Convergence of a Discretized Self-Adaptive Evolutionary Algorithm on Multi-Dimensional Problems

William E. Hart, John M. DeLaurentis

***Abstract*— We consider the convergence properties of a non-elitist self-adaptive evolutionary strategy (ES) on multi-dimensional problems. In particular, we apply our recent convergence theory for a discretized $(1, \lambda)$-ES to design a related $(1, \lambda)$-ES that converges on a class of seperable, unimodal multi-dimensional problems.**

I. Introduction

The distinguishing feature of self-adaptive evolutionary algorithms (EAs) is that the control parameters (like mutation step lengths) are evolved by the evolutionary algorithm. Thus the control parameters are adapted in an implicit manner that relies on the evolutionary dynamics to ensure that more effective control parameters are propagated during the search [3]. Self-adaptation is a central feature of EAs like evolutionary stategies (ES) and evolutionary programming (EP), which are applied to continuous design spaces.

Rudolph [7] summarizes theoretical results concerning self-adaptive EAs and notes that the theoretical underpinnings for these methods are essentially unexplored. In particular, convergence theories that ensure convergence to a limit point on continuous spaces have only been developed by Rudolph [6], Hart, DeLaurentis and Ferguson [5], and Auger et al. [1], [2].

In this paper, we illustrate how our analysis of a $(1, \lambda)$-ES for one-dimensional unimodal functions can be used to ensure convergence of a related ES on multi-dimensional functions. This $(1, \lambda)$-ES randomly selects a search dimension in each iteration, along which points generated. For a general class of separable functions, our analysis shows that the ES searches along each dimension independently, and thus this ES converges to the (global) minimum.

II. A Discretized ES

Figure 1 describes the self-adaptive $(1, \lambda)$-ES that we consider in this paper; $x^t \in \mathbf{R}^n$ denotes the point in iteration t, and x_k^t is the value of the k-th dimension of x^t. This ES generates λ new points $(x^{t,1}, \ldots, x^{t,\lambda})$ in each iteration by searching along the k-dimension, and it selects the best point generated for the next iteration. This ES updates the mutation scale and new points using $2n$ random variables D_k and B_k, $k = 1, \ldots, n$; $\sigma_k^{t,i}$ is updated with the value $d_k^{t,i}$, and $x_k^{t,i}$ is updated with the value $b_k^{t,i}$. Since we are searching each dimension in an independent manner, this allows for different randomization schemes for different dimensions.

Given x^0, σ^0
For $t = 1, \ldots$
  Select $k \in \{1, \ldots, n\}$ uniformly at random
  For $i = 1 : \lambda$
    $x^{t,i} = x^t$ and $\sigma^{t,i} = \sigma^t$
    $\sigma_k^{t,i} = \sigma_k^t \cdot d_k^{t,i}$
    $x_k^{t,i} = x_k^t + \sigma_k^{t,i} \cdot b_k^{t,i}$
  End
  $j = \arg\min_{i=1:\lambda} \ f(x^{t,i})$
  $x^{t+1} = x^{t,j}$
  $\sigma^{t+1} = \sigma^{t,j}$
End

Fig. 1. A self-adaptive $(1, \lambda)$-ES for multi-dimensional problems: $d_k^{t,i}$ and $b_k^{t,i}$ are generated from random variables D_k and B_k respectively.

We consider self-adaptive $(1, \lambda)$-ESs that use discrete random variables for D_k and B_k. Let $d_k^{t,i}$ be a realization of the random variable D_k: $d_k^{t,i} \in \{\gamma_k, 1, \eta_k\}$, where $\gamma_k < 1 < \eta_k$. Let $\nu_k^1 = P\{D_k = \gamma_k\}$, $\nu_k^2 = P\{D_k = 1\}$ and $\nu_k^3 = P\{D_k = \eta_k\}$ for all t; we assume that these probabilities are nonzero. Thus $\sigma_k^{t,i} = \sigma_k^t d_k^{t,i}$. A step length $\sigma_k^{t,i}$ is used to update the k-th dimension of $x^{t,i}$: $x_k^{t,i} = x_k^t + \sigma_k^{t,i} \cdot b_k^{t,i}$, where $b_k^{t,i}$ is the realization of the random variable B_k: $b_k^{t,i} \in \{-1, +1\}$ with probabilities $\{\frac{1}{2}, \frac{1}{2}\}$ respectively.

Let Algorithm A denote the self-adaptive $(1, \lambda)$-ES that employs these random variables. We make the following assumption concerning the parameterization of

Corresponding Author: William Hart, Sandia National Laboratories P. O. Box 5800, MS 1110 Albuquerque, NM 87185-1110 wehart@sandia.gov

John DeLaurentis, Sandia National Laboratories P. O. Box 5800, MS 1110 Albuquerque, NM 87185-1110 jdlauren@sandia.gov

Algorithm A:

Assumption 1 *For $k = 1, \ldots, n$, Algorithm A has the property that*

1) $1/2 < \gamma_k < 1$,
2) $1 < \eta_k < 2 - \gamma_k$,
3) $\gamma_k, \eta_k \in \mathbf{Q}$, *and*
4) ν_k^1, ν_k^3 *and* λ *are chosen so that* $p_k \log \eta_k + q_k \log \gamma_k > 0$, *where*
   - $p_k = 1 = \left(1 - \nu_k^3/2\right)^\lambda + \left(\nu_k^3/2\right)^\lambda$
   - $q_k = \left((1 + \nu_k^1)/2\right)^\lambda - \left((1 - \nu_k^1)/2\right)^\lambda$.

Our analysis applies when Algorithm A is applied to objective functions that satisfy the following assumption:

Assumption 2 *The function $f(x) = \sum_{k=1}^n g_k(x_k)$, where $g_k : \mathbf{R} \to \mathbf{R}$ has the property that*

1) *there exists a unique global minimum $x_k^* = 0$,*
2) *g_k is strictly monotonically increasing for $x_k \in (x_k^*, \infty)$.*
3) *g_k is strictly monotonically decreasing for $x_k \in (-\infty, x_k^*)$.*

The elements of Assumption 1 are consistent with the assumptions made to ensure convergence on one-dimensional unimodal problems in Hart et al. [5]. Assumption 1.1 and 1.2 ensure that the expansion and contraction factors are well-balanced. Assumption 1.3 is used to simplify the proof that the step lengths converge to zero. Assumption 1.4 ensures that the expected search behavior on g_k away from x_k^* should increase the step length.

Let $X_{\lambda,k}^t$ and $\Sigma_{\lambda,k}^t$ be random variables that describe the distribution of the values of x^t and σ^t respectively when a population of size λ is used by Algorithm A on a function that satisfies Assumption 2. The following two theorems restate our previous convergence results using our current notation. Note that these theorems apply to one-dimensional, unimodal functions that satisfy Assumption 2.

Theorem 1 (Theorem 2, [5]) *Suppose that Algorithm A satisfies Assumptions 1.1, 1.2 and 1.3, and suppose that the function $f : \mathbf{R} \to \mathbf{R}$ satisfies Assumption 2. Then there exists $\lambda' > 0$ such that for all $\lambda > \lambda'$, $\Sigma_{\lambda,1}^t \xrightarrow{a.s.} 0$.*

Theorem 2 (Theorem 3, [5]) *Suppose that Algorithm A satisfies Assumption 1, and suppose that the function $f : \mathbf{R} \to \mathbf{R}$ satisfies Assumption 2. Then there exists $\lambda'' \geq \lambda'$ such that for all $\lambda > \lambda''$, $X_{\lambda,1}^t \xrightarrow{a.s.} 0$.*

Note that we assume that $x_k^* = 0$ for each g_k only for convenience sake, since if an EA converges on a function that satisfies this condition, then we can show convergence for any other function h with nonzero global minimizer by considering the convergence of the function $f(x) = h(x_1 + x_1^*, \ldots, x_n + x_n^*)$.

III. Convergence Theory

The convergence theory that we describe considers the sequence of best points in each iteration of an implicitly self-adaptive $(1, \lambda)$-ES, and we show that these points converge *almost surely* (i.e. with probability one) on unimodal functions for a suitable choice of parameters. If Y and Y_t are random variables, then we say that the sequence $\{Y_t\}_{t \geq 0}$ converges almost surely to Y if $P\{\lim_{t \to \infty} Y_t = Y\} = 1$. We write this as $Y_t \xrightarrow{a.s.} Y$. See Grimmett and Stirzaker [4] for a thorough discussion of stochastic convergence.

The following theorem shows that Algorithm A converges to the minimum of f by showing that $X_{\lambda,k}^t \xrightarrow{a.s.} 0$ and $\Sigma_{\lambda,k}^t \xrightarrow{a.s.} 0$ for all k.

Theorem 3 *Suppose that Algorithm A satisfies Assumption 1 and that f satisfies Assumption 2. Then there exists λ''' such that for all $\lambda \geq \lambda'''$, $X_{\lambda,k}^t \xrightarrow{a.s.} 0$ and $\Sigma_{\lambda,k}^t \xrightarrow{a.s.} 0$ for all k.*

Proof: Let $X_{\lambda,k}^t$ and $\Sigma_{\lambda,k}^t$ be the stochastic process defined on some probability space $(\Omega, \mathcal{F}, P)$ that describes the behavior of Algorithm A on f. A simple application of the Borel-Cantelli Lemma demonstrates that each dimension is selected infinitely often with probability one.

For some event $\omega \in \Omega$, consider $x^t(\omega)$ and suppose that the k-th dimension has been selected for search. Any point generated in this iteration will have a function value

$$\sum_{i \neq k} g_i(x_i^t(\omega)) + g_k(x_k^t(\omega) + \Delta x_k^t(\omega)),$$

where $\Delta x_k^t(\omega)$ is some step taken along the k-th dimension. Note that the first term in this expression is independent of the value of $\Delta x_k^t(\omega)$. Since the selection criteria used to find the best of the λ points in this iteration solely depends on the *relative ranks* of the points generated, this selection is independent of the particular value of this constant term.

This argument applies in all cases, so we have shown that in each iteration, the selection of the next iterate only involves the value of the current dimension being modified. Consequently, the random variables $X_{\lambda,1}^t, \ldots, X_{\lambda,n}^t$ are independent and the random variables $\Sigma_{\lambda,1}^t, \ldots, \Sigma_{\lambda,n}^t$ are independent. Thus our previous analysis can be directly applied to describe the dynamics of $X_{\lambda,k}^t$ and $\Sigma_{\lambda,k}^t$ for any k [5]; for any k, Theorems 1 and 2 to show that there exists some λ_k such that for all $\lambda \geq \lambda_k$,

$X^t_{\lambda,k} \xrightarrow{a.s.} 0$ and $\Sigma^t_{\lambda,k} \xrightarrow{a.s.} 0$. It suffices to make $\lambda''' = \max_k \lambda_k$ to complete our proof. ■

Theorem 3 applies for a general class of unimodal problems in any dimension. Note that Algorithm A is not practically relevant if $\lambda_0 \geq 6$, since in this case you could simply enumerate all of the six points that can be generated in each iteration. However, our analysis of Algorithm A is effectively reduced to the independent convergence along each dimension of the $(1, \lambda)$-ES described by Hart et al. [5]. Consequently, we argue that our analysis of the value λ_0 for the one-dimensional case effectively demonstrates that Algorithm A can be practically relevant. That is, we have $\lambda_0 < 6$ in cases where $\lambda_k < 6$ for each dimension.

We conclude with a corollary that highlights the fact that our analysis demonstrates convergence on sphere functions, which are commonly considered in the analysis of ES. Consequently, this work complements the recent work by Auger [1] which proves convergence of the $(1, \lambda)$-ES with normal mutation on sphere functions.

Corollary 1 *Suppose that Algorithm A satisfies Assumption 1 and that $f(x) = \sum_i x_i^2$. Then there exists λ_0 such that for all $\lambda \geq \lambda_0$, $X^t_{\lambda,k} \xrightarrow{a.s.} 0$ and $\Sigma^t_{\lambda,k} \xrightarrow{a.s.} 0$ for all k.*

ACKNOWLEDGEMENTS

We thank Anne Auger for discussions that led us to think more carefully about the implications of our previous analysis. This work was supported by the DOE Office of Science, and it was performed at Sandia National Laboratories. Sandia is a multiprogram laboratory operated by Sandia corporation, a Lockheed Martin Company, for the United States Department of Energy under Contract DE-AC04-94AL85000.

REFERENCES

[1] A. Auger. Convergence results for (1,λ)-SA-ES using the theory of φ-irreducible Markov chains. In *Proc of the Evolutionary Algorithms workshop of the Thirtieth International Colloquium on Automata, Languages and Programming*, 2003. (to appear).

[2] A. Auger, C. Le Bris, and M. Schoenauer. Dimension-independent convergence rate for non-iostrophic $(1, \lambda)$-ES. In E. Cantu-Paz, J. A. Foster, K. Deb, L. D. Davis, R. Roy, U.-M. O'Reilly, H.-G. Beyer, R. Standish, G. Kendal, S. Wilson, M. Harman, J. Wegener, D. Dasgupta, M. A. Potter, A. C. Schultz, K. A. Dowsland, N. Jonoska, and J. Miller, editors, *Proc of the Genetic and Evolutionary Computation Conference*, pages 512–524, 2003.

[3] A. E. Eiben, R. Hinterding, and Z. Michalewicz. Parameter control in evolutionary algorithms. *IEEE Trans Evolutionary Computation*, 3(2):124 – 141, 1999.

[4] G. R. Grimmett and D. R. Stirzaker. *Probability and Random Processes, Second Edition*. Oxford University Press, Oxford, 1992.

[5] W. E. Hart, J. M. DeLaurentis, and L. A. Ferguson. On the convergence of an implicitly self-adaptive evolutionary algorithm on one-dimensional unimodal problems. *IEEE Trans Evolutionary Computation*, 2002. (to appear).

[6] G. Rudolph. Convergence of non-elitist strategies. In *Proc of the First IEEE Conf on Evolutionary Computation*, volume 1, pages 63–66, Piscataway, NJ, 1994. IEEE Press.

[7] G. Rudolph. Self-adaptive mutations may lead to premature convergence. *IEEE Trans Evolutionary Computation*, 5(4):410–414, 2001.