{\beta_1}e^{-3\sigma}| = \\
&= e^{\beta_0}|Q_{k_1}^{\beta_1}e^{3\sigma} - Q_{k_1}^{\beta_1}e^{-3\sigma} + Q_{k_1}^{\beta_1}e^{-3\sigma} - Q_{k_2}^{\beta_1}e^{-3\sigma}| \leq \\
&\leq e^{\beta_0}(Q_{k_1}^{\beta_1}|e^{3\sigma} - e^{-3\sigma}| + e^{-3\sigma}|Q_{k_1}^{\beta_1} - Q_{k_2}^{\beta_1}|) \\
&= e^{\beta_0}(T_1 + T_2) \qquad (4.17)
\end{aligned}
$$

Since EA has converged to the minimum, $Ref_k \to 0$ and, by the linear regression, also $Q_k \to 0$. If $\beta_1 > 0$, then $Q_{k_1}^{\beta_1}$ is also convergent to zero. Therefore, T_1 and T_2 are bounded by:

$$
\begin{aligned}
T_1 &\leq \epsilon|e^{3\sigma} - e^{-3\sigma}| \\
T_2 &\leq \epsilon e^{-3\sigma}
\end{aligned}
$$

for k_1, k_2 large enough. Substituting in (4.17) and taking into account that $e^{3\sigma} - e^{-3\sigma} > 0$, we have that:

$$|Ref_{k_1} - Ref_{k_2}| \leq \epsilon e^{\beta_0}e^{3\sigma}$$

Therefore $|Ref_{k_1} - Ref_{k_2}| \leq \epsilon$ holds provided that:

$$\epsilon e^{\beta_0}e^{3\sigma} \leq \epsilon$$

Finally, taking logarithms we have that:

$$\beta_0 + 3\sigma \leq 0$$

The relation between ε and the model accuracy given by S_R proves the Proposition.

□

We would like to note that Proposition 2 is proven using a set of inequalities $\leq$ which achieve strict values for some cases and, thus, it is a sufficient but not necessary condition. This means that any measure satisfying the inequality is Cauchy, but implies by no means that all Cauchy measures should fulfill it. In this context, Proposition 2 is a way of verifying that a measure is Cauchy rather than a means of discarding non-Cauchy measures.

Corollary 1 *If $\widehat{\beta_1} > 0$ and the coefficients for the regression model of $\widetilde{Q}$ over $\widetilde{Ref}$ satisfy the following inequality:*

$$\widehat{\beta_0} + 3S_R \leq 0 \qquad (4.18)$$

then $\widetilde{Q}$ satisfies the Cauchy-measure Condition.

Proof. The Corollary follows from applying the regression estimation formulae given in Chapter 2. □

It is worth noticing that in the case that EA has succeeded in converging to the global optimum, a quantity fulfilling Proposition 2 terminates EA ensuring a given accuracy of its solution. This follows from the fact that a sequence of real numbers converges if and only if it is a Cauchy sequence. In this case, we have that $\forall\epsilon$, $(Q_k)_k$ satisfies:

$$\begin{array}{ll} \exists k_0^1, \text{ such that } \forall k > k_0^1, \; ||Q_k - Q_0|| < \epsilon & \text{(Convergent Sequence)} \\ \exists k_0^2, \text{ such that } \forall k_1, k_2 > k_0^2, \; ||Q_{k_1} - Q_{k_2}|| < \epsilon & \text{(Cauchy Sequence)} \end{array} \tag{4.19}$$

for Q_0 the limit value of $(Q_k)_k$. By considering $k_0 = max(k_0^1, k_0^2)$, the limit and Cauchy conditions hold for the same ϵ and $\forall k_1, k_2 > k_0$:

$$||Q_{k_1} - Q_{k_2}|| < \epsilon \Leftrightarrow ||Q_{k_1} - Q_0|| < \epsilon \tag{4.20}$$

The above condition combined with the Cauchy-measure Condition, implies that:

$$Q \text{ steady} \Leftrightarrow ||Q_{k_1} - Q_{k_2}|| \leq \epsilon \Rightarrow ||Ref_{k_1} - Ref_{k_2}|| \leq \epsilon \Leftrightarrow ||Ref_{k_1} - d(P, Opt)|| < \epsilon \tag{4.21}$$

for $d(P, Opt)$ the distance from the best individuals to the optimum. In the case EA has converged, $d(P, Opt) = 0$ and ϵ bounds the algorithm accuracy.

Finally, the Cauchy condition given in definition 2 can also be stated in terms of the range formula approximating the deviation from a steady state. Let us consider the first generations $K_{Ter}^{\tilde{q}}$ and K_{Ter}^{Ref} such that Q and Ref have reached their steady states.

Definition 3 (Empirical Cauchy-measure condition) *A quantity is a Cauchy measure if for all possible cost functions the following termination inequality holds:*

$$K_{Ter}^{\tilde{q}} \leq K_{Ter}^{Ref} \tag{4.22}$$

We would like to note that, in the next section, we will check in case that EA converges to the optimum, the relation between the parameter ϵ_{st} in formula (3.8) and the algorithm accuracy, provided that formula (3.8) is an accurate approximation to $\tilde{q}$ steady state.

$$K_{Ter}^{\tilde{q}} := \min_k (Rng_{\{k,\ldots,k+n_{st}\}}(\tilde{q}) < \epsilon_{st}) \tag{4.23}$$

$$K_{Ter}^{Ref} := \min_{k}(Rng_{\{k,\ldots,k+n_{st}\}}(Ref) < \epsilon_{st}) \tag{4.24}$$

Then, the continuous Cauchy condition given in definition 2 can be stated by the equation:

$$K_{Ter}^{\tilde{q}} \leq K_{Ter}^{Ref} \tag{4.25}$$

4.2 Prediction Formula

In case that Q satisfies Proposition 2, its values across EA iterations behave like Ref and, thus, we can use them as a bound on Ref for functions having similar landscapes and properties. In the case that Proposition 2 does not hold but there is a valid regression model (i.e. $\beta_1 \neq 0$) we can still give a bound on Ref from the values achieved by Q, using the regression prediction intervals [41] for the model in logarithmic scale given by 4.2.

The prediction intervals in logarithmic scale $PI(log(Q))$ are computed for each $log(Q)$ and provide ranges for $log(Ref) = \beta_0 + \beta_1 log(Q)$ at a given confidence level $1 - \alpha$. That is, given $log(Q)$, the values of the response $log(Ref)$ are within $L_{PI}(log(Q)) \leq log(Ref) \leq U_{PI}(log(Q))$ with a $(1-\alpha)100\%$ confidence level.

Given $log(Q)$, the confidence interval at a confidence level (1-α) predicting $log(Ref)$ is given by:

$$\begin{aligned} PI_{log(Ref)} &= PI_{log(Ref)}(log(Q)) = [L_{PI_{log(Ref)}(log(Q))}, U_{PI_{log(Ref)}(log(Q))}] = \\ &= [\widehat{log(Ref)} - t_{\alpha/2}^{N-2} S_R \sqrt{1+h_0}, \widehat{log(Ref)} + t_{\alpha/2}^{N-2} S_R \sqrt{1+h_0}] \end{aligned} \tag{4.26}$$

for S_R given by 2.6, $t_{\alpha/2}^{N-2}$ the value of a T-Student distribution with $N-2$ degrees of freedom having a cumulative probability equal to $\alpha/2$ and $\widehat{log(Ref)}$, h_0 given by:

$$\begin{aligned} &\widehat{log(Ref)} = b_0 + b_1 Q \\ &h_0 = (1 \quad x_0)(X^T X)^{-1} \begin{pmatrix} 1 \\ Q \end{pmatrix} = a_0 + a_1 Q + a2Q^2 \end{aligned} \tag{4.27}$$

Where $(a_0.a_1, a_2)$ stand for the coefficients of the quadratic polynomial resulting from the previous algebraic expression. **formula per les as** The exponential of PI already provides (with confidence $1-\alpha$) an upper bound for the accuracy of EA solution given EA current state. In order to obtain the upper bound for $log(Q)$ ensuring a given accuracy $U_{PI}(Q) = \epsilon$, it suffices to find the value $log(Q)$ that solves:

$$\widehat{log(Ref)} + t_{\alpha/2}^{N-2} S_R \sqrt{1+h_0} = U_{PI}(log(Q)) = \epsilon \tag{4.28}$$

Using the expressions for $\widehat{log(Ref)}$ and h_0 in (4.27) and solving (4.28) for $log(Q)$, we obtain:

$$log(\tilde{Q}) = \frac{2b_0b_1 - t^{N-2}_{\alpha/2^2}S^2_R a_1 - 2b_1U_{PI}(log(Q)) + \sqrt{D}}{2(t_{\alpha/2^2}S^2_R a_2 - b^2_1)} \tag{4.29}$$

where the discriminant is given by:

$$D = (t^{N-2}_{\alpha/2^2}S^2_R a_1 - 2b_0b_1 + 2b_1U_{PI}(log(Q)))^2 - 4(t^{N-2}_{\alpha/2^2}S^2_R a_2 - b^2_1)(t^{N-2}_{\alpha/2^2}S^2_R(a_0+1) - \\ - b^2_0 + 2U_{PI}(log(Q))b_0 - U_{PI}(log(Q))^2) \tag{4.30}$$

The formula (4.29) gives the point estimation value $log(\tilde{Q})$ ensuring $log(Ref)$ is bounded with confidence (1-α). In order to find the value of the alternative quantity, $\tilde{Q}$ that actually bounds the Ref, we must take exponentials to invert the logarithmic scale and find the bound in the normal scale. By taking exponentials from (4.28) we get the value for $log(\tilde{Q})$ ensuring $Ref_0 < \epsilon$ with confidence $1-\alpha$ as:

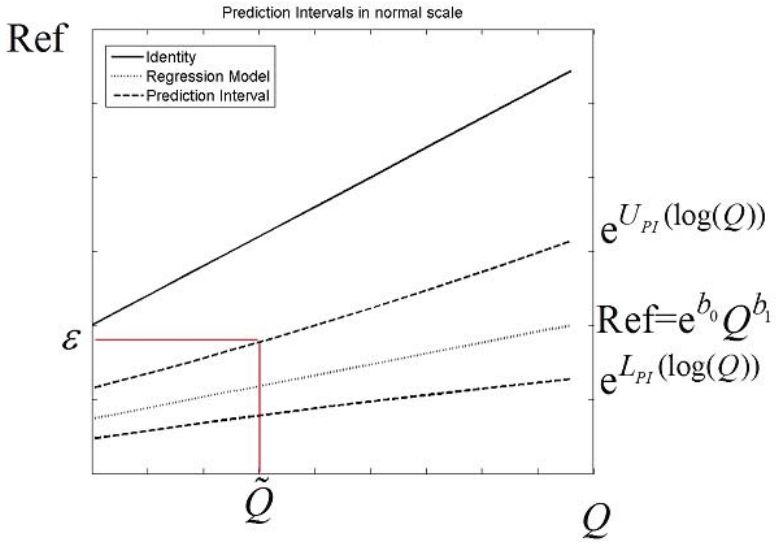
$$log(Ref) \leq U_{PI}(log(\tilde{Q})) \Rightarrow Ref \leq exp^{U_{PI}(log(\tilde{Q}))} = \epsilon \tag{4.31}$$

In figure 4.2 and figure 4.3, we illustrate the different casuistics that might appear in the definition of the bound P_{UP} depending on the deviation from identity of the regression model slope. We show prediction intervals (dashed lines), the regression model (dotted line) and the identity line (solid line) in the logarithmic scale as well as their transformation in the normal scale after taking exponentials. In the normal scale we observe a non-linearity for the prediction intervals with increasing prediction ranges due to the model heteroscedasticity. In both scales, we show the value of the independent variable, either $log(\tilde{Q})$ or $\tilde{Q}$, that ensures an upper bound on either $log(Ref)$ or Ref given by the relation (4.31). In case the regression model is under the identity, as illustrated in figure 4.2, $log(Ref) = log(\tilde{Q}) * b_1 + b_0 < log(\tilde{Q})$, the value $log(\tilde{Q})$ (and, thus, $\tilde{Q}$) is higher than the considered accuracy $log(\epsilon)$ (ϵ in normal scale). And, in fact, $\tilde{Q} = \epsilon$ always achieves the accuracy bound. In case the regression model is above the identity, as shown in figure 4.3), $log(Ref) = log(\tilde{Q}) * b_1 + b_0 > log(\tilde{Q})$, neither $log(\tilde{Q})$ nor $\tilde{Q}$ can not bound Ref by themselves, and we should iterate Q more than the number of iterations required to have Ref bound ϵ.

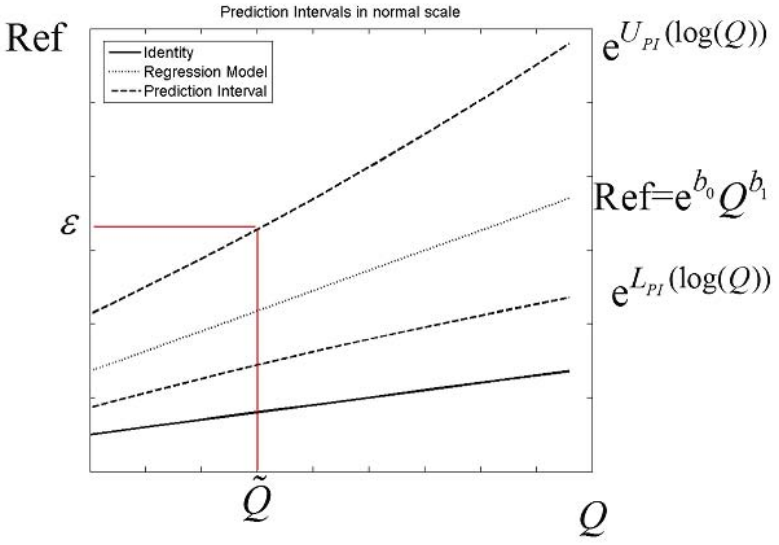
4.3 Experimental Set-Up

The goal of these experiment is to validate our framework for assessing the quality of existing stopping conditions and to check its applicability to unknown black-box problems.

The following experiments have been carried out:



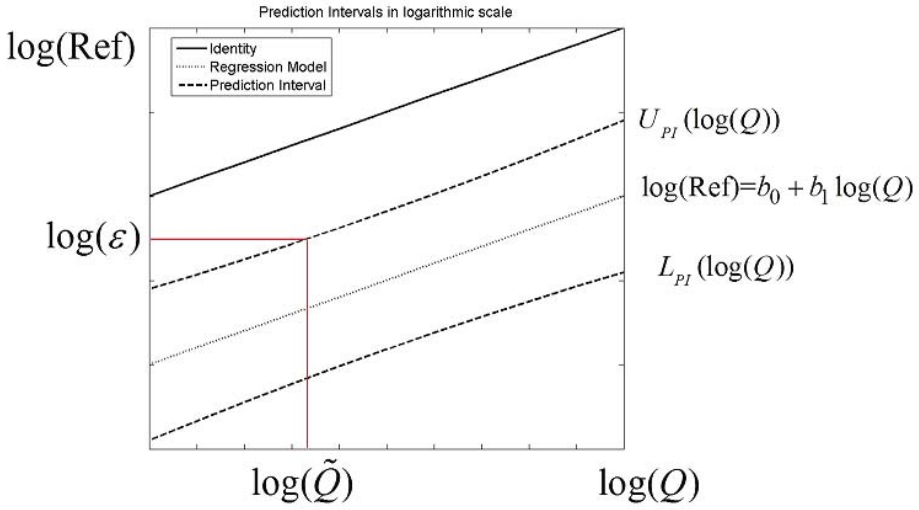
(a)



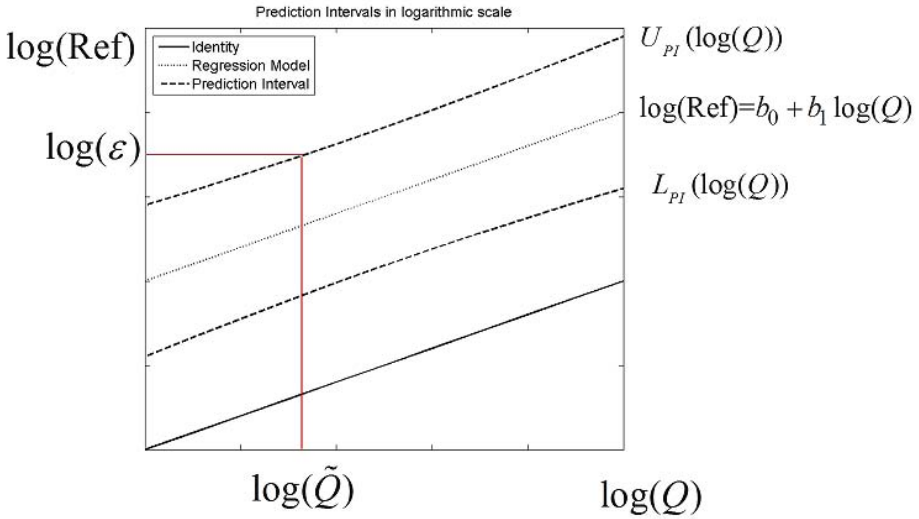
(b)

Figure 4.2: The predicted upper bound and the equivalence condition in case of a good, (a), and poor, (b), equivalence relation.

1. **Assessment of the Regression Model.** In order to correctly apply our quality framework, there must exist a true linear relation [2]. This is assessed



(a)



(b)

Figure 4.3: The predicted upper bound and the equivalence condition in case of a good, (a), and poor, (b), equivalence relation.

by the t-test on the slope $\widehat{\beta_1}$ and the R^2 statistic as described in Chapter 2

2. **Assessment of Current Termination Conditions.** We have applied our evaluation framework to those cases that have a reliable regression model. The

quality of termination quantities have been assessed using the strong, weak and empirical substitution conditions described in Section 4.1.

3. **Applicability to Black-Box Problems.** We have checked to what extent results generalize to more realistic settings. In particular we have checked that the weak condition holds in case of perturbations. Finally, we have also explored to what extent termination of EA using Q relates to its accuracy in case of convergence to the optimum by showing that prediction intervals provide an actual bound for Ref.

Like in the experiments of Chapter 3, the regression framework has been applied to the following quantities based on improvement in f-space, *Fitval*, distribution of EA population in x-space, *MxD*, as well as, the distance to the function minimum in x-space, *Ref*:

$$\begin{array}{l} Imp = \frac{1}{n_{best}} \sum_{j=1}^{n_{best}} f(Ind_j) \\ MxD = max_{j \in (1,\ldots,n_{best})} d(Ind_j, Ind_1) \\ Ref := max_{j \in (1,\ldots,n_{best})} d(Ind_j, Opt) \end{array} \tag{4.32}$$

for $(Ind_j)_{j=1}^{n_{best}}$ the set of the n_{best} best individuals and Opt the function optimum in x-space.

We have chosen Differential Evolution (DE) paradigm with the settings given in Chapter 3 for its dependency on diversity for achieving steady states (see experiments in Chapter 3) and its proven success for solving global optimization over continuous spaces [60, 13]. The range formula for checking the empirical condition has been computed using the optimal parameters selected in Chpater 3, $\epsilon_{st} = 10^{-6}, n_{st} = 500$.

We have applied our framework to test the 31 state-of-the-art evolutionary algorithms cited in Chapter 3 ([24, 25]). For a representative set of the functions, the benchmark also includes noisy versions with different degrees of noise [16]. These functions have been used in our third experiment. In table 4.1 we show the test function set with their main features. Each test function is checked with a $\surd$ if the property holds, with a $\surd\surd$ if the property strongly holds and with a X if the property does not hold. For illustration of the function landscapes associated to tehe main properties , figure 4.4 shows a pair of functions for each of the properties (multimodality, scalability and global structure) related to EA performance. At the top of the figure, we show functions favoring convergence of EA schemes to the function optimum, whereas functions favoring the opposite behavior are shown in bottom images. At the top of the figure, we show functions favoring convergence of EA schemes to the function optimum, whereas functions favoring the opposite behavior are shown in bottom images.

We would like to note that the experiments have been done excluding the functions LinearSlope, StepEllipsoidal, Weierstrass and Katsuura (figure 4.4). These functions have infinite global minima, either a discrete set (Weierstrass, Katsuura) or even a submanifold (plateau) of them (LinearSlope, StepEllipsoidal). These cases have to be dropped from our analysis for two main reasons. First, the gold-standard reference measure of the algorithm accuracy given by the distance to the global optimum cannot

be computed in a reliable manner. Second, functions having multiple global minima require specific EA paradigms and other stopping criteria different from reaching a steady state (see discussion in Section 4.5 for further details). The remaining set of 20 functions include multimodal functions having several local minima with different EA convergence behaviours: convergence to optimum (Rastrigin, Schaffers and Gallaghers 101) and premature convergence (Griewank-Rosenbrock, Schwefel, Gallagher21 and Lunacek).

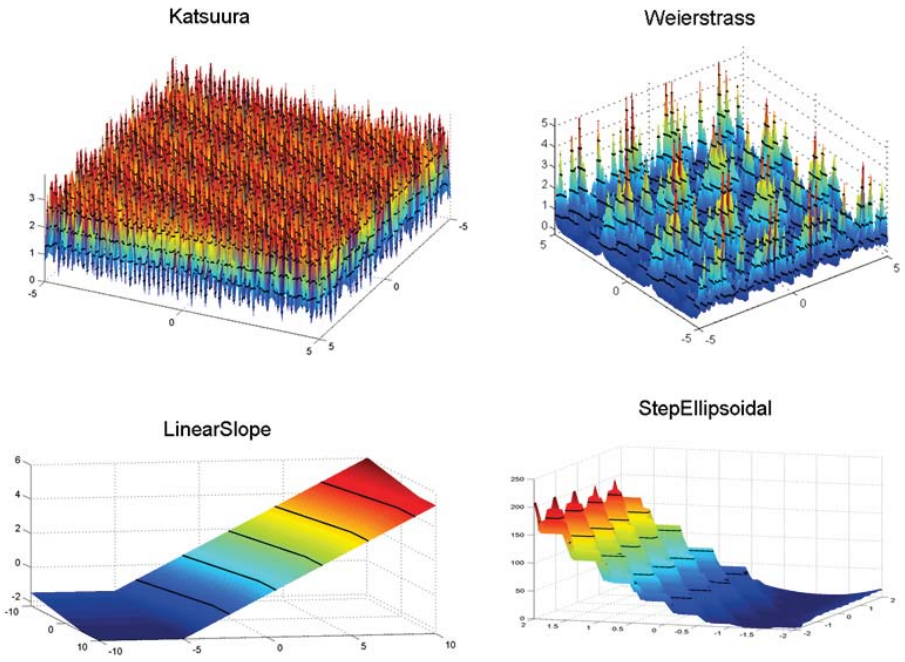


Figure 4.4: Dropped functions having multiple global minima, a finite discrete set functions in top images and an infinite set given by a manifold in bottom images

As in Chapter 3, in order to account for variability across initial population, a total number of 30 runs per function have been performed. For each run 3000 EA iterations were performed in order to ensure convergence to the steady state.

For each quantity and test function, the generic regression model is:

$$log(Ref_k) = \beta_0 + \beta_1 log(Q_k) + \varepsilon_k \tag{4.33}$$

for Ref_k, Q_k a random sampling of size 1000 of the values obtained across runs and iterations. In case EA steady state is reached in finite time (usually due to premature convergence as discussed in Section 4.5), repetitions of values at final iterations have been removed to preserve uncorrelation of the linear model.

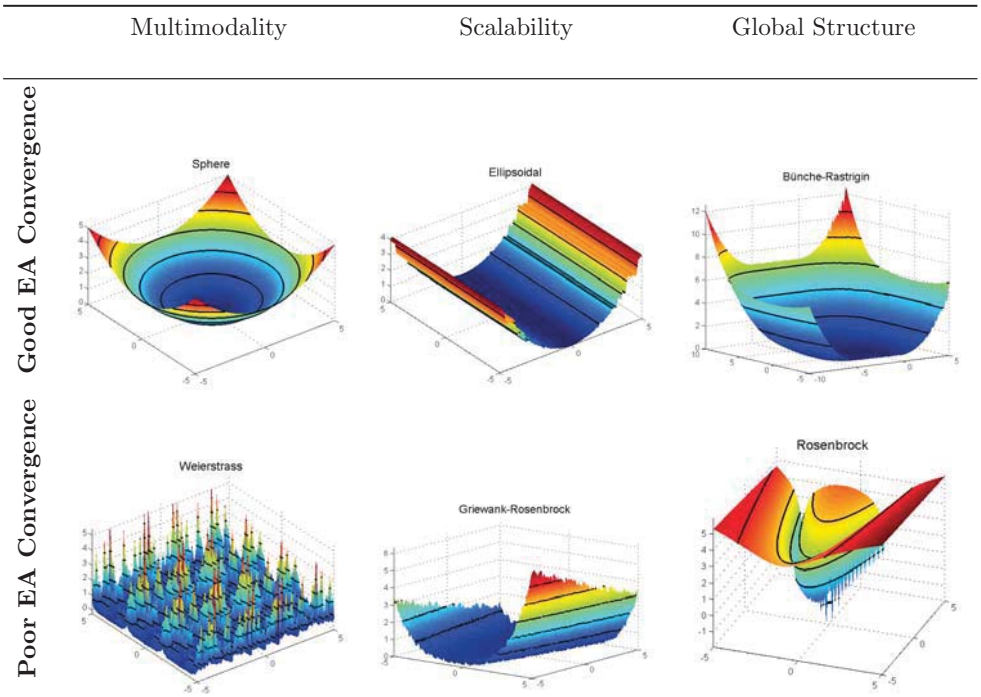


Figure 4.5: Representative set of the benchmark test functions.

We note that, by taking exponentials, the regression model in the original scale is polynomial with multiplicative errors:

$$Ref_i = e^{\beta_0} Q_i^{\beta_1} e^{\varepsilon_i} \tag{4.34}$$

4.4 Results

4.4.1 Assessment of Regression Models

Tables 4.2 and 4.3 report the $p-value$ of the t-test on the slope, R^2 and the parameters of the regression model ($\widehat{\beta_0}$, $\widehat{\beta_1}$ and S_R) for the 20 noiseless test functions having a unique global minimum. The functions are split in 5 groups: functions 1-4 are separable functions (group 1), functions 5-9 have a low (good) conditioning (group 2), functions 10-14 have a high (bad) conditioning (group 3), functions 15-19 have a strong global structure (group 4) and functions 20-24 have a weak global structure (group 5). Results for *MxD* are given in Table 4.2 and results for *Imp* in Table 4.3. Regressions having a either $p - value > 0.05$ or $R^2 < 0.9$ are discarded. For both quantities, the regression model is consistent for all functions.

Function	Mult.	gl.str.	separ	scaling	homog.	basins	gl.loc	plat.
Sphere	X	X	$\surd\surd$	X	$\surd\surd$	X	X	X
Ellipsoidal	X	X	$\surd\surd$	$\surd\surd$	$\surd\surd$	X	X	X
Rastrigin	$\surd\surd$	$\surd\surd$	X	$\surd$	$\surd\surd$	$\surd$	$\surd$	X
Bche-Rastrigin	$\surd\surd$	$\surd\surd$	$\surd\surd$	$\surd$	$\surd\surd$	$\surd$	$\surd$	X
Linear Slope	X	X	$\surd\surd$	X	$\surd\surd$	X	X	$\surd$
Attractive Sec.	X	X	$\surd\surd$	$\surd$	$\surd$	X	X	X
Step Ellipsoidal	X	X	$\surd\surd$	$\surd$	$\surd\surd$	X	X	$\surd$
Rosenbrock	$\surd$	X	X	X	$\surd$	$\surd$	$\surd$	X
Rosenbrock Rot.	$\surd$	X	X	X	$\surd$	$\surd$	$\surd$	X
Ellipsoidal	X	X	X	$\surd\surd$	$\surd\surd$	X	X	X
Discus	X	X	X	$\surd\surd$	$\surd\surd$	X	X	X
Bent Cigar	X	X	X	$\surd\surd$	$\surd\surd$	X	X	X
Sharp Ridge	X	X	X	$\surd$	$\surd$	X	X	X
Different Powers	X	X	X	$\surd$	$\surd$	X	X	X
Rastrigin	$\surd\surd$	$\surd\surd$	X	$\surd$	$\surd\surd$	$\surd$	$\surd$	X
Weierstrass	$\surd\surd\surd$	$\surd$	X	$\surd$	$\surd\surd$	$\surd$	$\surd$	X
Schaffers	$\surd\surd$	$\surd$	X	$\surd$	$\surd$	$\surd$	$\surd\surd$	X
Schaffers Conditioned	$\surd\surd$	$\surd$	X	$\surd\surd$	$\surd$	$\surd$	$\surd\surd$	X
Griewank-Rosenbrock	$\surd\surd$	$\surd\surd$	X	X	$\surd\surd$	$\surd$	$\surd$	X
Schwefel	$\surd$	*dec*	X	X	$\surd\surd$	$\surd$	$\surd$	X
Gallagher's 101	$\surd$	X	X	$\surd$	$\surd\surd$	$\surd$	$\surd$	X
Gallagher's 21	$\surd$	X	X	$\surd$	$\surd\surd$	$\surd$	$\surd$	X
Katsura	$\surd\surd\surd$	X	X	X	$\surd\surd$	$\surd$	$\surd$	X
Lunacek	$\surd\surd$	$\surd$	X	$\surd$	$\surd\surd$	$\surd$	$\surd$	X

Table 4.1: Test Functions and their features

For the distribution-based quantity, the goodness-of-fit is excellent, given that S_R is extremely small compared to the variable ranges. It is worth noticing that the slope β_1 is very close to 1 and $\widehat{\beta_0}$ is always negative. This implies that the relation in logarithmic scale is a translation of the identity and the regression model in the original scale is also linear. Besides a negative $\widehat{\beta_0}$ ensures that *MxD* might guarantee an upper bound for EA accuracy. It follows that *MxD* is a promising alternative quantity to stop EA ensuring a given accuracy. Regarding results for the improvement-based quantity reported in Table 4.3, the goodness-of-fit is also excellent. Unlike *MxD*, the slope $\widehat{\beta_1}$ takes any value in $[0, 1]$ and $\widehat{\beta_0}$ is not negative for all cases. This implies that the quantity does not provide an upper bound for EA accuracy and, thus, is worse suited for stopping EA.

4.4.2 Assessment of Current Termination Conditions

For this experiment we have used the noiseless test functions in order to test the two following issues:

Table 4.2: Distribution-based Quantity. Regression Model.

	Function	Model Verification		Model Parameters		
		$p-value$	$R-sq$	$Constant$	$slope$	SR
1	Esphere	$6.4E-208$	0.99	-3.50	1.00	0.59
2	Ellipsoidal	$1.4E-201$	0.99	-3.7	1.00	0.71
3	Rastrigin	$6.6E-03$	0.99	-1.94	1.10	0.26
4	Bunche-Rastrigin	$1.75E-212$	0.99	-3.59	1.00	0.70
6	Attractive-Sector	$7.99E-20$	0.99	-3.30	1.00	0.86
8	Rosenbrock	$2.08E-19$	0.99	-3.65	0.99	1.38
9	Rosenbrock-Rotated	$1.26E-28$	0.99	-2.90	1.03	0.84
10	Ellipsoidal-HC	$2.59E-16$	0.99	-3.71	1.02	0.88
11	Discus	$5.98E-16$	0.99	-4.20	0.96	0.58
12	Bent-Cigar	$4.16E-131$	1.00	-3.38	1.00	1.33
13	Sharp-Ridge	$8.35E-37$	0.96	-2.83	1.00	0.95
14	Different-Powers	$2.41E-25$	0.92	-2.03	1.03	1.32
15	Rastrigin-Multimodal	$1.28E-14$	0.99	-3.46	1.00	1.07
17	Schaffers	$2.58E-17$	0.99	-3.56	1.02	1.03
18	Schaffers-Conditioned	$5.52E-28$	0.98	-3.11	1.05	1.23
19	Griewank-Rosenbrock	$2.66E-04$	0.94	1.26	1.67	2.51
20	Schwefel	$7.66E-04$	0.91	-7.61	0.72	3.10
21	Gallagher-101	$5.57E-06$	0.93	-4.32	0.96	2.18
22	Gallagher-21	$2.60E-07$	0.99	-3.15	1.04	0.11
24	Lunacek	$3.79E-09$	0.94	-4.38	0.94	3.23

1. **Strong Interchange Condition.** We have checked the definition 1 for distribution-based (MxD) and improvement-based ($Fitval$) measures using our benchmark test functions. Table 4.5 and 4.4 contains the ϵ values for MxD and show to what extent the slope β_1 is equal to 1 in the case of distributed-based measure, and, thus, to what extent our measure can substitute our standard goal measure Ref. As we can expect, the strong condition does not hold for all the functions but we can see that most of them keep ϵ bellow 10^{-2}. It is worth noticing that functions with a poor ϵ was precisely poor R^2 ones. That fact indicates that more strong linear relation more the slope closes to 1. Then, more alternative and gold-standard measures are interchangeable. In the case of improvement-based measure, $Fitval$, practically all of ϵ values are greater than 10^{-1} and, thus, the substitute condition does not hold.

2. **Weak and Empirical Interchange Condition.** We have checked that the statistical requirements given in Proposition 2 ensure the Cauchy-measure condition given by **Ref Definici** for a set of benchmark test functions. The Cauchy-measure condition given in Proposition 2 has been computed over the meaningful regression models. The theoretical condition has been compared to the empirical termination given by (4.22). For each case, the empirical termination has

Table 4.3: Improvement-based Quantity. Regression Model.

	Function	**Model Verification**		**Model Parameters**		
		$p-value$	$R-sq$	$Constant$	$slope$	SR
1	Esphere	0	1	0	0.5	0.00
2	Ellipsoidal	$2.42E-256$	0.99	-0.34	0.5	0.05
3	Rastrigin	$3.93E-05$	0.99	-2.56	0.51	0.01
4	Bunche-Rastrigin	$1.80E-218$	0.99	-2.20	0.50	0.53
6	Attractive-Sector	$6.85E-22$	0.99	1.67	0.28	0.52
8	Rosenbrock	$3.81E-24$	0.99	0.12	0.50	0.44
9	Rosenbrock-Rotated	$1.98E-30$	0.99	1.38	0.50	0.62
10	Ellipsoidal-HC	$1.14E-19$	0.99	0.65	0.50	0.33
11	Discus	$2.46E-19$	0.99	-0.69	0.50	0.22
12	Bent-Cigar	$3.26E-140$	0.99	0.03	0.50	0.83
13	Sharp-Ridge	$1.28E-39$	0.97	1.22	0.50	0.73
14	Different-Powers	$1.59E-29$	0.95	2.47	0.37	0.84
15	Rastrigin-Multimodal	$2.79E-22$	0.99	0.41	1.01	0.09
17	Schaffers	$6.48E-18$	0.99	-0.70	1.00	0.86
18	Schaffers-Conditioned	$4.42E-29$	0.99	-1.21	1.04	1.04
19	Griewank-Rosenbrock	$2.4E-04$	0.97	3.15	0.52	0.95
20	Schwefel	$1.5E-03$	0.98	-5.07	0.58	0.63
21	Gallagher-101	$5.98E-07$	0.98	-0.77	0.24	0.72
22	Gallagher-21	$4.24E-07$	0.99	-0.39	0.24	0.13
24	Lunacek	$6.26E-09$	0.95	-3.28	0.47	2.96

been checked using a proportion test for the variable:

$$X_{Dif} := \begin{cases} 1 & \text{if } K^{Q}_{Ter} - K^{Ref}_{Ter} > 0 \\ 0 & \text{otherwise} \end{cases}$$

using $q_0 = 0.95$ and $\alpha = 0.05$.

Tables 4.8 and 4.9 report the average termination iterations $\mu(K^{Q}_{Ter})$ and $\mu(K^{Ref}_{Ter})$ computed across EA runs, p-values for the termination condition (4.22) and the Cauchy condition given by Proposition 2. We use a $\surd$ if conditions are satisfied and a X otherwise. In the case of the termination given by (4.22) a $\surd$ is given if the p-value is less that $\alpha = 0.05$. In the case of Proposition 2 a $\surd$ indicates that the inequality of its Corollary is satisfied. Results for *MxD* are given in Table 4.8 and results for *Imp* in Table 4.9. For all cases, a steady state for *MxD* also implies Ref steady state as p-values are below 0.01. It follows that the quantity is a good a candidate for terminating EA, at least for the benchmark test set. Besides the theoretical condition ensuring generalization of results is fulfilled in $13/20 = 65\%$ cases. In fact, if we consider the joint regression model for the 20 functions, we obtain a regression parameters $\beta_0 = -3.079$, $\beta_1 = 1.000$, $S_R = 0.813$ and $R-sq = 0.999$ that also satisfy the Cauchy condition. This indicates that the quantity is also well-conditioned for

Table 4.4: β_1 Condition MxD

	Function	S_{β_1}	$CI-$	$CI+$	ϵ
1	Esphere	$5.61E-04$	0.99	1.00	$1.47E-03$
2	Ellipsoidal	$6.53E-04$	0.99	1.00	$1.75E-03$
3	Rastrigin	$4.64E-02$	0.90	1.30	$2.35E-01$
4	Bunche-Rastrigin	$5.98E-04$	0.99	1.00	$1.41E-03$
6	Attractive-Sector	$2.32E-02$	0.95	1.05	$4.29E-02$
8	Rosenbrock	$3.06E-02$	0.93	1.05	$6.38E-02$
9	Rosenbrock-Rotated	$1.93E-02$	0.99	1.07	$6.52E-02$
10	Ellipsoidal-HC	$2.81E-02$	0.96	1.07	$6.52E-02$
11	Discus	$2.27E-02$	0.91	1.01	$7.96E-02$
12	Bent-Cigar	$4.20E-03$	0.99	1.01	$1.07E-02$
13	Sharp-Ridge	$2.71E-02$	0.94	1.05	$4.87E-02$
14	Different-Powers	$5.29E-02$	0.93	1.14	$1.21E-01$
15	Rastrigin-Multimodal	$3.20E-02$	0.94	1.08	$6.54E-02$
17	Schaffers	$2.34E-02$	0.97	1.07	$5.74E-02$
18	Schaffers-Conditioned	$2.82E-02$	0.99	1.10	$9.41E-02$
19	Griewank-Rosenbrock	$2.92E-01$	0.93	2.43	1.26
21	Schwefel	$1.75E-01$	0.27	1.17	$6.31E-01$
21	Gallagher-101	$1.34E-01$	0.65	1.27	$2.90E-01$
22	Gallagher-21	$1.37E-02$	1.01	1.07	$6.69E-02$
24	Lunacek	$1.23E-01$	0.68	1.21	$2.74E-01$

terminating EA for other functions presenting similar landscapes. In the case of *Imp*, only 3 functions (Bunche-Rastrigin, Schaffers-Conditioned and Schwefel) satisfy the empirical termination condition. However only two of the former functions satisfy the theoretical condition given in Proposition 2. This indicates that the capability of *Imp* for detecting EA steady state will be very low in general.

4.4.3 Applicability to Black-Box Problems

In order to assess the capability of the statistical requirements for accurate termination we have computed the proportions test (3.13) in the set of noisy functions [16]. This set is representative of the main properties characterizing cost functions but includes several degrees of noise that distort the expected profile like in real-world applications. There is a function for each of the 5 groups [16]: Sphere for group 1, Rosenbrock for group 2, Ellipsoidal for group 3, Schaffers for group 4 and Gallagher101 for group 5. We want to note that DE converged to the optimum in all cases.

The generalization capability is validated in the measure that the termination condition agrees with the statistical requirements of Proposition 2 computed in experiment 4.4.2. In order to assess the accuracy of the EA solution given by the experimental termination condition, we have also compared the distances between

Table 4.5: β_1 Condition $ImpFunc$

	Function	S_{β_1}	$CI-$	$CI+$	ϵ
1	Esphere	$1.11E-30$	0.5	0.5	$5.00E-01$
2	Ellipsoidal	$2.23E-05$	0.49	0.49	$5.00E-01$
3	Rastrigin	$1.25E-04$	0.5	0.5	$4.95E-01$
4	Bunche-Rastrigin	$2.25E-04$	0.49	0.49	$5.02E-01$
6	Attractive-Sector	$3.86E-03$	0.27	0.28	$7.31E-01$
8	Rosenbrock	$4.85E-03$	0.49	0.51	$5.12E-01$
9	Rosenbrock-Rotated	$7.05E-03$	0.49	0.52	$5.06E-01$
10	Ellipsoidal-HC	$5.19E-03$	0.48	0.50	$5.15E-01$
11	Discus	$4.37E-03$	0.48	0.50	$5.16E-01$
12	Bent-Cigar	$1.30E-03$	0.49	0.50	$5.01E-01$
13	Sharp-Ridge	$1.05E-02$	0.48	0.52	$5.14E-01$
14	Different-Powers	$1.18E-02$	0.34	0.39	$6.55E-01$
15	Rastrigin-Multimodal	$2.58E-03$	1.00	1.02	$1.94E-02$
17	Schaffers	$1.92E-02$	0.96	1.04	$3.73E-02$
18	Schaffers-Conditioned	$2.37E-02$	0.99	1.09	$8.43E-02$
19	Griewank-Rosenbrock	$4.12E-02$	0.41	0.64	$5.65E-01$
21	Schwefel	$4.08E-02$	0.44	0.70	$5.21E-01$
21	Gallagher-101	$1.19E-02$	0.21	0.27	$7.83E-01$
22	Gallagher-21	$3.69E-03$	0.23	0.25	$7.69E-01$
24	Lunacek	$5.66E-02$	0.35	0.60	$6.27E-01$

the best individual and the function optimum at the termination iterations K^{Q}_{Ter} and K^{Ref}_{Ter}. The empirical termination condition ensures a given accuracy in the measure that K^{Q}_{Ter} distance is smaller than the one achieved at K^{Ref}_{Ter}. A one-tailed paired t-test for the difference in means is used to check if the accuracy achieved by K^{Q}_{Ter} bounds EA actual accuracy.

Tables 4.10 and 4.11 report the average termination iterations $\mu(K^{Q}_{Ter})$ and $\mu(K^{Ref}_{Ter})$ for the 100 EA runs and the proportion test for the empirical termination condition as in Tables 4.8 and 4.9. We also report the average distances (noted by $\mu(D^{Q})$ and $\mu(D^{Ref})$) between the best individual and the function optimum at the termination iterations K^{Q}_{Ter} and K^{Ref}_{Ter}, as well as, p-values for the one-tailed paired t-test on their difference. Results for *MxD* are given in Table 4.10 and results for *Imp* in Table 4.11.

As predicted by the Cauchy condition, *MxD* terminates EA at its steady state, while *Imp* always fails to terminate EA at the steady state. It is worth noticing the case of the Schaffers function because it is illustrative of the usefulness and reliability of the inference provide by our Cauchy condition. The empirical termination condition was successful in the noiseless function (case 18 in Table 4.9), but the Cauchy condition was not satisfied. This indicated that the empirical conclusion was not reliable enough to be generalized. This is proved by the empirical failure reported in Table 4.11.

Concerning distances, the accuracy achieved by K^{MxD}_{Ter} always bounds (with p-

Table 4.6: β_0 Condition MxD

	Function	S_{β_0}	$CI-$	$CI+$	¶$value$
1	Esphere	$1.27E-01$	-3.75	-3.24	$1.40E-47$
2	Ellipsoidal	$1.44E-01$	-3.98	-3.41	$1.55E-44$
3	Rastrigin	$5.14E-01$	-4.15	0.26	0.032
4	Bunche-Rastrigin	$1.26E-01$	-3.84	-3.34	$1.66E-49$
6	Attractive-Sector	$3.78E-01$	-4.09	-2.51	$2.20E-08$
8	Rosenbrock	$5.83E-01$	-4.86	-2.42	$2.63E-06$
9	Rosenbrock-Rotated	$3.53E-01$	-3.62	-2.17	$3.02E-09$
10	Ellipsoidal-HC	$4.00E-01$	-4.56	-2.86	$3.77E-08$
11	Discus	$3.42E-01$	-4.92	-3.47	$7.40E-10$
12	Bent-Cigar	$2.61E-01$	-3.90	-2.86	$2.26E-22$
13	Sharp-Ridge	$3.53E-01$	-3.54	-2.12	$8.68E-11$
14	Different-Powers	$5.34E-01$	-3.11	-0.95	$2.21E-04$
15	Rastrigin-Multimodal	$5.32E-01$	-4.60	-2.31	$7.02E-06$
17	Schaffers	$3.83E-01$	-4.38	-2.74	$6.51E-08$
18	Schaffers-Conditioned	$4.03E-01$	-3.94	-2.29	$6.57E-09$
19	Griewank-Rosenbrock	2.22	-4.43	6.95	0.70
21	Schwefel	$2.41E-01$	-13.81	-1.41	0.01
21	Gallagher-101	$1.22E-01$	-7.13	-1.52	0.003
22	Gallagher-21	$9.97E-02$	-3.39	-2.90	0.88
24	Lunacek	1.83	-8.33	-0.43	0.02

values under 10^{-20}) the actual EA accuracy given by the distance at iteration K_{Ter}^{Ref}. This agrees with the theoretical properties of Cauchy-measures. On the other side, distances given by *Fitval* do not present a clear relation with the ones given by the reference measure Ref. As reported in the literature [72], the large variability in their values strongly depends on the properties of the cost functions. A comparison between Rosenbrock and Gallagher101 functions illustrates this behavior. Although, for both functions, *Fitval* stopped EA before Ref in around 170 iterations (see Table 4.11), their accuracies are very different. It is also worth noticing that differences in accuracies do not seem to be related to the differences between *Fitval* and Ref iterations. This is illustrated again by comparison between Ellipsoidal and Gallagher101 functions. Ellipsoidal is the function having the least number of difference in iterations, but, it is worse than Gallagher101 in terms of accuracy loss.

4.5 Discussion

There are several topics arising from the theoretics and experiments presented in this paper that deserve a detailed discussion.

1. Behavior in case of Premature Convergence

Table 4.7: β_0 Condition $ImpFunc$

	Function	S_{β_0}	$CI-$	$CI+$	¶$value$
1	Esphere	$5.12E-28$	$9.26E-15$	$9.26E-15$	1.00
2	Ellipsoidal	$9.97E-03$	-0.35	-0.31	$5.49E-55$
3	Rastrigin	$2.89E-03$	-2.57	-2.55	$6.35E-07$
4	Bunche-Rastrigin	$9.57E-02$	-2.39	-2.01	$1.27E-41$
6	Attractive-Sector	$2.90E-01$	1.06	2.27	0.99
8	Rosenbrock	$2.16E-01$	-0.33	0.57	0.70
9	Rosenbrock-Rotated	$3.17E-01$	0.72	2.02	0.99
10	Ellipsoidal-HC	$1.92E-01$	0.25	1.06	0.99
11	Discus	$1.57E-01$	-1.02	-0.35	0.00
12	Bent-Cigar	$1.70E-01$	-0.30	0.37	0.57
13	Sharp-Ridge	$3.50E-01$	0.51	1.92	0.99
14	Different-Powers	$4.75E-01$	1.51	3.43	0.99
15	Rastrigin-Multimodal	$5.12E-02$	0.30	0.52	0.99
17	Schaffers	$3.60E-01$	-1.47	0.06	0.03
18	Schaffers-Conditioned	$3.77E-01$	-1.98	-0.44	0.00
19	Griewank-Rosenbrock	1.03	0.30	6.00	0.98
21	Schwefel	$6.92E-01$	-7.26	-2.86	0.00
21	Gallagher-101	$5.46E-01$	-2.06	0.52	0.10
22	Gallagher-21	$1.57E-01$	-0.77	0.00	0.02
24	Lunacek	1.79	-7.17	0.60	0.04

Concerning cases in which algorithms diverge, EA population stagnates and cannot reach a steady state and the iterative process does not converge to any point. Divergent cases failing to reach a steady state are always problematic and often require stopping EA using a maximum number of iterations, given that EA population is prone to behave erratically and there is no clear quantity reaching a steady state. Divergent cases usually arise due to the choice of an EA paradigm that fails to sufficiently explore the function profile. This is the case of DE and the functions LinearSlope, StepEllipsoidal, Weierstrass and Katsuura that have multiple global minima. These functions require niche search EA paradigms, like crowding [1], which present a different asymptotic behaviour given by the multiple global minima searched by the algorithm. Our methodology could be applied provided that a different kind of measure of EA steady state taking into account the multiple convergence was defined. This is future research out of the scope of the presented work. The quality of the fit depends on the properties of EA asymptotic behavior. There are 3 possible general asymptotic behaviors for EA: convergence to optimum, premature convergence and stagnation. Each case has a characteristic regression model and EA population configuration illustrated in figure 4.10. We show the regression in logarithmic scale for $Q = MxD$ (middle scatter plots) and $Q = Imp$ (right scatter plots) in right plots. The left most scatter shows EA population configuration for the first (blue dots) and last (black crosses) iterations, together with

Table 4.8: Distribution-based Quantity. Cauchy-measure Condition.

	Function	$\mu(K^{Q}_{Ter})$	$\mu(K^{Ref}_{Ter})$	*Termination*	*Proposition 2*
1	Esphere	100.94	73.7	0.01, $\surd$	$\surd$
2	Ellipsoidal	101.89	73.76	0.01, $\surd$	$\surd$
3	Rastrigin	114.75	87.68	0.01, $\surd$	$\surd$
4	Bunche-Rastrigin	118.33	88.62	0.01, $\surd$	$\surd$
6	Attractive-Sector	376.66	297.49	0.01, $\surd$	$\surd$
8	Rosenbrock	556.51	480.52	0.01, $\surd$	$\surd$
9	Rosenbrock-Rotated	528.25	446.4	0.01, $\surd$	$\surd$
10	Ellipsoidal-HC	204.37	146.88	0.01, $\surd$	$\surd$
11	Discus	388.83	304.53	0.01, $\surd$	$\surd$
12	Bent-Cigar	483.14	382.39	0.01, $\surd$	X
13	Sharp-Ridge	494.23	402.37	0.01, $\surd$	X
14	Different-Powers	499.72	395.3	0.01, $\surd$	X
15	Rastrigin-Multimodal	626.78	560.19	0.01, $\surd$	$\surd$
17	Schaffers	198.89	129.73	0.01, $\surd$	$\surd$
18	Schaffers-Conditioned	451.98	321.36	0.01, $\surd$	X
19	Griewank-Rosenbrock	1274.59	1096.43	0.01, $\surd$	X
21	Schwefel	120.73	82.85	0.01, $\surd$	$\surd$
21	Gallagher-101	301.23	230.71	0.01, $\surd$	X
22	Gallagher-21	294.96	228.9	0.01, $\surd$	$\surd$
24	Lunacek	386.6	328.79	0.01, $\surd$	X

the function optimum (red circle) for a 2D case. Each asymptotic behavior of EA population is better appreciated in the squared close-ups around function optima.

In the case of convergence to optimum (first row in figure 4.10), EA population converges to a steady state that gathers around the function optimum to finally coincide with it. For this cases there is an excellent fit of the regression model and all theory applies.

In the case of premature convergence (second row in figure 4.10), EA population reaches a steady state gathered around a point different from the function optimum. The functions Griewank-Rosenbrock, Schwefel, Gallagher21 and Lunacek belong to this group. Due to the premature convergence, the limit point of EA steady state might be achieved in a finite number of iterations. Before reaching such limit point, there is a perfect linear regression (see close-up of figure 4.10). However the moment EA has actually reached the limit point, quantities are prone to repeat their values. Such repetitions create a fake asymptote that distorts the linear pattern and, in fact, violate the uncorrelation assumption for errors required for the regression model. In order to apply our regression framework the final repeated values have to be removed. Since infinite values are excluded from the computation of regression models, repetitions in final iterations are discarded by applying the regression model to $Y = Ref - Ref(N_{it})$

Table 4.9: Improvement-based Quantity. Cauchy-measure Condition.

	Function	$\mu(K^{Q}_{Ter})$	$\mu(K^{Ref}_{Ter})$	*Termination*	*Proposition 2*
1	Esphere	45.74	73.7	1, X	X
2	Ellipsoidal	48.78	73.76	1, X	X
3	Rastrigin	80.21	87.68	1, X	X
4	Bunche-Rastrigin	94.49	88.62	0.01, $\surd$	$\surd$
6	Attractive-Sector	124.74	297.49	1, X	X
8	Rosenbrock	388.03	480.52	1, X	X
9	Rosenbrock-Rotated	331.87	446.4	1, X	X
10	Ellipsoidal-HC	76.25	146.88	1, X	X
11	Discus	226.66	304.53	1, X	X
12	Bent-Cigar	279.82	382.39	1, X	X
13	Sharp-Ridge	213.23	402.37	1, X	X
14	Different-Powers	79.01	395.3	1, X	X
15	Rastrigin-Multimodal	559.23	560.19	1, X	X
17	Schaffers	139.42	129.73	0.67, X	X
18	Schaffers-Conditioned	371.65	321.36	0.01, $\surd$	X
19	Griewank-Rosenbrock	755.96	1096.43	1, X	X
21	Schwefel	98.24	82.85	0.01	$\surd$
21	Gallagher-101	99.44	230.71	1, X	X
22	Gallagher-21	120.75	228.9	1, X	X
24	Lunacek	294.52	328.79	1, X	X

Table 4.10: Distributed-based Quantity. Applicability to Black Box Problems.

	Predictive Power			EA Actual Accuracy		
Function	$\mu(K^{Q}_{Ter})$	$\mu(K^{Ref}_{Ter})$	*Termination*	$\mu(D^{Q})$	$\mu(D^{Ref})$	p-val
Esphere	198.8	131.9	0.01, $\surd$	$2e-8$	$5e-7$	$9e-41$
Rosenbrock	456.1	314.1	0.01, $\surd$	$3e-8$	$4e-7$	$8e-28$
Ellipsoidal	200.9	132.7	0.01, $\surd$	$3e-8$	$5e-7$	$1e-29$
Schaffers	1601.4	868.3	0.01, $\surd$	$3e-8$	$4e-7$	$4e-22$
Gallagher101	364.6	287.2	0.01, $\surd$	$3e-8$	$5e-7$	$4e-34$

and $X = Q - Q(N_{it})$, for N_{it} the maximum number of EA iterations.

Finally, in the case of stagnation (third row in figure 4.10), EA population cannot reach a steady state and the iterative process does not converge to any point. The functions LinearSlope, StepEllipsoidal, Weierstrass and Katsuura belong to this group. These cases have to be dropped from the further analysis, given that they require stopping EA using other criteria different from reaching a steady state. First, we would like to note that these functions do not have a unique global minimum and, thus, require multimodal search EA paradigms, like niching [56]. Multimodal algorithms present a different asymptotic behavior (given by the multiple minima searched by the algorithm) and are out of the

Table 4.11: Improvement-based Quantity. Applicability to Black Box Problems.

	Predictive Power			EA Actual Accuracy		
Function	$\mu(K^{Q}_{Ter})$	$\mu(K^{Ref}_{Ter})$	*Termination*	$\mu(D^{Q})$	$\mu(D^{Ref})$	p-val
Esphere	57.1	131.9	1, X	$4e-5$	$5e-7$	1
Rosenbrock	123.2	314.8	1, X	$7e-5$	$4e-7$	1
Ellipsoidal	60.0	132.	1, X	$4e-5$	$5e-7$	1
Schaffers	439.04	868.3	1, X	$1e-5$	$4e-7$	1
Gallagher101	110.2	287.2	1, X	$1e-6$	$5e-7$	1

Table 4.12: Noiseless Functions Upper bound

	MaxDist		FitVal	
Function	$\epsilon = 10^{-6}$	$\epsilon = 10^{-9}$	$\epsilon = 10^{-6}$	$\epsilon = 10^{-9}$
Esphere	$6.01e-6$	$5.98e-9$	$1.00e-12$	$1.00e-18$
Rosenbrock	$1.07e-6$	$1.22e-9$	$3.89e-14$	$4.64e-20$
Ellipsoidal	$5.19e-6$	$5.13e-9$	$6.53e-13$	$6.48e-19$
Schaffers	$2.27e-6$	$1.59e-9$	$1.61e-7$	$1.49e-10$
Gallagher101	$1.39e-6$	$3.36e-10$	$1.56e-7$	$1.44e-10$

Table 4.13: Noisy Functions Upper bound

	MaxDist		FitVal	
Function	$\epsilon = 10^{-6}$	$\epsilon = 10^{-9}$	$\epsilon = 10^{-6}$	$\epsilon = 10^{-9}$
Esphere	$3.65e-6$	$3.32e-9$	$5.70e-14$	$5.52e-20$
Rosenbrock	$1.10e-6$	$1.04e-9$	$6.22e-15$	$3.59e-21$
Ellipsoidal	$2.76e-6$	$2.76-9$	$8.23e-14$	$8.06e-20$
Schaffers	$3.28e-7$	$2.45e-10$	$7.56e-9$	$4.68e-12$
Gallagher101	$1.82e-7$	$2.59e-11$	$9.88e-11$	$6.04e-14$

scope of the presented work. Second, divergent cases failing to reach a steady state are always problematic and often require stopping EA using a maximum number of iterations [72].

2. Applicability to Black-Box Problems

   The proposed evaluation mechanisms are given in statistical regression terms (Proposition 2) that are checked for known test functions. If the set of test functions cover enough features of potential landscapes, then statistical inference can be applied to extend results to unknown functions presenting similar properties. Knowledge about the properties of unknown optimization problems can be obtained using ELA [36] and, thus, suitability of a quantity validated with our method could be assessed in real settings. We would like to note that the chosen dataset already includes most landscapes described for optimization

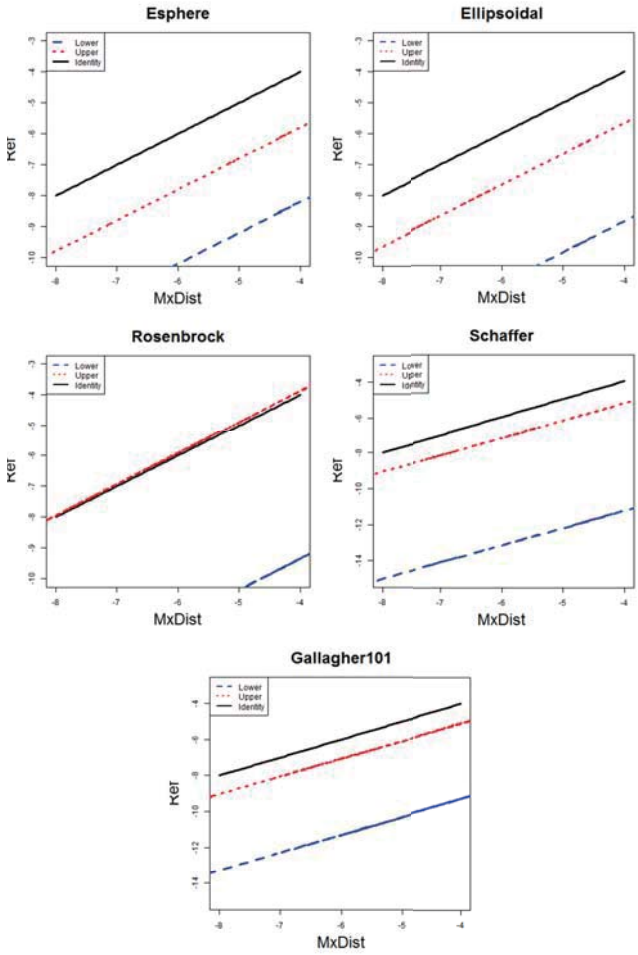


Figure 4.6: Confidence Intervals and Identity line for MxD noiseless functions in logarithmic scale

problems [36]. Thus, if a given quantity Q satisfies the Cauchy measure condition for the BBOB'09/'10 sets, this quantity will be likely to be suitable for stopping EA in practice using the range formula (3.8), provided that EA has reached a steady state. Further, our methodology can be applied to any data set of test functions, so results are easily reproducible for benchmarks more complete regarding the whole space of possible optimization tasks.

Once the quantity and its parameters have been chosen using the validation benchmark, these are the settings that should be used for terminating unknown problems without further verifications. We would like to point that whether

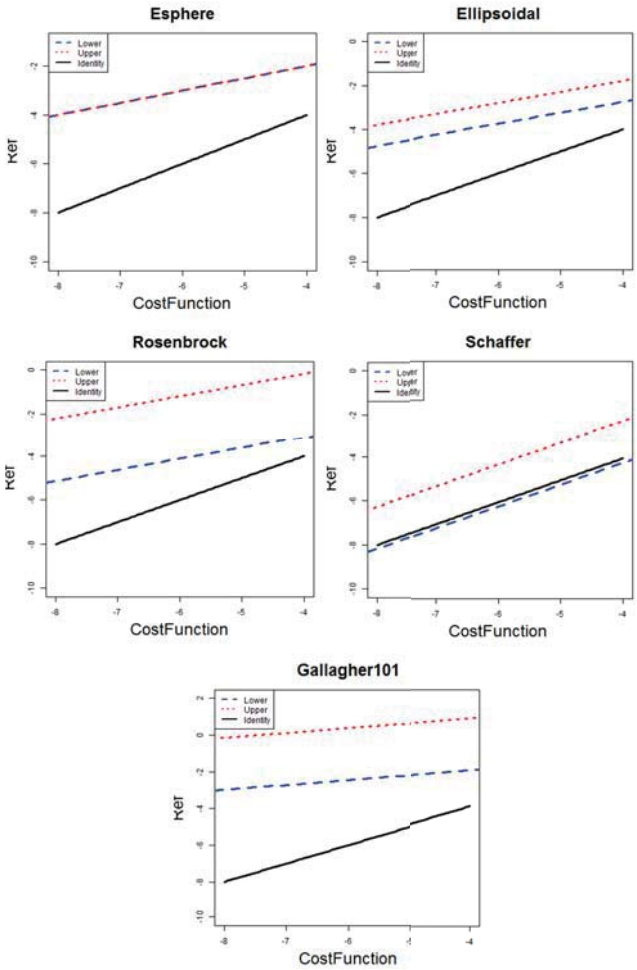


Figure 4.7: Confidence Intervals and Identity line for FitVal noiseless functions in logarithmic scale

EA should be restarted or has reached the optimum cannot be predicted by our evaluation mechanisms and, it is out of the scope of this paper. In fact, information about the EA state optimality is impossible when no gradient information is available. Given that this could be the usual case for EA optimization problems, whether EA has reached the optimum can only be answered for very specific algorithms and optimization problems [34]. Therefore, in practice, the most sensible stop criterion should ensure EA has reached a steady state rather than its convergence to the optimum [65].

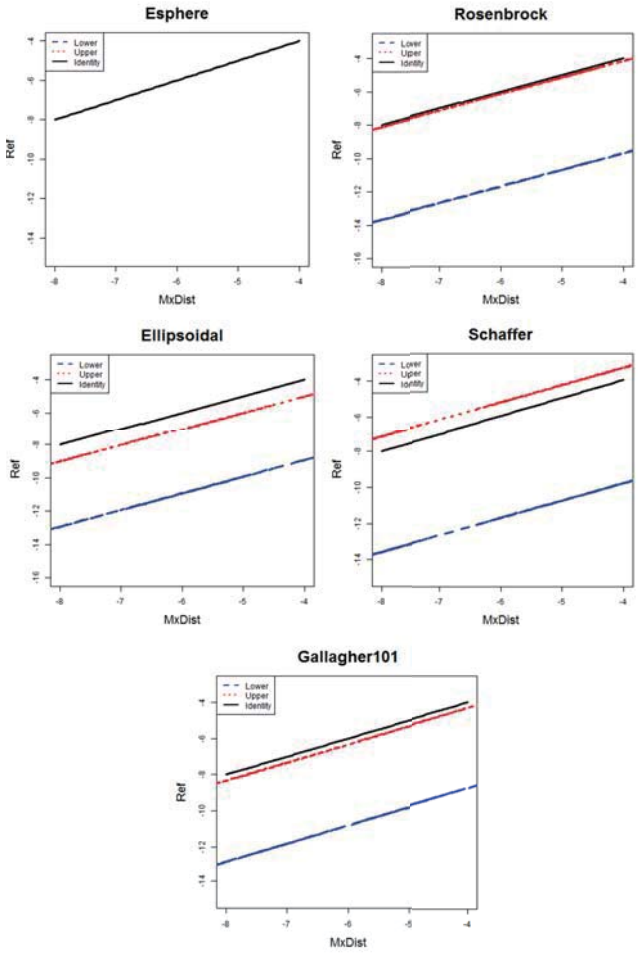


Figure 4.8: Confidence Intervals and Identity line for MxD noisy functions in logarithmic scale

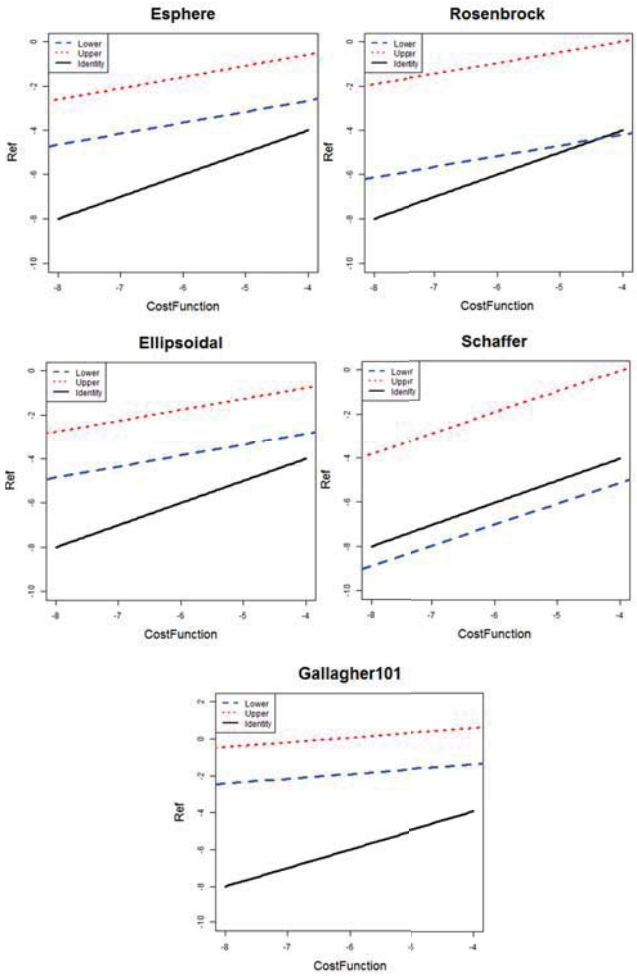


Figure 4.9: Confidence Intervals and Identity line for FitVal noisy functions in logarithmic scale

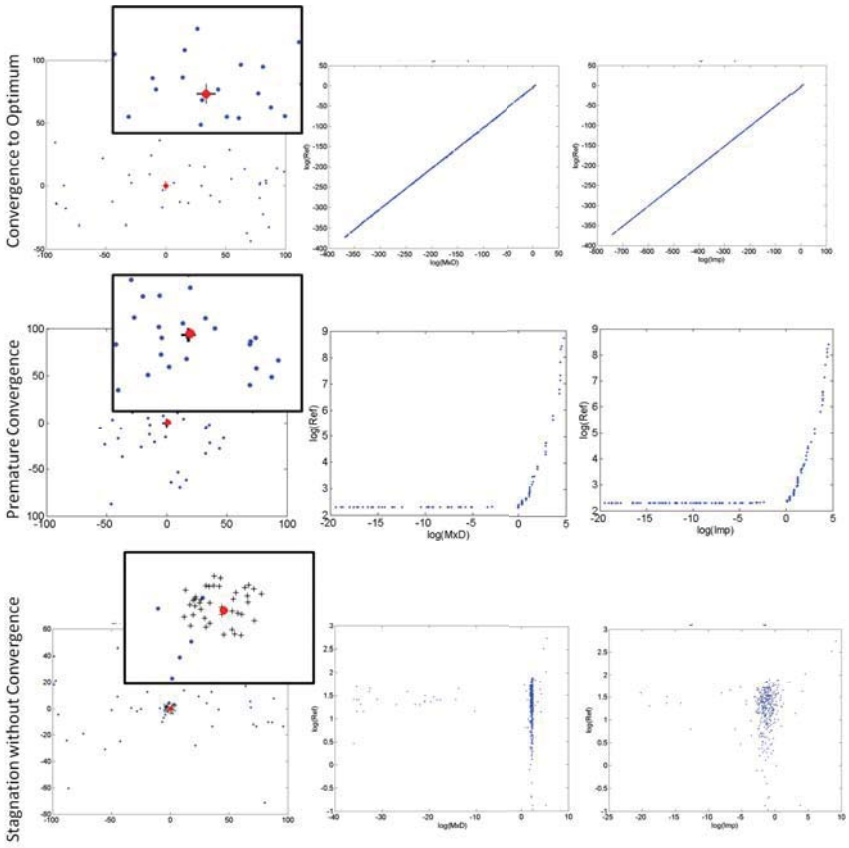


Figure 4.10: EA convergence asymptotic behaviors.

Chapter 5

Application to Pharmacology

5.1 Preliminaries

This thesis is aimed to develop and apply mechanistic models because they usually include many parameters in the equation with the consequent necessity of algorithmic applications for correct estimation. In particular we consider the Operational model of allosterism including receptor activity.

Knowledge of biological systems can be gained by examining the physiological effects that they produce under particular experimental conditions. These effects are the result of the transduction by the system of the signal embodied in the molecular structure of a drug, where for drug we mean, in a general sense, any substance able to perturb a biological system in a concentration-dependent fashion. Because the true transduction function is generally unknown an estimate must be obtained. These functional estimates, expressed as mathematical models, are useful for the characterization of the biological system and for the classification and discovery of new drugs. The perturbation that the drug exerts on the biological system can, in theory, be described by a true theoretical function $\phi(x)$ of all the parameters present in the system (figure 5.1). Because this function is unknown an $f(x)$ estimate must be obtained from $E/[A]$ experimental data. The $f(x)$ estimate can be used for both characterization of the biological system and classification and design of new drugs.

Yet what we understand by a mathematical model, particularly in pharmacology? A mathematical model of pharmacological effect is a mathematical equation $E = f([A])$, in which E represents the pharmacological effect, $[A]$ is the concentration of the drug and f is a mathematical function containing a number of parameters. It is the nature of the parameters what defines the approach we are using for modelling the experimental data. We distinguish between two types of mathematical models: (i) empirical, if the parameters lack physical meaning and (ii) mechanistic, if the parameters embody part of the biological information the function is intended to represent [26].

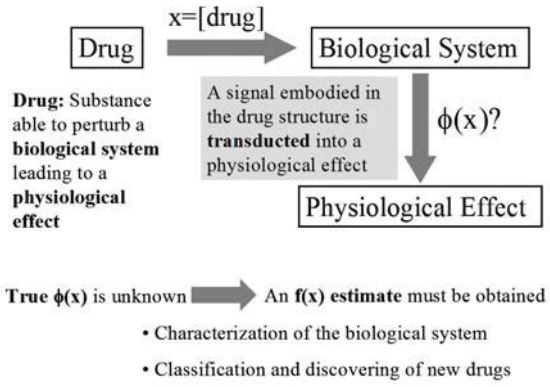


Figure 5.1: Schematic representation of the drug action process.

Empirical models. Customarily, drugs and receptors are compared and classified by monitoring the pharmacological effect the receptor produces with increasing $[A]$. $E/[A]$ data are commonly depicted in a semi-logarithmic scale, E/x with $x = log[A]$, typically leading to sigmoid-shaped plots. To properly compare the experimental scatter plots, $E = f(x)$ functions are fitted to the data points and their parameters estimated. The parameter estimates allow the quantification of the geometric characteristics of the E/x curves. If, for simplicity, we assume that basal response is absent ($E = 0$ for $[A] = 0$) then the shapes of E/x curves can be characterized by four quantities, the upper asymptote (maximum response), the mid-point (curve location), the mid-point slope (steepness) and the inflection point (symmetry of the curve) . Each of these properties can be mathematically defined and pharmacologically interpreted.

The upper asymptote, top(figure 5.2 **A**), reflects the efficacy of the agonist-receptor system and is defined as the value towards the effect tends as [A] increases, $E \xrightarrow[x\to\infty]{} Top$. The mid-point(5.2 **B**), x_{50}, measures the agonist potency and is defined as the x for half the top. The mid-point slope (5.2 **C**)is the value of the slope of the E/x curve at the mid-point, $(\frac{dE}{dx})_{x=x_{50}}$, and displays the sensitivity of the system to small changes in agonist concentration. Rectangular hyperbolic curves give a typical mid-point slope of 0.576 when they are normalized (the derivative is divided by top) while non-hyperbolic curves can be steep ($(\frac{dE}{dx})_{x=x_{50}} > 0.576$) or flat ($(\frac{dE}{dx})_{x=x_{50}} < 0.576$) . The point of inflection ((5.2 **C**)),X_I , is a point on a curve at which the curvature changes from convex to concave or vice versa. For an E/x curve, this is a point at which the first derivative of the function is a maximum whereas the second derivative is equal to zero. Importantly, the location of the point of inflection serves for the assessment of the symmetry of the curve. An E/x curve is symmetric if the point of inflection matches the mid-point, $X_I = X_50$, and asymmetric if it does not,$X_I \neq X_50$. Geometric characterization of the experimental E/x plots is what empirical models are asked for. The above commented geometric features, upper asymptote, location, steepness and symmetry, are obtained after fitting data with a mathematical model. There are a number of $E = f(x)$ functions for data fitting but, with no doubt, the

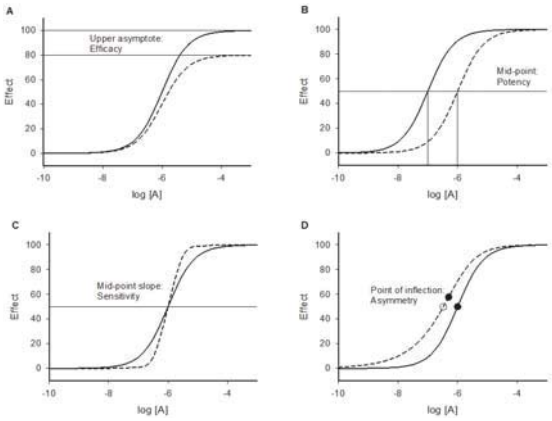


Figure 5.2: Geometric parameters characterizing $E/[A]$ curves. The data are represented using logarithm values for the X axis. The effect for $[A] = 0$ is 0. The solid curve is symmetric. The dashed curve is asymmetric(open and solid circles are used for the mid-point and the inflection point, respectively).

mostly used in pharmacology is the Hill equation.

Hill equation: a model for symmetric concentration-effect curves.

The Hill equation [26] is the three-parameter (5.1).

$$E = \frac{a}{1 + 10^{m(x_b - x}} \tag{5.1}$$

where $x = Log[A]$ and $m > 0$, being m the Hill coefficient.

The upper asymptote is a, the mid-point is $x_50 = x_b$, the mid-point slope is $\frac{dE}{dx})_{x=x_{50}} = \frac{amln10}{4} = 0.576am$ (notice that for m=1, rectangular hyperbola, the mid-point slope is equal to 0.576 after normalization by dividing by a) and the point of inflection is $x_I = x_b$. Because there is an identity between the inflection point and the mid-point, the Hill equation produces symmetric curves in all cases and, therefore, is not appropriate for fitting asymmetric E/x data.

But what to do with asymmetric concentration-effect curves?. There are several empirical models capable of dealing with asymmetric E/x data, namely the Richards model, the Gompertz model and the modified Hill equation [19].

The Richards model The Richards model [46] is a generalization of the Hill equation by including an additional parameter (5.2).

$$E = \frac{a}{1 + 10^{m(x_b - x)^s}} \tag{5.2}$$

with $s > 0$ The upper asymptote is a, the mid-point is $x_{50} = x_b - \frac{1}{m} log(2^{frac1s} - 1)$

, the mid-point slope is $(\frac{dE}{dx})_{x=x_{50}} = \frac{amln(10)s(1-12^{\frac{1}{s}})}{2}$, and the point of inflection is $x_I = x_b + \frac{1}{m}log(s)$. The new parameter, s, allows for asymmetry. If $s = 1$, 5.2 is equivalent to 5.1 and the theoretical curve is symmetric. Consistently with this feature, we see that if $s = 1$ then $x_I = x_50 = x_b$. However, if $sneq1$ then x_Ineqx_50 , and the theoretical curve is asymmetric. Interestingly, for $sneq1$, the degree of asymmetry of the curve, measured as the difference between x_I and x_50, relies on both s and m parameters, and $x_I - x_50 = \frac{1}{m}log(s(2^{\frac{1}{s}} - 1))$. If $s > 1$ then $x_I < x_50$, the point of inflection is located before the mid-point, whereas if If $s < 1$ then $x_I > x_50$, the point of inflection is placed after the mid-point. We see that for a given s value the degree of asymmetry decreases as the parameter m increases. In addition, it has been shown [26] that the degree of asymmetry is higher for Richards equations with $s < 1$ than for those equations with $s > 1$. The Richards equation may detect and quantify the asymmetry present in E/x data. However, the correlation between parameters, that inclusion of parameter s brings in, poses additional difficulties in data fitting [63]. The Richards equation is a typical example of an overparameterized model (see Appendix for a discussion on data fitting). To account for asymmetry without increasing the number of parameters of the Hill equation led to the proposal of two new functions, the Gompertz model and the modified Hill equation.

Connecting empirical and mechanistic models The asymmetry of the curves and the Hill coefficient. In an earlier work [18] a connection between empirical and mechanistic models was shown by using a mechanistic model taken from the ion channel field, a ligand-gated ion channel with 4 binding sites. Nevertheless, the model can be applied to a general tetrameric receptor R that is activated after all the binding sites are occupied. By defining the effect as the proportion of receptors in the open/active state, the following expression was obtained

where $K1, K2, K3$ and $K4$ are the microscopic equilibrium dissociation constants and $K_E = \frac{[A_4R*]}{[A_4R]}$ is the equilibrium constant for the opening/activation (ion channel/receptor) reaction. By defining the effect as the proportion of receptors in the open/active state, the following expression was obtained

$$E = \frac{A^4 K_E}{K_1K_2K_3K_4 + 4K_2K_3K_4[A] + 6K_3K_4[A]^2 + 4K_4[A]^3 + [A]^4(1 + K_E)} \tag{5.3}$$

By supposing that the efficacy is very low ($KE << 1$) and no cooperativity between the binding sites ($K_i = K$), the previous mechanistic equation simplifies to

$$E = \frac{K_E}{(1 + \frac{K}{[A]})^4} = \frac{K_E}{(1 + 10^{logK-x})^4} \tag{5.4}$$

where $x = log[A]$. The latter equation corresponds to a Richards model 5.2(with $a = K_E, m = 1, x_b = logKands = 4$). Thus, we see that empirical models (phenotype) may reflect some of the features that characterize mechanistic models (genotype). In addition, this value of s is indicative of asymmetry and may represent an example

of the necessity of expanding the set of empirical models and consider that in some cases other models apart from the symmetric Hill model are needed.We have shown an example in which a mechanistic model may be expressed as an empirical model under particular mechanistic features. It can be hypothesized that if a systematic analysis of mechanisms were done in a particular biological/pharmacological research area and the corresponding set of empirical equations were identified then we could proceed in the inverse order and try to propose some mechanistic conditions from the application of one or other empirical model. For instance, in the case of GPCRs asymmetric curves are found when total receptor and total G protein concentrations are not negligible one relative to the other . This indicates that the stoichiometry of the biological species in a proposed mechanistic GPCR system must be consistent with the symmetry of the experimental curves. Likewise many other molecular properties associated to GPCR function will be commented from the models presented on the next section.

Mechanistic models We have seen that there is a relationship between the mechanism underlying experimental data and the shape of the curves they produce. Empirical models may be powerful enough to reveal that "something" at the biological level is happening if this affects any of the geometric characteristics of the curves. Yet to properly analyze mechanistic hypotheses mechanistic models need to be used.

5.2 The Operational Model of agonism including constitutive receptor activity

To allow the operational model of agonism to account for constitutive receptor activity, an extension of the model has been recently proposed in [59]. Constitutive receptor activity was included in the model by defining a stimulus, S, as $S = [R] + \epsilon[AR]$, which is connected with the observed effect by the rectangular hyperbolic function $\frac{E}{Em} = \frac{S}{KE+S}$ (figure 5.3).

1. **The equilibrium constant of the model**

$$A + R \xrightarrow{K} AR \tag{5.5}$$

2. **The fractional observed effect** A receptor stimulus is defined which includes the concentration of both the free and the ligand-bound receptors.

$$S = [R] + \epsilon[AR] \tag{5.6}$$

A rectangular hyperbolic equation is proposed for the relationship between the receptor stimulus and the observed effect.

$$f = \frac{E_m S}{K_E + S} = \frac{K\aleph + \epsilon\aleph[A]}{K(1+\aleph) + (1+\epsilon\aleph)[A]} \tag{5.7}$$

with $\aleph = \frac{[R_T]}{K_E}$ and $[R_T] = [R] + [AR]$

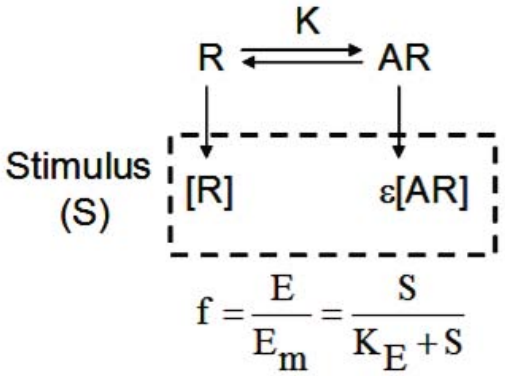


Figure 5.3: The Operational Model of Agonism Including Constitutive Receptor Activity

3. **The Geometric Descriptors of the Curves**

   (a) **Left asymptote** Basal response for $[A] = 0$, $Basal = \frac{E_m}{1+\frac{1}{\aleph}}$

   (b) **Right asymptote**. The asymptotic f-values when $[A]$ increases. $Top = \frac{E_m}{1+\frac{1}{\epsilon\aleph}}$

   (c) **The midpoint**. The $[A]$ value for half maximum effect $[A_{50}] = \frac{K(1+\aleph)}{1+\epsilon\aleph}$. $[A_{50}]$ is lower, equal and greater than K for agonists ($\epsilon > 1$), neutral agonists ($\epsilon = 1$) and invers agonists ($\epsilon < 1$) respectively.

Equation 5.7 contains 3 parameters, $\aleph = \frac{[R_T]}{K_E}$ (a parameter determining the capacity of the free receptors of generating the (basal) response), K (the equilibrium dissociation constant of the ligand for the receptor) and ϵ (a parameter measuring the different capacity of the ligand-bound receptor of generating a stimulus relative to the free receptor). Values of greater, equal and lower than one lead to asymptotic top values greater, equal and lower than basal response, which are the effects found for agonists, neutral antagonists and inverse agonists, respectively.

5.2.1 The Operational Model of allosterism

. The operational model of allosterism [15, 27, 28, 30, 38] was constructed as an extension of the operational model of agonism [10] by including a second binding site, that for the allosteric modulator. This leads to three independent dissociation constants, two for the binding of the agonist A and the allosteric modulator B to the free receptor (K and M, respectively) and a third one involving the doubly occupied receptor. The dissociation constant regulating the latter equilibrium can be expressed

in terms of K or M, depending on which is the complexed receptor to which the second ligand binds and a cooperativity binding factor α. Following the rationale [30] a total stimulus S is defined as the sum of the stimulus of all the ligand-bound receptor species, $S = \epsilon_A[AR] + \epsilon_B[RB] + \epsilon_{AB}[ARB]$. The connection between stimulus and fractional effect is made by the rectangular hyperbolic function $\frac{E}{Em} = \frac{S}{(KE+S)}$.

5.8 embodies the concentration-fractional effect relationship for the operational model of allosterism. Apart from the K and M dissociation constants and α binding cooperativity, the equation includes the operational efficacies $\tau_A = \frac{\epsilon_A[R]_T}{K_E}$, $\tau_B = \frac{\epsilon_B[R]_T}{K_E}$ and $\tau_{AB} = \frac{\epsilon_{AB}[R]_T}{K_E} = \beta\frac{\epsilon_A[R]_T}{K_E} = \beta\tau_A$. As in the operational model of agonism, τ includes both molecular and tissue components. It is worth noting that ϵ_{AB} was defined as $\epsilon_{AB} = \epsilon_A\beta$ another possibility being $\epsilon_{AB} = \epsilon_A\epsilon_B\delta$. The chosen parameters have a different meaning, β describes the ability of B to alter the signalling capacity of A (asymmetric interaction: its value would be different if it had been defined as the ability of A to alter the signalling capacity of B, that is $\epsilon_{AB} = \epsilon_B\beta$) whereas δ describes the activation cooperativity between A and B in the ARB complex (symmetric interaction: it does not measure the influence of one compound on the other but their mutual effects on each other). However, only the former ($\epsilon_{AB} = \epsilon_A\beta$) definition is satisfactory in this modelling approach in terms of curve fitting as the latter ($\epsilon_{AB} = \epsilon_A\epsilon_B\delta$) leads to an additional parameter in the final equation, after intrinsic efficacies are combined with system parameters into operational (τ) efficacies.

$$E = \frac{E}{E_m} = \frac{K\tau_B[B] + \tau_A(M + \alpha\beta[B])[A]}{K(M + (1 + \tau_B)[B]) + (M(1 + \tau_A) + \alpha(1 + \tau_A\beta)[B])[A]} \tag{5.8}$$

As constitutive receptor activity is not included in the model, the effect for $[A]$ and $[B]$ equal to 0 (basal response) is 0. To account for this issue without changing the mechanistic nature of the model, basal response has been included as an ad hoc parameter as it was also done in the operational model of agonism (5.9) [58]. However, as it also happened with the operational model of agonism, effects lower than the basal response are outside the scope of the model.

$$E = Basal + \frac{(E_m - Basal)(K\tau_B[B] + \tau_A(M + \alpha\beta[B])[A])}{K(M + (1 + \tau_B)[B]) + (M(1 + \tau_A) + (\alpha(1 + \tau_A\beta)[B])[A]} \tag{5.9}$$

It is worth noting that constitutive receptor activity was not incorporated into the model because as the authors stated [30] the resultant number of parameters makes the model impractical for fitting to experimental data. Interestingly, the operational model of agonism including constitutive receptor activity (vide supra) has been applied to describe experimental data, the effects of agonists at CC-chemokine receptor 4, with overall satisfactory fitting parameters [59]. These positive results encouraged us to incorporate constitutive receptor activity into the operational model of allosterism. This would allow us to model the behaviour of allosteric modulators with intrinsic negative agonist efficacy (inverse agonists).

5.2.2 The operational model of allosterism including constitutive receptor activity

. This is the model for our Evolutionary Algorithm application. To incorporate constitutive receptor activity in the operational model of allosterism we followed a rationale similar to that taken by Slack and Hall in the operational model of agonism [59].

1. The equilibrium constant of the model

$$A + R \xrightarrow{K} AR; K = \frac{[A][R]}{[AR]} \tag{5.10}$$

$$B + R \xrightarrow{K} RB; M = \frac{[B][R]}{[RB]} \tag{5.11}$$

$$AR + B \xrightarrow{\frac{M}{\alpha}} ARB; \alpha = \frac{M[ARB]}{[B][AR]} \tag{5.12}$$

$$RB + A \xrightarrow{\frac{K}{\alpha}} ARB; \alpha = \frac{K[ARB]}{[A][RB]} \tag{5.13}$$

2. The fractional observed effect the concentration of free receptors was included in the definition of the total stimulus S, $S = [R] + \epsilon_A[AR] + \epsilon_B[RB] + \epsilon_{AB}[ARB]$, which, in turn, was connected with the observed effect by the rectangular hyperbolic function $\frac{E}{Em} = \frac{S}{KE+S}$ (5.4).

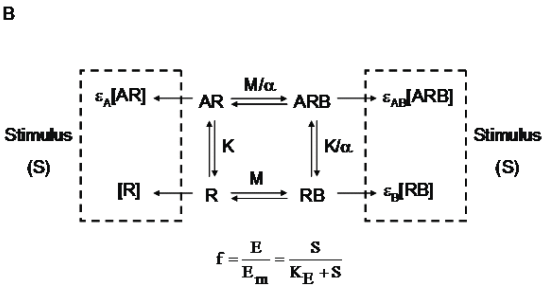


Figure 5.4: The operational model of allosterism including constitutive receptor activity.

Equation 5.14 embodies the concentration-fractional effect relationship for the operational model of allosterism including constitutive receptor activity. It is worth mentioning that the equation is an alternative derivation of one previously obtained in [22]. By comparing 5.8 and 5.14, we see that including constitutive receptor activity leads to a model in which intrinsic efficacies (ϵs) are included instead of operational efficacies (τs) and a parameter ($\aleph = \frac{[R]_t}{K_E}$) to account for basal response appears. Ligands A and B are agonists, neutral antagonists or

inverse agonists if they increase, not change or decrease the basal response (ϵ_A or ϵ_B greater, equal or lower than 1, respectively). Finally, for proper comparison with the allosteric two-state model, the intrinsic efficacy of ARB is defined as $\epsilon_{AB} = \epsilon_A \epsilon_B \delta$, with δ measuring the activation cooperativity between A and B in the ARB complex.

$$f = \frac{E}{E_m} = \frac{K\aleph(M + \epsilon_B[B]) + \epsilon_A\aleph(M + \epsilon_B\alpha\delta[B])[A]}{K(M(1+\aleph) + (1+\epsilon_B\aleph)[B]) + (M(1+\epsilon_A\aleph) + \alpha(1+\epsilon_A\epsilon_B\delta\aleph)[B])[A]} \tag{5.14}$$

The operational model of allosterism including constitutive receptor activity contains four chemical equilibria, corresponding to the binding of the agonist or the allosteric modulator to the free receptor or to a receptor occupied by the other compound.

3. **Geometric descriptors of the curves**

   (a) Left asymptote in absence of A and B Basal: response f for $[A] = 0$ and $[B] = 0$. $Basal = \frac{E_m}{1+\frac{1}{\aleph}}$

   (b) Left asymptote in absence of A Bottom: f for $[A] = 0$. $Bottom = \frac{\aleph(M+\epsilon_B[B])}{M(1+\epsilon)+[B](1+\epsilon_B\aleph)}$

   (c) Right asymptote, the asymptote f-value as $[A]$ increases ($f \xrightarrow[[A}{\rightarrow} \infty] Top$)
   $Top = \frac{\aleph\epsilon_A(M+\epsilon_B\alpha\delta[B])}{M(1+\epsilon_A\aleph)+\alpha(1+\epsilon_A\epsilon_B\delta\aleph)[B]}$

   (d) The mid-point. The [A] values for half maximum effect. $[A]_{50} = \frac{K(\aleph(M+\epsilon_B[B])+M+[B])}{\aleph\epsilon_A(M+\epsilon_B\delta\alpha[B])+M+\alpha[B]}$

In 5.1 the pharmacological properties of the parameters of the operational model of allosterism is presented.

Pharmacological Property	**Parameter**
Constitutive receptor activity (basal response)	$\aleph$
Agonist Concentration	$[A]$
Allosteric modulator concentration	$[B]$
Agonist dissociation constant	K
Allosteric modulator dissociation constant	M
Intrinsic efficacy of A	ϵ_A
Intrinsic efficacy of B	ϵ_B
Binding cooperativity between A and B	α
Activation cooperativity between A and B	δ

Table 5.1: Operational model of allosterism parameters with their pharmacological properties.

5.3 Theoretical Simulation

Over-parameterization is a characteristic of operational models. It has been shown that for this kind of models is not possible to directly fit a single experimental $E/[A]$ curve if Em, the maximum effect of the system, is unknown [59, 33, 50]. Because of the correlation between parameters, to obtain reliable parameter estimates at least two $E/[A]$ curves with differentiable asymptotic maximum responses are needed and this can be done by using the receptor inactivation method [17]. Decreasing the number of available receptors has the effect of lowering the $\aleph$ parameter in the operational model of allosterism including constitutive receptor activity. This leads to concentration-effect curves with asymptotic maximum values lower than that of the control curve.

To exemplify this case we will assume that we are working with the same input data as in the previous case (the same tissue, the same receptor and the same agonist) but in the presence of an AM at a defined (10^{-4}) concentration 5.5.

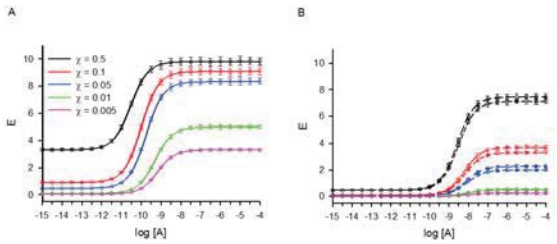


Figure 5.5: Simulated experiment in which curves in figure 5A (the allosteric modulator B is absent) are the control curves for figure 5B (the allosteric modulator B is present)

On the fitting procedure the parameter values estimated in the previous case (E_m, $\aleph$, ϵ_A, K) were assumed known and kept fixed. Figure 5.5 illustrates a simulated experiment in which curves in figure 5A (the allosteric modulator B is absent) are the control curves for figure 5B (the allosteric modulator B is present). The whole set of curves of the experiment is separated into two (figure 5A and figure 5B) for clarity. **A**. Monte Carlo data (mean $\pm$ SD) and curve fitting (solid lines) under the operational model of agonism with constitutive receptor activity. 50 sets each composed of 5 $\aleph$-varied ($\aleph = 0.5, 0.1, 0.05, 0.01 and 0.005$) curves were generated for an agonist-receptor system, in which $Em = 10$, $K = 10^{-9}$ and $\epsilon = 10^2$, by assuming that responses follow Equation 4 under a normal distribution with mean equal to the former theoretical values and standard deviation equal to $3 percent$ of the mean for $log[A]$ ranging between -15 and -4. For curve fitting a hybrid approach between a global Evolutionary Computation method and a local gradient-based nonlinear approach was used (see Table 2). **B**. Monte Carlo data generation and curve fitting under the operational model of allosterism with constitutive receptor activity. As in figure 5A, 50 sets each composed of 5 $\aleph$-varied ($\aleph = 0.5, 0.1, 0.05, 0.01 and 0.005$) curves were generated for an agonist-receptor system with parameters equal to those in figure 5A ($Em = 10$, $K = 10^{-9}$ and $\epsilon = 10^2$) in the presence of an allosteric modulator

with $M = 10^{-9}$ and $\epsilon_\beta = 10^{-1}$ that interacts with the agonist as determined by $\alpha = 10^{-1}$ and $\delta = 0.5$, at two fixed concentrations, $[B] = 10^{-6}$ and $[B] = 10^{-4}$. Monte Carlo response data were generated as in figure 5A using Equation 8 and the defined system, agonist and modulator parameters. Curve fitting was performed as in figure 5A keeping fixed the parameter estimates obtained for the agonist when acting alone (see Table 3): long-dashed lines $[B] = 10^{-6}$ and short-dashed lines $[B] = 10^{-4}$, colour code as in figure 5A.

As it can be seen in Table 5.2 the hybrid DE-NLR approach produced excellent results. Comparison between figures 5.5 A and B shows the effect of compound B as a NAM with negative intrinsic efficacy and negative cooperativities both for binding and activation: left and right asymptotes are lowered and a right-shift displacement of the curves is observed. System- (Em and $\aleph$) and agonist-dependent (K, ϵ_A) parameters were taken from parameter estimates in Table 2 and kept fixed (Curves from Table 2 (figure 5A) represent the control curves from those from Table 3 (figure 5B)). Two concentrations (10^{-6} and 10^{-4}) were used for the allosteric modulator. Details for Monte Carlo simulated data generation and curve fitting as described at the bottom of Table 2. All parameters were assumed log-normally distributed. The parameter space for exploration with DE was defined as: $log(M)$, $(-15, -3)$; $log(\epsilon_B)$, $(-4, 0)$; $log(\alpha)$, $(-4, 4)$; and $log(\delta)$, $(-4, 4)$.

Parameter	**Theoretical values**	**Parameter estimates(mean $\pm$ SD)**
$Log(M)$	-9	-9.01 ± 0.05
$Log(\epsilon_B)$	-1	-1.00 ± 0.01
$Log(\alpha)$	-1	-1.00 ± 0.02
$Log(\delta$	-0.30	-0.30 ± 0.01

Table 5.2: Results for parameter estimation

Chapter 6

Conclusions and Future Work

A main concern in EAs is determining a termination condition ensuring EA has reached its steady state so that useless iterations are not performed. We have introduced a regression framework for assessing the capability of existing stopping conditions for terminating EA at its steady state. Our framework relies on a statistical inference regression model, which guarantees that the conclusions derived from any particular experiments are generalizable with a given confidence. We have also provided with an empirical termination condition based on the rate of convergence of quantities. The framework has been applied to DE paradigm and two quantities, one defined in f-space and the other defined in x-space.

We would like to highlight some interesting conclusions that arise from the analysis of our experiments.

1. **Paradigms suitable for loss of diversity termination**

   Termination of EA in black-box applications is still an open issue. This work presents a formulation of EA steady state inspired in the concept of Cauchy sequences and a practical stop criterion based on an implementation of this concept through a 2-parameter criterion together with statistical tools for adjusting parameters optimal values.

   Our experiments concluded that loss of population diversity seems to be a good candidate for PSO and DE (at least for low dimensions), but might not be the most appropriate one for CMA. This result was not unexpected and shows that the tools presented constitute an appealing basis for the definition of a general framework for EA termination criteria analysis.

   Our criterion can be used to detect loss of diversity for EA termination. We have applied our method to 3 EA paradigms (DE, PSO and CMA) and 5 representative test functions up to dimension 10. Experiments show that the tools presented constitute an appealing basis for the definition of a general framework for EA termination criteria analysis. However, more research is needed in order

to fully validate our framework as a solid methodology for the implementation of termination strategies.

2. **Selection of a Stopping Condition for DE Paradigm**

   We have applied our framework for quality assessment to two generic stopping conditions, one based on function improvement in f-space, *Imp*, and another one based on EA population distribution in x-space, *MxD*. These quantities have been used to terminate a DE paradigm for benchmark test functions [23]. Two experiments concerning the quality of the termination conditions are presented: capability for detection of DE steady-state and applicability to black-box problems.

   Our first experiment shows that the proportion test is rejected whenever Proposition 2 is satisfied, which validates our theoretical regression framework. This experiment also selects *MxD* as the better candidate for terminating than *Imp*, given that it satisfies the proportion test in 95% of the functions and Proposition 2 regardless of the function profile. Our second experiment shows the necessity of the statistical inference in order to guarantee reproducibility of results in black-box applications. As predicted, *MxD* terminates DE at its steady state in the presence of noise, while *Imp* fails to produce the expected results (the p-value for Schaffers worsens). Besides, *MxD* always bounds DE accuracy, which is one of the theoretical properties of Cauchy measures.

   Therefore, we conclude that measures based on EA population distribution in x-space are well-suited for terminating DE at its steady state. This agrees with theoretical studies on DE properties which relate its convergence to the steady state with a lost in diversity [68, 70] and, thus, it reinforces the usefulness of our framework for selecting termination conditions.

   Our framework strongly depends on the definition of EA steady state. In this paper we have restricted to single objective EA paradigms reaching a steady state. However, the results obtained encourage extending our definitions of steady state (such as the ones arising in multi-objective problems) and EA paradigms to select quantities that guarantee that EA has reached some sort of steady state and their relation to the convergence rate to the steady state.

3. **Applicability of the framework to Pharmacology**

6.1 Future Work

Termination of EA in black-box applications is still an open issue. This work presents a 2-parameter criterion together with statistical tools for adjusting parameters optimal values. Experiments show that the tools presented constitute an appealing basis for the definition of a general framework for EA termination criteria analysis. However, more research is needed in order to fully validate our framework as a solid methodology for the implementation of termination strategies. The results obtained on EA paradigms depending on loss of diversity (DE in particular) indicate the strength

of our statistical approach to fully determine its flexibility and capabilities for EA termination. Although this preliminary work constitutes a new effort in the use of statistical analysis as a tool for termination of EA algorithms and illustrates their potential for analyzing the behavior of EA algorithms, several issues should be further investigated:

1. **Scalability to high dimensions**

   A limitation of this study is that we have only considered 5 functions up to dimension 10 with a number of EA runs that fall shortly to achieve steady states for some cases. We are aware that to fully generalize results more functions and iterations should be considered. This is a matter of computational resources, time and an efficient parallel implementation and it is our top issue in our to-do list.

2. **Dependency on parameters**

   Another interesting topic to be further investigated is the variability of the number of generations n_{st} under different configurations in functions, as well as, the dependency on other parameters involved in the computation of the termination condition. In the first case, it would be of interest to determine the variability of generations across landscapes and dimensions, in order to check if the diversity criterion is still useful. Although a preliminary study [48] suggests that the parameters involved in the computation of the quantity selected for measuring diversity (in particular the percentage of best individuals used for computing MxD) is not a critical issue, at least for DE, its influence should be further explored. Finally, the impact that the accuracy, ϵ_{st}, required by the application has on the number of generations, and, thus, EA executions requires a deeper study, especially for high dimensions. Variability under different conditions, can be assessed with ANOVA test using functions and dimensions properties for inter-group variability. In the case of absence of normality and/or homoscedasticity a non-parametric test would be used.

3. **EA paradigms internal mechanisms for ensuring convergence** The question of how to select the quantities depending on the behavior of the algorithm is at the very core of the methods used as termination criteria. This question is not addressed in this work, but the proposed framework allows exploring which quantities are better related to EA internal mechanisms.

   In this context, we think that a quantity is suitable for EA termination (under the hypothesis that EA converges to a steady state) if the parameters of the range formula do not change across dimensionality of the search space. This hypothesis is supported by the results obtained for CMA in the experiments of Chapter3. We recall that such experiments indicated that loss of diversity was not the best suited quantity for CMA termination as the stopping parameters were not stable across function dimensionality. Therefore a list of methods and quantities that could be used in order to apply our framework will be further investigated.

4. **Further analyze non-convergent cases** As the results show, there is a relation between the convergence cases and the R^2 statistics. Practically all the convergent cases present a high R^2 value and, thus, a goodness of fit. Consequently, for the non-convergent cases, a good convergence version of the paradigms exploration must be done to ensure the convergence and check if the linear relation between measures and the R^2 statistics improve.

5. **Adaptation to a Multi-objective scheme** Practically all real-world optimization problems are multi-objective. The study and the applications of those schemas are increasing rapidly. An adaptation or modification of our framework is needed to extend all their mechanisms and utilities to the multi-objective situation. Perhaps, new statistical tools and concepts will be able to lead with that new and important paradigms.

Publications

Journals

- Debora Gil,**David Roche**, Agnes Borras,Jesús Giraldo. Terminating Evolutionary Algorithms Ensuring Accuracy. *Computational Optimization and Applications*, Springer US.
- **David Roche**, Debora. Gil, Jesús Giraldo. Mechanistic analysis of the function of agonists and allosteric modulators: Reconciling two-state and operational models. *British Journal of Pharmacology*, 2012-BJP-1042-RU.R1.
- **David Roche**, Jesús Giraldo, Debora Gil. Multiple active receptor conformation, agonist efficacy and maximum effect of the system: The conformationally-based operational model of agonism. *Drug Discovery Today*, DRUDIS1129.
- X Rovira*, M Viv*, J Serra, **D. Roche**, PG Strange and J Giraldo. Modelling the interdependence between the stoichiometry of receptor oligomerization and ligand binding for a coexisting dimer/tetramer receptor system. *British Journal of Pharmacology*,2009
- Juan A.L. Garcia*, Frederic Bartumeus,**David Roche**, Jesús Giraldo,H. Eugene Stanley , Emilio O. Casamayor. Ecophysiological significance of scale-dependent patterns in prokaryotic genomes unveiled by a combination of statistic and genometric analyses. *Genomics*,2008
- Xavier Rovira*, **David Roche**, Juan Serra, Julie Kniazeff, Jean-Philippe Pin, and Jesús Giraldo. Modelling the Binding and Function of Metabotropic Glutamate Receptors. *The Journal of Pharmacology and Experimental Therapeutics*,2008
- Jesús Giraldo*, Juan Serra, **David Roche** and Xavier Rovira. Assessing Receptor Affinity for Inverse Agonists: Schild and Cheng-Prusoff Methods Revisited. *Current Drug Targets*,2007
- Jesús Giraldo*, **David Roche**, Xavier Rovira, Juan Serra. The catalytic power of enzymes: Conformational selection or transition state stabilization?. *Febs letters*,2006, Volume 580 , Issue 9 , 2170 - 2177

Book Chapters

- **David Roche**, Debora Gil,Jesús Giraldo. Mathematical modeling of G protein-coupled receptor function: What can we learn from empirical and mechanistic models?. *Computational Optimization and Applications*, (Springer).

Congress Contributions

- **David Roche**, Debora Gil,Jesús Giraldo. Detecting loss of diversity for an efficient termination of EAs. In book *In 15th International Symposium on Symbolic and Numeric Algorithms for Scientific Computing.*, 270-277, 2013.
- **David Roche**, Debora Gil,Jesús Giraldo. An inference model for analyzing termination conditions of Evolutionary Algorithms. In book *14th Congrs Catal en Intelligencia Artificial*, 216-225, 2011.
- A. Hernandez, D. Gil, **David Roche**, Monica M.S. Matsumoto, Sergio S. Furuie (Eds). Inferring the Performance of Medical Imaging Algorithms. In book *14th International Conference on Computer Analysis of Images and Patterns*,(Vol. 6854, pp. 520528). L. Berlin: Springer-Verlag Berlin Heidelberg.
- **David Roche**, Debora Gil,Jesús Giraldo..Using statistical inference for designing termination conditions ensuring convergence of Evolutionary Algorithms In book *11th European Conference on Artificial Life*

Bibliography

[1] A. Agapie. Theoretical analysis of mutation-adaptive evolutionary algorithms. *Evolutionary Computation*, 9:127–146, 2001.

[2] S.F. Arnold. *The theory of linear models and multivariate observations*. Wiley, 1997.

[3] M. Ashish, S.and Srivastava. *Regression analysis. Theory, methods and applications.* Springer-Verlag, 1990.

[4] A. Auger and N. Hansen. A restart cma evolution strategy with increasing population size. In *CEC 2005*, pages 1769–1776, 2005.

[5] H. Aytug and G.J Koehler. New stopping criterion for genetic algorithms. *Eur J Op Research*, 126:662–674, 2000.

[6] T. Bäck, U. Hammel, and H.P. Schwefel. Evolutionary computation: Comments on the history and current state. *IEEE Transactions on Evolutionary Computation*, 1:3–17, 1997.

[7] Claus Bendtsen. *pso: Particle Swarm Optimization*, 2012. R package version 1.0.3.

[8] H.-G. Beyer and H.-P. Schwefel. Evolution strategies: A comprehensive introduction. *Journal Natural Computing*, 1(1):3–52, 2002.

[9] Trautmann H. Bischl B., Mersmann O. and M. Preuss. Algorithm selection based on exploratory landscape analysis and cost-sensitive learning. In *GECCO, 14th annual conference on Genetic and Evolutionary Computation*, pages 313–320. ACM New York USA, 2012.

[10] J.W. Black and P. Leff. Operational models of pharmacological agonism. *Proc.R.Soc.Lond.B.Biol.Sci.*, 220:141–162, 1983.

[11] S. Cagnoni, E. Lutton, and G. Olague. *Genetic and Evolutionary Computation for Image Processing and Analysis*. Hindawi Publishing Corporation, 2008.

[12] W. Cheney and D. Kincaid. *Numerical Mathematics and Computing, Sixth edition.* Bob Pirtle, USA, 2008.

[13] S. Das and A. Konar. An improved differential evolution scheme for noisy optimization problems. In *Pat Rec Mac Int*, number 3776 in LNCS, pages 417–421, 2005.

[14] K. Deb and S Jain. Running performance metrics for evolutionary multi-objective optimization. In *SEAL*, pages 13–20, 2002.

[15] FJ. Ehlert. Analysis of allosterism in functional assays. *Journal of Pharmacology Exp. Ther.*, 315:740–754, 2005.

[16] S. Finck, N. Hansen, R. Ros, and A. Auger. Real paraemter black-box optimization benchmarking 2009: Presentation of the noiseless functions. Technical Report 2009/20, Research Center PPE, 2009.

[17] R.F. Furchgott. The use of b-haloalkylamines in the differentiation of receptors and in the determination of dissociation constants of receptor-agonist complexes. *Advances in Drug Research.*, eds. Harper N.J., Simmonds A.B.:21–55, 1966.

[18] J. Giraldo. Empirical models and hill coefficients. *Trends. Pharmacology*, 24:63–65, 2003.

[19] J. et al. Giraldo. Assessing the (a)symmetry of concentration-effect curves: empirical versus mechanistic models. *Pharmacology and Therapy*, 95:21–45, 2002.

[20] T. Goel and N Stander. A non-dominance-based online stopping criterion for multiobjective evolutionary algorithms. *International Journal for Numerical Methods in Engineering*, page Online access, 2010.

[21] J.L. Guerrero, Marti L., J. Garcia, A Berlanga, and J.M Molina. Introducing a robust and efficient stopping criterion for moeas. In *CEC*, pages 1–8, 2010.

[22] DA. Hall. Predicting dose-response curve behavior: Mathematical models of allosteric receptor-ligand interactions. *Allosteric receptor modulation in drug targeting.*, ed. Bowery N.G.:39–77, 2006.

[23] N. Hansen. Benchmarking a bi-population cma-es on the bbob-2009 function testbed. In *ACM-GECCO*, 2009.

[24] N. Hansen, A. Auger, R. Ros, S. Finck, and P. Posík. Comparing results of 31 algorithms from the black-box optimization benchmarking. In *GECCO*, pages 1689–1696, 2010.

[25] N. Hansen, S. Finck, R. Ros, and A. Auger. Real-parameter black-box optimization benchmarking 2009: Noiseless functions definitions. Technical Report RR-6829, INRIA, 2009.

[26] AV. Hill. The combinations of haemoglobin with oxygen and with carbon monoxide. biochem. *Biochem*, 7:471–480, 1913.

[27] T. Kenakin. New concepts in drug discovery: collateral efficacy and permissive antagonism. *Nat.Rev.Drug Discov.*, 4:919–927, 2005.

[28] T. Kenakin. Allosteric agonist modulators. *J Recept.Signal.Transduct.Res.*, 27:247–259, 2007.

[29] J. Kennedy and R. Eberhart. Particle swarm optimization. In *Neural Networks, 1995. Proceedings., IEEE International Conference on*, volume 4, pages 1942–1948 vol.4, 1995.

[30] PM. Sexton K.Leach and A. Christopoulos. Allosteric gpcr modulators: taking advantage of permissive receptor pharmacology. *Trends Pharmacol Sci.*, 28:382–389, 2007.

[31] Marcel Kronfeld, Hannes Planatscher, and Andreas Zell. The EvA2 optimization framework. In *LION-SWOP*, number 6073 in LNCS, pages 247–250. Springer Verlag, 2010.

[32] N. M. Kwok, Q. P. Ha, D. K. Liu, G. Fang, and K. C. Tan. Efficient particle swarm optimization: A termination condition based on the decision-making approach. In *CEC 2007*, pages 3353–3360, 2007.

[33] P. Leff. Can operational models of agonism provide a framework for classifying hormone receptors? *Perspectives on Receptor Classification*, eds. Black J.W., Jenkinson D.R., Gerskowitch V.P.:157–167, 1987.

[34] Ashyraliyev M., Fomekong-Nanfack Y, Kaandorp JA, and Blom JG. Systems biology: parameter estimation for biochemical models. In *FEBS*, pages 886–902, 2009.

[35] L. Marti, J. Garcia, A. Berlanga, and J.M Molina. An approach to stopping criteria for multi-objective optimization evolutionary algorithms: The mgbm criterion. In *CEC*, pages 1263–1270, 2009.

[36] O. Mersmann, B. Bischl, H. Trautmann, M. Preuss, C. Weihs, and Rudolph. Exploratory landscape analysis. In *GECCO'11*, pages 829–836, 2011.

[37] O. Mersmann, M. Preuss, and H. Trautmann. Benchmarking evolutionary algorithms: Towards exploratory landscape analysis. In *PPSN XI, Part I, LNCS 6238*, pages 73–82, 2010.

[38] A. Thomas et al. MR. Price, GL. Baillie. Allosteric modulation of the cannabinoid cb1 receptor. *Mol. Pharmacol.*, 68:1484–1495, 2005.

[39] Katharine Mullen, David Ardia, David Gil, Donald Windover, and James Cline. DEoptim: An R package for global optimization by differential evolution. *J Stat Soft*, 40 (6):1–26, 2011.

[40] Hansen N. The cma evolution strategy: a comparing review. In *Towards a new evolutionary computation. Advances on estimation of distribution algorithms.*, pages 75–102. Springer, 2007.

[41] P. Newbold, W.L. Carlson, and B. Thorne. *Statistics for Business and Economics*. Pearson Education, 6th edition, 2007.

[42] V. Ravikumar Pandi and Bijaya K. Panigrahi. Comparative study of evolutionary computing methods for parameter estimation of power quality signals. *IJAEC*, 1(2):28–59, 2010.

[43] K.V. Price, R.M. Storn, and J.A Lampinen. *Differential Evolution: A Practical Approach to Global Optimization*. Springer, 2005.

[44] I Recehnberg. Evolutionsstrategie. *Optimierung Technischer Systeme nach Prinzipien der Biologischen Evoution*, 1973.

[45] I Recehnberg. Cybernetic solution path of an experimental problem. *Wiley-IEEE Press.*, pages 301–310, 1998.

[46] FJ. Richards. A flexible growth function for empirical use. *Journal of Experimental Botany*, 10:290–300, 1959.

[47] Mateen Rizki, Michael Zmuda, and Louis Tamburino. Evolving pattern recognition systems. *IEEE Trans Evol Comp*, 6 (6):594–609, 2002.

[48] D. Roche, D. Gil, and J. Giraldo. An inference model for analyzing termination conditions of evolutionary algorithms. In *CCiA Proc.*, pages 216–225, 2011.

[49] D. Roche, D. Gil, and J. Giraldo. Using statistical inference for designing termination conditions ensuring convergence of evolutionary algorithms. In *Proc. ECAL*, pages 680–687, 2011.

[50] D. Roche, D. Gil, and J. Giraldo. Multiple active receptor conformation, agonist efficacy and maximum effect of the system: the conformation-based operational model of agonism. *Drug Discov.Today*, 18:365–371, 2013.

[51] O. Rudenko and M Schoenauer. A steady performance stopping criterion for paretobased. In *Multi-Objective Programming and Goal Programming Conference*, 2004.

[52] W. Rudin. *Principles of Mathematical Analysis*. McGraw-Hill Science, 1976 (3rd edition).

[53] G Rudolph. Convergence analysis of canonical genetic algorithms. *IEEE Trans Neural Net*, 5:96–101, 1994.

[54] G. Rudolph. Finite markov chain results in evolutionary computation: A tour d'horizon. *Fundamenta Informaticae*, pages 1–22, 1998.

[55] M. Safe, J. Carballido, and I. Ponzoni. On stopping criteria for genetic algorithms. In *SBIA 2004*, number 3171 in LNCS, pages 405–413, 2004.

[56] Ofer M. Shir. *Niching in Evolutionary Algorithms*. Springer, 2013.

[57] Dan Simon. *Evolutionary Optimization Algorithms. Biologically Inspired and Population-Based Approaches to Computer Intelligence*. Wiley, New York, 2013.

[58] JM. Watson SJ. Bradley, CJ. Langmead and RA. Challiss. Quantitative analysis reveals multiple mechanisms of allosteric modulation of the mglu5 receptor in rat astroglia. *Mol.Pharmacol.*, 79:874–885, 2011.

[59] R. Slack and D. Hall. Development of operational models of receptor activation including constitutive receptor activity and their use to determine the efficacy of the chemokine ccl17 at the cc chemokine receptor ccr4. *Br.J.Pharmacol.*, 166:1774–1792, 2012.

[60] R. Storn and K. Price. Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces. *Journal of Global Optimization*, 11:341–359, 1997.

[61] M.F. Tagetiren and P. Suganthan. A multi-populated differential evolution algorithm for solving contrained optimization problem. In *CEC*, pages 340–347, 2006.

[62] Heike Trautmann, Olaf Mersmann, and David Arnu. *cmaes: Covariance Matrix Adapting Evolutionary Strategy*, 2011. R package version 1.0-11.

[63] P. H. Van Der Graaf and R. C. Schoemaker. Analysis of asymmetry of agonist concentration-effect curves. *Journal of Pharmacology and Toxicological methods*, 41:107–115, 1999.

[64] J. Vesterstrom and R. Thomsen. A comparative study of differential evolution, particle swarm optimization, and evolutionary algorithms on numerical benchmark problems. In *CEC*, pages 1980 – 1987, 2004.

[65] T Wagner, H. Trautmann, and L. Martí. A taxonomy of online stopping criteria for multi-objective evolutionary algorithms. In *EMO'11*, pages 16–30, 2011.

[66] T. Wagner, H. Trautmann, and B Naujoks. Ocd: Online convergence detection for evolutionary multi-objective algorithms based on statistical testing. In *EMO*, pages 198–215, 2009.

[67] Weijun Xia and Zhiming Wu. An effective hybrid optimization approach for multi-objective flexible job-shop scheduling problems. *Computers and amp; Industrial Engineering*, 48(2):409 – 425, 2005.

[68] D. Zaharie. Critical values for the control parameters of differential evolution algorithm. In *MENDEL*, pages 62–67, 2002.

[69] D. Zaharie. Falten les dades reals d'any i tot. In *FALTEN DADES REALS*, pages 62–67, 2002.

[70] D. Zaharie. A compartive analysis of crossover variants in differential evolution. In *Proc. Int. Conf. Comp. Sci. Inf. Tech.*, pages 171–181, 2007.

[71] Daniela Zaharie and Dana Petcu. Adaptive pareto differential evolution and its parallelization. In Roman Wyrzykowski, Jack Dongarra, Marcin Paprzycki,

and Jerzy Waniewski, editors, *Parallel Processing and Applied Mathematics*, volume 3019 of *Lecture Notes in Computer Science*, pages 261–268. Springer Berlin Heidelberg, 2004.

[72] K. Zielinski and R. Laur. Stopping criteria for differential evolution in constrained single-objective optimization. In *Studies in Computational Intelligence*, number 143 in Adv Diff Evol, pages 111–138, 2008.