





PROCEEDINGS OF SPIE

SPIE—The International Society for Optical Engineering

# ***Applications and Science of Neural Networks, Fuzzy Systems, and Evolutionary Computation V***

**Bruno Bosacchi**  
**David B. Fogel**  
**James C. Bezdek**  
*Chairs/Editors*

**9–10 July 2002**  
**Seattle, Washington, USA**

*Sponsored and Published by*  
SPIE—The International Society for Optical Engineering

*Cooperating Organizations*  
The Boeing Company (USA)  
Pacific Northwest National Laboratory (USA)  
Washington State University/College of Sciences  
and College of Engineering and Architecture (USA)  
WTC—Washington Technology Center (USA)  
University of Washington/College of Engineering (USA)  
University of Washington/Center for Nanotechnology (USA)



**Volume 4787**

SPIE is an international technical society dedicated to advancing engineering and scientific applications of optical, photonic, imaging, electronic, and optoelectronic technologies.



The papers appearing in this book comprise the proceedings of the meeting mentioned on the cover and title page. They reflect the authors' opinions and are published as presented and without change, in the interests of timely dissemination. Their inclusion in this publication does not necessarily constitute endorsement by the editors or by SPIE.

Please use the following format to cite material from this book:

Author(s), "Title of paper," in *Applications and Science of Neural Networks, Fuzzy Systems, and Evolutionary Computation V*, Bruno Bosacchi, David B. Fogel, James C. Bezdek, Editors, Proceedings of SPIE Vol. 4787, page numbers (2002).

ISSN 0277-786X  
ISBN 0-8194-4554-1

Published by  
**SPIE—The International Society for Optical Engineering**  
P.O. Box 10, Bellingham, Washington 98227-0010 USA  
Telephone 1 360/676-3290 (Pacific Time) • Fax 1 360/647-1445

Copyright © 2002, The Society of Photo-Optical Instrumentation Engineers.

Copying of material in this book for internal or personal use, or for the internal or personal use of specific clients, beyond the fair use provisions granted by the U.S. Copyright Law is authorized by SPIE subject to payment of copying fees. The Transactional Reporting Service base fee for this volume is \$15.00 per article (or portion thereof), which should be paid directly to the Copyright Clearance Center (CCC), 222 Rosewood Drive, Danvers, MA 01923. Payment may also be made electronically through CCC Online at <http://www.directory.net/copyright/>. Other copying for republication, resale, advertising or promotion, or any form of systematic or multiple reproduction of any material in this book is prohibited except with permission in writing from the publisher. The CCC fee code is 0277-786X/02/\$15.00.

Printed in the United States of America.

## Conference Committee

### *Conference Chairs*

**Bruno Bosacchi**, Princeton University (USA)  
**David B. Fogel**, Natural Selection, Inc. (USA)  
**James C. Bezdek**, University of West Florida (USA)

### *Program Committee*

**Peter J. Angeline**, Natural Selection, Inc. (USA)  
**Piero P. Bonissone**, GE Global Research Center (USA)  
**Carlos Brody**, Cold Spring Harbor Research Laboratory (USA)  
**H. John Caulfield**, Fisk University (USA)  
**Kumar Chellapilla**, Natural Selection, Inc. (USA)  
**Christos Douligeris**, University of Piraeus (Greece)  
**Thomas M. English**, Tom English Computing Innovations (USA)  
**Jack J. Gelfand**, Princeton University (USA)  
**Lawrence O. Hall**, University of South Florida (USA)  
**Fredric M. Ham**, Florida Institute of Technology (USA)  
**Laurent Itti**, University of Southern California (USA)  
**Andrew A. Kostrzewski**, Physical Optics Corporation (USA)  
**Daniel D. Lee**, University of Pennsylvania (USA)  
**Evangelia Micheli-Tzanakou**, Rutgers University (USA)  
**Sunanda Mitra**, Texas Tech University (USA)  
**Witold Pedrycz**, University of Alberta (Canada)  
**Anthony V. Sebald**, University of California/San Diego (USA)  
**Mohan M. Trivedi**, University of California/San Diego (USA)  
**Athanasios V. Vasilakos**, ICS-FORTH (Greece)

### *Session Chairs*

- 1 Evolutionary Computation  
**Piero P. Bonissone**, GE Global Research Center (USA)
- 2 Computational Neuroscience  
**Joseph Shamir**, Technion-Israel Institute of Technology (Israel)
- 3 Applications I  
**Arthur J. Decker**, NASA Glenn Research Center (USA)
- 4 Applications II  
**Bruno Bosacchi**, Princeton University (USA)

- 5 New Directions  
**Bruno Bosacchi**, Princeton University (USA)
- 6 Soft Computing  
**Igor Maslov**, CUNY/City College (USA)
- 7 Applications III  
**Bruno Bosacchi**, Princeton University (USA)
- 8 Applications IV  
**Bruno Bosacchi**, Princeton University (USA)

# Contents

vii *Conference Committee*

---

## SESSION 1 EVOLUTIONARY COMPUTATION

---

- 1 **Extending self-adaptation in evolutionary algorithms (Invited Paper)** [4787-01]  
D. B. Fogel, Natural Selection, Inc. (USA)
- 10 **Exploring the Pareto frontier using multisexual evolutionary algorithms: an application to a flexible manufacturing problem (Invited Paper)** [4787-02]  
S. R. Bonissone, Consultant (USA); R. Subbu, GE Global Research Ctr. (USA)
- 23 **Control of quantum phenomena through evolutionary engineering: let the system do the thinking (Invited Paper)** [4787-03]  
H. H. Rabitz, B. Bosacchi, Princeton Univ. (USA)

---

## SESSION 2 COMPUTATIONAL NEUROSCIENCE

---

- 37 **Toward highly capable neuromorphic autonomous robots: beobots (Invited Paper)** [4787-06]  
L. Itti, Univ. of Southern California (USA)
- 46 **Dimensionality reduction for sensorimotor learning in mobile robotics (Invited Paper)** [4787-07]  
D. D. Lee, Univ. of Pennsylvania (USA)

---

## SESSION 3 APPLICATIONS I

---

- 53 **When will it break? A hybrid soft computing model to predict time-to-break margins in paper machines (Invited Paper)** [4787-08]  
P. P. Bonissone, K. Goebel, GE Global Research Ctr. (USA)
- 65 **Reliable neural modeling of pHEMT from a smaller number of measurement data** [4787-11]  
M. Joodaki, G. Kompa, Univ. of Kassel (Germany)

---

## SESSION 4 APPLICATIONS II

---

- 74 **Genetic algorithms and support vector machines for time series classification** [4787-12]  
D. R. Eads, Los Alamos National Lab. (USA) and Rochester Institute of Technology (USA);  
D. Hill, Rochester Institute of Technology (USA); S. Davis, S. J. Perkins, J. Ma, R. B. Porter,  
J. P. Theiler, Los Alamos National Lab. (USA)

- 86 **Network-centric decision architecture for financial or 1/f data models** [4787-13]  
H. M. Jaenisch, J. W. Handley, SPARTA, Inc. (USA); S. Massey, Alabama A&M Univ. (USA);  
C. T. Case, C. G. Songy, SPARTA, Inc. (USA)
- 98 **Enhancement of color image obtained from PCA-FCM technique using local area histogram equalization** [4787-15]  
S. Chitwong, T. Boonmee, F. Cheevasuvit, King Mongkut's Institute of Technology  
Ladkrabang (Thailand)
- 107 **GA-based registration of temporal and multimodal ocular fundus image pairs** [4787-16]  
S. S. Myint, O. Chutatape, Nanyang Technological Univ. (Singapore)

---

#### SESSION 5 NEW DIRECTIONS

- 114 **Optical implementation of neural networks (Invited Paper)** [4787-17]  
F. T. S. Yu, R. Guo, The Pennsylvania State Univ. (USA)
- 124 **Spin chaos in magnetic resonance (Invited Paper)** [4787-18]  
Y.-Y. Lin, S. Y. Huang, Univ. of California/Los Angeles (USA); N. Lisitz, W. S. Warren,  
Princeton Univ. (USA)
- 132 **Nature's alternative to hyperspectral imaging and why nature is right (Invited Paper)**  
[4787-19]  
H. J. Caulfield, Fisk Univ. (USA)

---

#### SESSION 6 SOFT COMPUTING

- 137 **Soft computing, visual communication, spatial/temporal events, and optimum bandwidth utilization (Invited Paper)** [4787-20]  
T. P. Jansson, A. A. Kostrzewski, B. K. Sun, W. Wang, K. Kolesnikov, S. A. Kupiec, Physical  
Optics Corp. (USA)
- 147 **Object-oriented soft-computing video compression and digital processing (Invited Paper)**  
[4787-21]  
A. A. Kostrzewski, S. A. Kupiec, T. P. Jansson, Physical Optics Corp. (USA)
- 158 **Manifold geometric invariants and object-centric approach** [4787-22]  
T. P. Jansson, Physical Optics Corp. (USA)
- 174 **WLAN visual sensor networking** [4787-23]  
A. A. Kostrzewski, W. Wang, T. P. Jansson, Physical Optics Corp. (USA)
- 179 **Fuzzy-logic adaptive genetic algorithm (FLAGA) in optical design** [4787-24]  
I. P. Agurok, A. A. Kostrzewski, T. P. Jansson, Physical Optics Corp. (USA)

---

**SESSION 7      APPLICATIONS III**

---

- 186    **Multimode intelligent data compressor/decompressor (MICE) optimized for speed and storage capacity** [4787-25]  
M. A. Piliavin, L. D'Adderio, T. M. Aye, J. L. Jansson, A. A. Kostrzewski, Physical Optics Corp. (USA)
- 198    **Efficient image reduction for image registration with evolutionary algorithm** [4787-27]  
I. V. Maslov, CUNY/Graduate Ctr. (USA); I. Gertner, CUNY/City College (USA)
- 210    **Using neural nets to measure ocular refractive errors: a proposal** [4787-28]  
A. V. Netto, M. C. Ferreira de Oliveira, Univ. de São Paulo (Brazil)

---

**SESSION 8      APPLICATIONS IV**

---

- 218    **Applying fuzzy logic in a complex temperature control system** [4787-29]  
S. D. Kaehler, Boeing Co. (USA)
- 232    **MLP neural network super-resolution restoration for the undersampled low-resolution image** [4787-31]  
B. H. Su, W. Q. Jin, L. H. Niu, G. R. Liu, Beijing Institute of Technology (China)

---

**POSTER SESSION**

---

- 236    **Architectures and application of data fusion system based on nervous system evolution** [4787-34]  
Z. H. Fang, T. Chen, B. C. Zhai, Y. Zhao, Z. J. Tan, Changchun Institute of Optics, Fine Mechanics and Physics (China)
- 244    **Fast parallel 3D stereovision measurement system using an artificial neural network** [4787-36]  
D. Feng, Y. B. Yan, Tsinghua Univ. (China); N. G. Lu, W. Y. Deng, Beijing Institute of Machinery (China)
- 250    **Phase-only filter solution based on fuzzy logic for reducing the defocus problem** [4787-40]  
E. Gur, Z. Zalevsky, D. Elkind, D. Mendlovic, Tel Aviv Univ. (Israel)
- 259    **CPU and memory allocation optimization using fuzzy logic** [4787-41]  
Z. Zalevsky, E. Gur, D. Mendlovic, Tel Aviv Univ. (Israel)
- 267    **Improving classification of neural networks by reducing lens aperture** [4787-42]  
I. Stainvas, Z. Zalevsky, D. Mendlovic, N. Intrator, Tel Aviv Univ. (Israel)
- 277    *Extended Abstract: Dynamic brain-state-in-a-box* [4787-4]  
N. H. Farhat, Univ. of Pennsylvania (USA)
- 281    *Addendum*
- 282    *Author Index*



# Extending Self-Adaptation in Evolutionary Algorithms

David B. Fogel\*

Natural Selection, Inc., 3333 N. Torrey Pines Ct., Suite 200, La Jolla, CA 92037 USA

## ABSTRACT

Self-adaptation in evolutionary algorithms concerns processes in which individuals incorporate information on how to search for new individuals. Instead of detailing the means for searching the space of possible solutions a priori, a process of random variation is applied both in terms of searching the space and searching for strategies to search the space. In one common implementation, each individual in the population is represented as a pair of vectors  $(\mathbf{x}, \sigma)$ , where  $\mathbf{x}$  is the candidate solution to an optimization problem scored in terms of a function  $f(\mathbf{x})$ , and  $\sigma$  is the so-called *strategy parameter* vector that influences how offspring will be created from the individual. Typically,  $\sigma$  describes a variance or covariance matrix under Gaussian mutations. Experimental evidence suggests that the elements of  $\sigma$  can sometimes become too small to explore the given search space adequately. The evolutionary search then stagnates until the elements of  $\sigma$  grow sufficiently large as a result of random variation. Several methods have been offered to remedy this situation. This paper reviews recent results with one such method, which associates multiple strategy parameter vectors with a single individual. A single strategy vector is active at any time and dictates how offspring will be generated. Experiments on four 10-dimensional benchmark functions are reviewed, in which the number of strategy parameter vector is varied over 1, 2, 3, 4, 5, 10, and 20. The results indicate advantages for using multiple strategy parameter vectors. Furthermore, the relationship between the mean best result after a fixed number of generations and the number of strategy parameter vectors can be determined reliably in each case.

**Keywords:** Evolutionary computation, evolutionary algorithm, optimization, self-adaptation, parameter control

## 1. INTRODUCTION

In nature, individuals and species can evolve mechanisms that affect the mutational distribution of new offspring. This can be modelled in evolutionary algorithms by including *self-adaptation*, in which each individual in a population encodes not only a candidate solution to a particular optimization problem, but also information on how to generate offspring as well. Empirical evidence and mathematical theory have long suggested that any fixed distribution of variation cannot be optimal for optimization (Bremermann, 1958; Wolpert and Macready, 1997). Unfortunately, many efforts in evolutionary computation during the period from 1975-1990 concentrated on applying fixed probabilities of crossover and mutation (denoted  $P_c$  and  $P_m$ , with values of about 0.6-1.0 and 0.0001-0.01, respectively), following incorrect theory regarding schema propagation and the effects of that propagation on optimal search (Holland, 1975; Macready and Wolpert, 1998). Self-adaptation offers the possibility for adapting the parameters that control the search online, as the search is conducted. From the perspective of optimization, there have been few reasonable alternatives for online variation control.

One alternative is Rechenberg's 1/5 rule (Rechenberg, 1973). This rule asserts that the convergence rate can be optimized when 1/5 of all offspring are superior to their parents. The 1/5 rule, however, was derived as a result of analysis on linear and quadratic surfaces, and represents a compromise solution. Moreover, the result holds for the case of infinite dimension size, for populations comprising only one parent and one offspring. Chellapilla and Fogel (1999) showed that the 1/5 rule may perform far less than optimal on many common benchmark functions that are not linear or convex. Other rules to control parameters are available, but these generally require significant handtuning. For a comprehensive review, see Eiben et al., 1999.

Self-adaptation was apparently first applied in evolutionary algorithms in independent investigations by Reed et al. (1967) and Rosenberg (1967). Rechenberg (1994) remarked that he conceived of this concept in 1967 and incorporated it subsequently into evolution strategies (Rechenberg, 1973). The most well-known form of self-adaptation follows Schwefel (1981, 1995), and is restricted to the case of optimizing a real-valued functional  $f(\mathbf{x})$ . A strategy parameter vector  $\sigma$  is associated with each individual  $\mathbf{x}$  in the population and guides the generation of offspring via the rule:

$$\sigma'_i = \sigma_i \times \exp(\tau N(0,1) + \tau' N_i(0,1)) \quad (1)$$

$$\mathbf{x}'_i = \mathbf{x}_i + \sigma'_i N(0,1) \quad (2)$$

where  $i$  denotes the  $i$ th dimension of the solution vector  $\mathbf{x}$  or strategy parameter vector  $\sigma$ ,  $N(0,1)$  is a standard Gaussian random variable,  $N_i(0,1)$  designates that a standard Gaussian random variable is sampled anew for each  $i$ th dimension, and  $\tau$  and  $\tau'$  are constants, typically set proportional (or equal to)  $1/\sqrt{2n}$  and  $1/\sqrt{2\sqrt{n}}$ , respectively, where  $n$  is the number of dimensions in  $\mathbf{x}$  and  $\sigma$ . Variations of this procedure can: 1) restrict the length of  $\sigma$  to one control element governing all dimensions, 2) include the use of correlated values of  $\sigma_i$  and  $\sigma_j$ , where  $i \neq j$ , or 3) change the order of the update equations (1) and (2) and use  $\sigma_j$  in place of  $\sigma_i$  (Gehlihaar and Fogel, 1996).

A less well-known but quite similar method was offered independently in Fogel et al. (1991), within the rubric of evolutionary programming (a *meta-evolutionary programming*), where for the same task of optimizing  $f(\mathbf{x})$ , the update equations were:

$$x'_i = x_i + \sigma_i N(0,1) \quad (3)$$

$$\sigma'_i = \sigma_i + c\sigma_i N(0,1) \quad (4)$$

where  $c$  is a positive scaling constant with a recommended value of  $1/6$  (this was intended to ensure that the vast majority of cases would maintain positive values of  $\sigma'_i$ ). In the case where a new value  $\sigma'_i$  were to go below zero, the value would be reset to a small positive constant  $\epsilon$ . This constant was often set to 0.0001 or lower. This method was also extended to include correlated mutations (Fogel et al. 1992). Comparisons to Schwefel's method above were offered in Saravanan and Fogel (1994) and Saravanan et al. (1995) on benchmark functions that were of interest at the time. The results generally (but not always) favored Schwefel's method and subsequent efforts within both evolution strategies and evolutionary programming have generally followed along this line. The no free lunch theorem (Wolpert and Macready, 1997; Köppen et al., 2001) and other related theory (Fogel and Ghoseil, 1997) indicate that neither Schwefel's nor Fogel's update rule can be optimal on every problem and across all representations. Assessing when to use either rule, or another rule, remains an open question (see also Herrera and Lozano, 2001; Rudolph, 2001; Beyer and Deb, 2001).

The update equations (1) and (2) impose some difficulties. One such difficulty involves the potential for having the elements of  $\sigma$  fall too small too quickly. Each element of  $\sigma$  describes a value related to the average step size in the corresponding dimension. Therefore, having elements of  $\sigma$  that are too small leads to an inefficient search. There is an inherent problem of determining when the step sizes are "too small" -- this is not only problem dependent but also depends on each individual's local neighborhood on the response surface. One technique (similar to Fogel et al. 1991) to overcome this is to mandate a lower bound on the elements of  $\sigma$ . Thus, there will always be at least some persistent search as mandated by the lower bound. This method, however, begs the question of setting the appropriate lower bound, and the optimal bound should vary by dimension.

Alternatively, Fogel et al. (2001) have proposed a method that extends the self-adaptive procedure to include  $m$  vectors  $\sigma_k$ ,  $k = 1, \dots, m$ , along with each candidate solution. Thus, instead of the typical representation of  $(\mathbf{x}, \sigma)$ , consider now  $(\mathbf{x}, \sigma_1, \dots, \sigma_m)$ . Only one of the  $m$  strategy vectors associated with an individual is *active* at any time. The active vector is used to control how to generate offspring using equations (1) and (2). Prior to creating an offspring, there is a probability  $p_s$  of changing the active vector by selecting randomly from among the other  $m-1$  choices. In this way, when a strategy parameter vector is successful in generating improved offspring, it is reinforced because (1) applies only to the active vector. When an individual fails to generate improved offspring there is a chance of changing the strategy to an alternative that may work better in the current circumstance.

The approach of Fogel et al. (2001) may offer the following advantages:

- 1) Under traditional self-adaptation of  $(\mathbf{x}, \sigma)$ , when the parameters of  $\sigma$  become too small the user must wait for mutation via (1) to increase them. This can be time consuming. With the adoption of  $(\mathbf{x}, \sigma_1, \dots, \sigma_m)$  it may be possible to switch randomly to an alternative  $\sigma$  vector that has larger elements and accomplish this more coarse search with a greater probability than the option of mutating smaller values into larger values above. Essentially, including multiple  $\sigma$  vectors may make it easier to "forget" strategy parameters that are no longer useful.
- 2) The  $m$  strategy vectors may serve as a memory, indicating successful strategies for searching a response surface that have worked in previous generations. Should an individual need to alternate between strategies for generating offspring, the use of multiple strategy vectors allows for faster switching between those strategies than might be possible if only a single strategy vector had to be mutated into the desired alternative.

There are also the following possible disadvantages:

- 1) The optimum number  $m$  and optimum probability  $p_s$  are unknown and problem dependent.

2) There is no mathematical analysis that suggests how to set  $m$  and  $p_s$  reasonably.

In the absence of a guiding mathematical theory, experimentation can provide initial evidence on the potential utility of this concept. Such experimentation has been conducted and reported in Fogel et al. (2001).

## 2. REVIEW OF RESULTS REPORTED

Fogel et al. (2001) conducted one of the first studies on multiple-vector self-adaptation. The reported results are reviewed here. The following experiments were conducted. Attention was given to four typical benchmark functions in  $n = 10$  dimensions (see Yao et al., 1999):

$$\text{F1: } \sum x_i^2, i = 1, \dots, n$$

$$\text{F2: } \sum [100(x_i^2 - x_{i+1})^2 + (x_i - 1)^2] + [100(x_n^2 - x_1)^2 + (x_n - 1)^2], i = 1, \dots, n-1$$

$$\text{F3: } \sum [x_i^2 - 10 \cos(2\pi x_i) + 10], i = 1, \dots, n$$

$$\text{F4: } -20 \exp[-0.2 \sqrt{n^{-1} \sum x_i^2}] - \exp[n^{-1} \sum \cos(2\pi x_i)] + 20 + e, i = 1, \dots, n$$

Function F1 is strongly convex with only a single optimum (no local optima). Function F2 (Rosenbrock) has only a single optimum as well, but has a deep groove when depicted in three dimensions that can make optimization difficult for a gradient method, and F3 and F4 are highly multimodal but exhibit regularities in the placement of the local optima. Independent trials were conducted on each function. The initial populations comprised 50 parents, distributed uniformly over the continuous interval  $[-5, 5]$  (denoted  $U(-5,5)$ ) in each dimension. Each parent also was assigned from 1 to  $m$  strategy vectors, with  $m$  varied systematically over the integers  $\{1, 2, 3, 4, 5, 10, 20\}$ . That is, first a series of trials was conducted with  $m = 1$ , and then  $m = 2$ , and so forth. Elements in the parents' initial strategy vectors were distributed  $U(0,1)$ . One of strategy vectors was selected as the active vector uniformly at random. One offspring was created from each parent by:

- (i) Determining if the active vector should be changed. With probability  $p_s = 0.1$ , the active vector was changed to one of the other  $m - 1$  strategy vectors selected uniformly at random. (This was omitted in the case of  $m = 1$ .)
- (ii) Using equations (1) and (2) to generate the offspring's vector  $\mathbf{x}'$ , active  $\sigma'$ , with the remaining inactive vectors  $\sigma$  being copied without alteration from the parent. Note that no lower bound was imposed on the strategy parameter values.

Selection used the tournament formulation offered in Fogel (1995, p. 137) with 10 competitions per individual. The 50 individuals with the most wins in the tournament were saved as parents for the next generation. Evolution was halted after 200 generations and the score of the best individual in the population was recorded. Each of these parameters governing population size, number of competitions,  $p_s$ , or number generations was chosen because it represented a reasonable starting place for the investigation and Fogel et al. (2001) made no claim about the optimality of the choices.

Five hundred independent trials were conducted at each setting of  $m$ . The mean best score in the population after 200 generations was recorded. However, the mean can be affected by outliers and the distributions of statistics of extreme values (such as the best sample from a collection of samples) may not tend to follow a Gaussian distribution. To reduce the influence of potential outliers, Fogel et al. (2001) adopted the standard practice of trimming outliers (5 percent).

In an effort to verify functional relationships between  $m$  and the average best score obtained on each function, a second set of 2000 independent trials was again conducted at each setting of  $m$  on every function. Comparisons were made between the behavior observed over 500 trials and that observed over 2000 trials, as will be indicated. The results on functions F1-F4 are shown in Figures 1-4, respectively (from Fogel et al. 2001). In each case the figure in part (a) shows the mean trimmed best error (MBE) score over 500 trials and the prospective functional relationship between  $m$  and MBE. Part (b) then shows the data for MBE as a function of  $m$  over 2000 trials at each value of  $m$ . Across all four functions, the data obtained for 2000 trials validates the relationship observed in the independent set of 500 trials. The best (least square) linear or nonlinear regression models relating the data for each function are in close agreement and a large degree of the variability of the data is explained by each model. The results indicate that the MBE generally decreases with increasing  $m$ , but that the specific mathematical relationship between MBE and  $m$  varies by function.

The MBE is one important statistic in determining the effectiveness of an optimization algorithm, but it is inherently limited because it is a single point descriptor. A better sense of the benefit of increasing  $m$  can be obtained by examining the cumulative

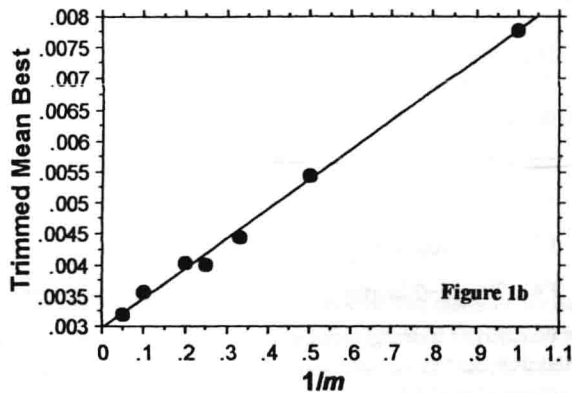
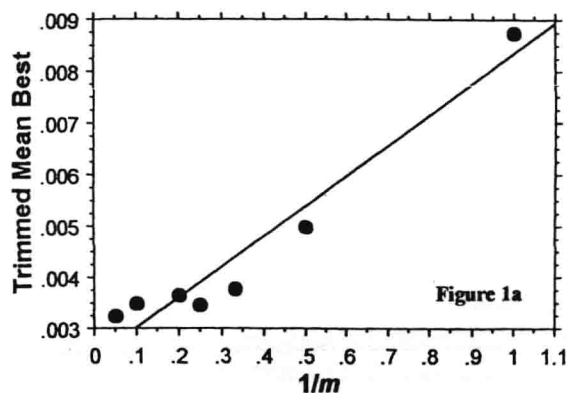


Figure 1. For the function F1, a good prediction of the mean trimmed best error (MBE) after trimming five percent of the outliers can be obtained by examining the inverse of the number of self-adaptive vectors used. (a) The data based on 500 trials at each setting of  $m$  with a five percent trim. An analysis of various transforms of the data revealed the relationship  $MBE = 0.002 + 0.006m^{-1}$ , which yielded an  $R^2$  value of 0.94. (b) The relationship was validated by 2000 additional trials. The new relationship was  $MBE = 0.003 + 0.005m^{-1}$ , which had an  $R^2$  value of 0.995. Thus there was good agreement between the coefficients of both models and the functional relationship explains a very large amount of the variation in the data.

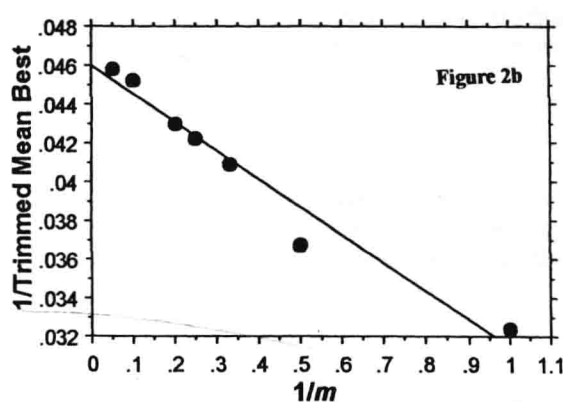
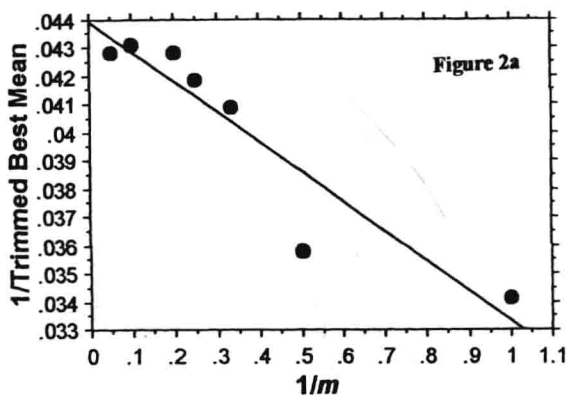
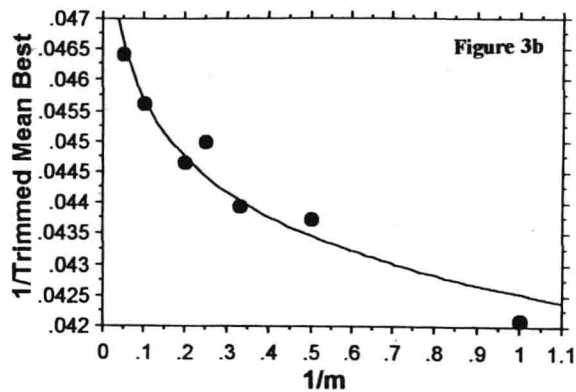
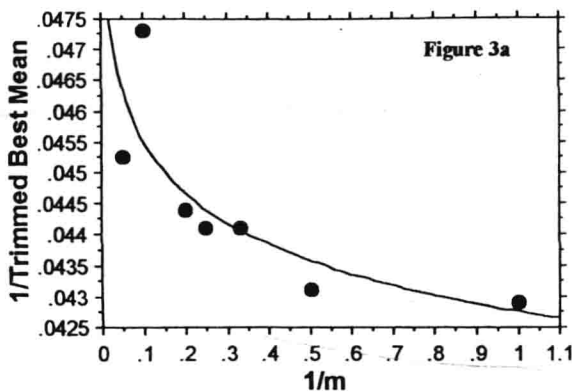
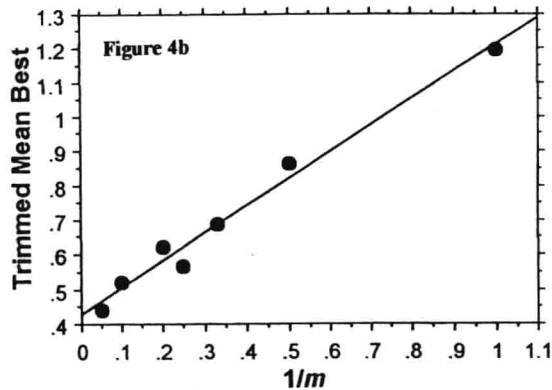
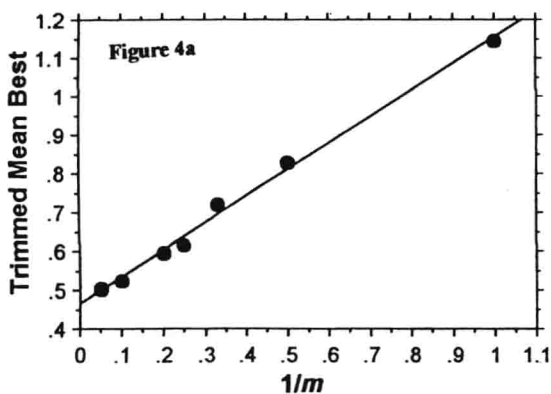


Figure 2. For the function F2, a good prediction of the inverse of the MBE after trimming five percent of the outliers can be obtained by examining the inverse of the number of self-adaptive vectors used. (a) The data based on 500 trials at each setting of  $m$  with a five percent trim. An analysis of various transforms of the data revealed the relationship  $MBE^{-1} = 0.044 - 0.011m^{-1}$ , which yielded an  $R^2$  value of 0.87. (b) The relationship was validated by 2000 additional trials. The new relationship was  $MBE^{-1} = 0.046 - 0.015m^{-1}$ , which had an  $R^2$  value of 0.96. There was good agreement between the coefficients of both models and the functional relationship explains a very large amount of the variation in the data.



**Figure 3.** For the function F3, a good prediction of the inverse of the MBE after trimming five percent of the outliers can be obtained by examining the natural logarithm of the inverse of the number of self-adaptive vectors used. (a) The data based on 500 trials at each setting of  $m$  with a five percent trim. An analysis of various transforms of the data revealed the relationship  $MBE^{-1} = 0.043 - 0.001 \ln(m^{-1})$ , which yielded an  $R^2$  value of 0.633. (b) The relationship was validated by 2000 additional trials. The new relationship was  $MBE^{-1} = 0.043 - 0.001 \ln(m^{-1})$ , which had an  $R^2$  value of 0.948. There was good agreement between the coefficients of both models and the functional relationship explains a very large amount of the variation in the data.



**Figure 4.** For the function F4, a good prediction of the MBE after trimming five percent of the outliers can be obtained by examining the inverse of the number of self-adaptive vectors used. (a) The data based on 500 trials at each setting of  $m$  with a five percent trim. An analysis of various transforms of the data revealed the relationship  $MBE = 0.43 + 0.783m^{-1}$ , which yielded an  $R^2$  value of 0.981. (b) The relationship was validated by 2000 additional trials. The new relationship was  $MBE = 0.465 + 0.69m^{-1}$ , which had an  $R^2$  value of 0.995. There was good agreement between the coefficients of both models and the functional relationship explains a very large amount of the variation in the data.

distribution function (CDF) of the best error score on each of the 2000 trials as a function of  $m$ . For example, Figure 5 shows these data for F1 for the case of 2000 trials with  $m \in \{1, 2, 3, 4, 5, 10, 20\}$  (taken from Fogel et al. 2001). The x-axis indicates the best value obtained and the y-axis indicates the percentile of that value. From Figure 5, when  $m = 1$ , approximately 60 percent of the 1900 trials (2000 less the 5 percent that were trimmed) achieved a best value that was less than or equal to 0.005. In contrast, for  $m = 20$ , approximately 80 percent of the trials achieved this same level of performance. There is a consistent improvement in the CDF with increasing  $m$ ; furthermore, the MBE is lower for increased  $m$  primarily because the search avoids poor solutions when using a greater number of self-adaptive parameters. Figure 6 shows similar results for function F4, where the local minima are reflected clearly in the CDF and entrapment in those minima is avoided to greater degree by increasing  $m$ . Fogel et al. (2001) provides the complete set of figures for all the functions studied.

### 3. CONCLUSIONS

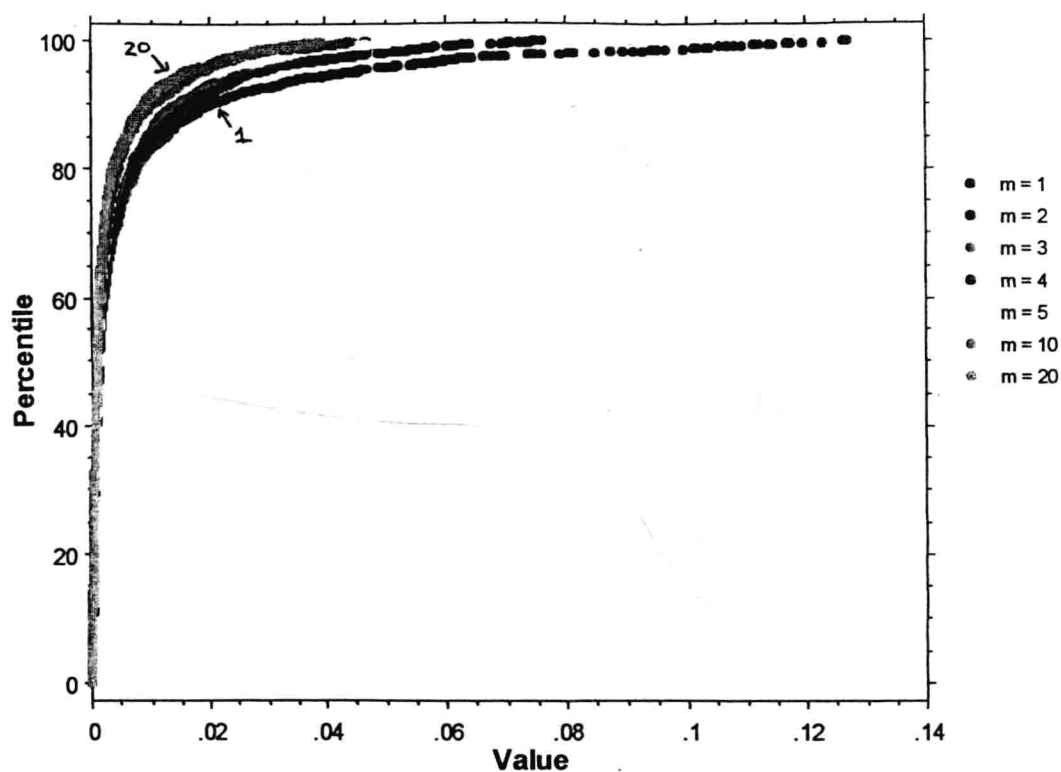
Just as the use of multiple parents (beyond two) in recombination may offer a more efficient optimization strategy in some problems (Fogel and Fogel, 2001), the use of multiple strategy parameter vectors may also offer advantages. Fogel et al. (2001) have shown that there is a consistent advantage for using more than one strategy vector when applied to four familiar test functions. A cost-benefit analysis can be undertaken to determine an appropriate number of strategy vectors to employ based on sample trials conducted on prospective new functions to examine. Figures 5-6 indicate that there may be significant overall performance differences in the range of  $m \in \{1, \dots, 5\}$  but less significant differences in the range  $m \in \{5, \dots, 20\}$ . The evidence suggests that the increased performance that arises from multiple strategy vectors stems most directly from a smaller probability of stalling at a poor solution, as opposed to finding superior solutions. This evidence is consistent with the perspective that the elements of the strategy parameter vector under the update equations (1) and (2) may decrease too rapidly, leading to stagnation. Switching between search strategies can overcome this problem. Future efforts should concern finding the optimal settings of  $p_s$ , as well as the potential for determining appropriate variations on this procedure on a problem-dependent basis and for problems of greater dimension.

### Acknowledgments

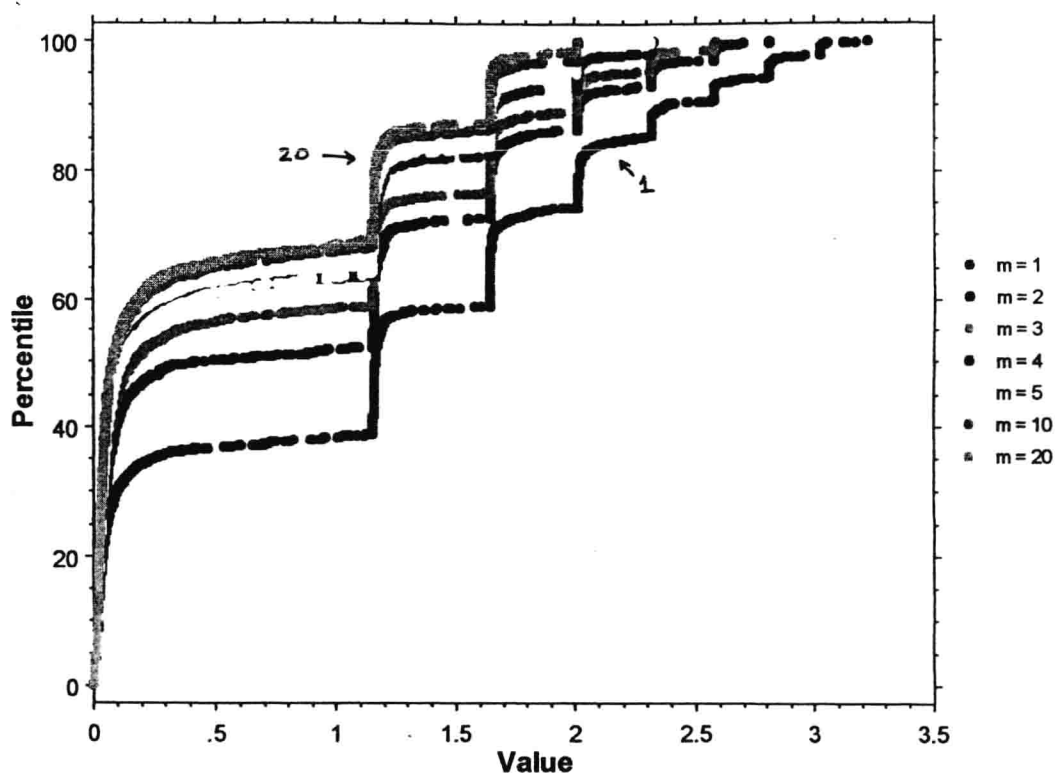
The author thanks Elsevier Science Publishers for rights to reprint portion of Fogel et al. (2001) [the author's previous publication from *BioSystems*] and thanks B. Bossachi for inviting a presentation to review this work.

### REFERENCES

- H.-G. Beyer and K. Deb (2001) "On self-adaptive features in real-parameter evolutionary algorithms," *IEEE Transactions on Evolutionary Computation*, Vol. 5:3, pp. 250-270.
- H.J. Bremermann, "The evolution of intelligence. The nervous system as a model of its environment," Technical report, no. 1, contract no. 477(17), Dept. Mathematics, Univ. Washington, Seattle, July, 1958.
- K. Chellapilla and D.B. Fogel (1999) "Fitness Distributions in Evolutionary Computation: Motivation and Examples in the Continuous Domain," *BioSystems*, Vol. 54:1-2, pp. 15-29.
- A.E. Eiben, R. Hinterding, and Z. Michalewicz, (1999) "Parameter control in evolutionary algorithms," *IEEE Transactions on Evolutionary Computation*, Vol. 3:2, pp. 124-141.
- D.B. Fogel (1995) *Evolutionary Computation: Toward a New Philosophy of Machine Intelligence*, IEEE Press, Piscataway, NJ.
- D.B. Fogel and G. B. Fogel (2001) "Revisiting Bremermann's genetic algorithm. II. Experiments with multiparent discrete recombination on linear systems of equations," *Applications and Science of Computational Intelligence IV*, K. Priddy, P. Keller, and P.J. Angeline (eds.), SPIE, Bellingham, WA.
- D.B. Fogel, G.B. Fogel, and K. Ohkura (2001) "Multiple-Vector Self-Adaptation in Evolutionary Algorithms," *BioSystems*, Vol. 61:2-3, pp. 155-162.
- D.B. Fogel, L.J. Fogel, and J.W. Atmar (1991) "Meta-evolutionary programming," *Proc. 25th Asilomar Conference on Signals, Systems, and Computers*, R.R. Chen (Ed.), Maple Press, pp. 540-545.



**Figure 5.** The cumulative distribution functions (CDFs) of the best value obtained in each of the 1900 trials (2000, less the 5 percent that were trimmed) for each setting of  $m$  on F1. The x-axis indicates the selected value and the y-axis indicates the percentile of that value. A percentile of 60 indicates that 60 percent of the trials achieved a value that was less than or equal to the selected value. The graphs indicate a general trend of improved performance as  $m$  increases. Moreover, the improvement comes mostly from avoiding poor solutions.



**Figure 6.** The CDFs of the best value obtained in each of the 1900 trials (2000, less the 5 percent that were trimmed) for each setting of  $m$  on F4. There is a general trend of improved performance as  $m$  increases, and a greatly reduced chance of stagnating in poor solutions for large  $m$ . F4 has local minima and these are reflected clearly in the CDFs.



- D.B. Fogel, L.J. Fogel, W. Atmar, and G.B. Fogel (1992) "Hierarchical models of evolutionary programming," *Proceedings of the First Annual Conference on Evolutionary Programming*, D.B. Fogel and W. Atmar (eds.), Evolutionary Programming Society, La Jolla, CA, pp. 175-182.
- D.B. Fogel and A. Ghoseil (1997) "A note on representations and variation operators," *IEEE Transactions on Evolutionary Computation*, Vol. 1:2, pp. 159-161.
- D.K. Gehlhaar and D.B. Fogel (1996) "Tuning evolutionary programming for conformationally flexible molecular docking," *Evolutionary Programming V: Proceedings of the Fifth Annual Conference on Evolutionary Programming*, L.J. Fogel, P.J. Angeline, and T. Bäck (eds.), MIT Press, Cambridge, MA, pp. 419-429.
- F. Herrera and M. Lozano (2001) "Adaptive genetic operators based on coevolution with fuzzy behaviors," *IEEE Transactions on Evolutionary Computation*, Vol. 5:2, pp. 129-148.
- J.H. Holland (1975) *Adaptation in Natural and Artificial Systems*, Univ. Michigan Press, Ann Arbor, MI.
- M. Köppen, D.H. Wolpert, and W.G. Macready (2001) "Remarks on a recent paper on the 'no free lunch' theorems," *IEEE Transactions on Evolutionary Computation*, Vol. 5:3, pp. 295-296.
- W.G. Macready and D.H. Wolpert (1998) "Bandit problems and the exploration/exploitation tradeoff," *IEEE Trans. Evol. Comp.*, Vol. 2:1, 2-22.
- I. Rechenberg (1973) *Evolutionstrategie: Optimierung technischer Systeme nach Prinzipien der biologischen Evolution*, Fromman-Holzboog, Stuttgart, Germany.
- I. Rechenberg (1994) personal communication, Technical University of Berlin.
- J. Reed, R. Toombs, and N.A. Barricelli (1967) "Simulation of biological evolution and machine learning. I. Selection of self-reproducing numeric patterns by data processing machines," *Journal of Theoretical Biology*, Vol. 17, pp. 319-342.
- R. Rosenberg (1967) "Simulation of genetic populations with biochemical properties," Ph.D. Dissertation, University of Michigan, Ann Arbor, MI.
- G. Rudolph (2001) "Self-adaptive mutations may lead to premature convergence," *IEEE Transactions on Evolutionary Computation*, Vol. 5:4, pp. 410-414.
- N. Saravanan and D.B. Fogel (1994) "Learning strategy parameters in evolutionary programming: An empirical study," *Proceedings of the Third Annual Conference on Evolutionary Programming*, A.V. Sebald and L.J. Fogel (eds.), World Scientific Press, River Edge, NJ, pp. 269-280.
- N. Saravanan, D.B. Fogel, and K.M. Nelson (1995) "A comparison of methods for self-adaptation in evolutionary algorithms," *BioSystems*, Vol. 36, pp. 157-166.
- H.-P. Schwefel (1981) *Numerical Optimization of Computer Models*, John Wiley, Chichester, U.K.
- H.-P. Schwefel (1995) *Evolution and Optimum Seeking*, John Wiley, NY.
- D.H. Wolpert and W.G. Macready (1997) "No free lunch theorems for optimization," *IEEE Transactions on Evolutionary Computation*, Vol. 1:1, pp. 67-82.
- X. Yao, Y. Liu, and G. Lin (1999) "Evolutionary programming made faster," *IEEE Transactions on Evolutionary Computation*, Vol. 3:2, pp. 82-102.

\*-David B. Fogel, Natural Selection, Inc., 3333 N. Torrey Pines Ct., Suite 200, La Jolla, CA 92037, tel: (858) 455-6449, fax: (858) 455-1560, [dfogel@natural-selection.com](mailto:dfogel@natural-selection.com), [www.natural-selection.com](http://www.natural-selection.com)