SOME THEORY OF EVOLUTION STRATEGIES

by THOMAS BÄCK†, BORIS NAUJOKS‡

†NuTech Solutions GmbH
Martin-Schmeisser-Weg 15
D-44227 Dortmund
Email: Thomas.Baeck@nutechsolutions.com
and
Leiden Institute for Advanced Computer Science
Niels Bohrweg 1
NL-2333 CA Leiden

‡Center for Applied Systems Analysis (CASA)
Informatik Centrum Dortmund (ICD)
Joseph-von-Fraunhofer-Str. 20
D-44227 Dortmund
Email: naujoks@icd.de

Abstract

This contribution provides some of the theoretical background for understanding the power of evolution strategies from the point of view of global, direct optimization algorithms. In particular, the two aspects of convergence reliability (i.e., global convergence with probability one) and convergence velocity (i.e., how fast is the search progressing towards an optimum) are discussed and some of the known results for evolution strategies are presented, including some indication how this theory can be transferred to genetic algorithms.

The summary of results presented in this paper clarifies that a solid theoretical basis for understanding the power of evolution strategies has been build during the past thirty years, with much progress in the last decade.

1 INTRODUCTION

Unlike other evolutionary algorithms, e.g. genetic algorithms, the theory of evolution strategies has from the very beginning focused on the questions of *convergence velocity* and *convergence reliability* of these algorithms. Loosely speaking, the former concentrates on the speed of the algorithm when a local optimum is approached (and therefore it provides insight into the local behaviour of an evolution strategy) while the latter concentrates on proving that the algorithm is capable of finding the global optimum of the objective function. The convergence velocity analysis requires strong simplifications regarding the objective functions that can be analyzed – so far, the convex case can be handled – while the convergence reliability (i.e., global convergence with probability one) analysis yields

a result for $t \to \infty$ (i.e., an asymptotical result as the running time of the algorithm goes to infinity) independent of the objective function.

These two types of analysis provide important benchmark information regarding the properties of evolution strategies as an optimization algorithm: With respect to the convergence velocity, they should be competitive with gradient-based methods and with respect to convergence reliability, they should prove to be global optimization algorithms, i.e., have the property of global convergence with probability one. The following sections 2 and 3 deal with these two aspects of evolution strategies, and discuss also some related topics such as selection schemes and the transfer of theoretical results to genetic algorithms.

2 CONVERGENCE VELOCITY

Traditionally, evolution strategies have been analyzed in terms of their *convergence velocity*, i.e., the expected improvement in fitness (1) or distance from the optimum $\vec{x}^*$ (2) between two consecutive generations:

$$\varphi = \mathbf{E}(|f(\vec{x}^*) - f(\vec{x}_t)| - |f(\vec{x}^*) - f(\vec{x}_{t+1})|) \quad (1)$$

$$\varphi = \mathbf{E}(\|\vec{x}^* - \vec{x}_t\| - \|\vec{x}^* - \vec{x}_{t+1}\|) \quad (2)$$

The second alternative, distance to the optimum, has been used in evolution strategies with enormous success to analyze algorithmic instances such as the (1+1)- [22], (1+λ)- [26] and (1,λ)- [26], (μ,λ)-[1] [8] and (μ/μ,λ)-strategies[2] [11] on simple quadratic objective functions, namely the sphere model $f(\vec{x}) = f((x_1, \ldots, x_n)^T) = \sum_{i=1}^{n}(x_i - x_i^*)^2 = R^2$ for the assumption of one fixed mutation step size σ and sufficiently large dimensionality n. Without going into the details here, it is worthwhile to mention that for comma-strategies the results can usually be expressed as simple quadratic equations such as

$$\varphi'_{(1,\lambda)} = c_{1,\lambda} \cdot \sigma' - \frac{\sigma'^2}{2} \quad (3)$$

for a (1,λ)-strategy, where $\varphi' = \varphi n/R$ and $\sigma' = \sigma n/R$ denote normalized variables, R is the current distance of the parent $\vec{x}$ to the optimum location $\vec{x}^*$, and $c_{1,\lambda}$ denotes the *progress coefficient* of the strategy [23]. For a (1,λ)-strategy, this constant depends only on λ and characterizes the selective pressure of the (1,λ)-selection method. It is defined as the expectation of the largest of λ independent random variables Z_i with identical, standardized and normalized Gaussian distribution $Z_i \sim N(0,1)$, which are rearranged in increasing order and relabeled by $Z_{1:\lambda} \leq Z_{2:\lambda} \leq \ldots \leq Z_{\lambda:\lambda}$:

$$c_{1,\lambda} = \mathbf{E}(Z_{\lambda:\lambda}) = \int_{-\infty}^{\infty} z \cdot \frac{d}{dz}[\Phi(z)]^\lambda dz \; . \quad (4)$$

[1] With $\mu > 1$, but without recombination for the sake of theoretical analysis.

[2] With global intermediary or global discrete recombination.