Evolutionary Optimization Algorithms
Errata

Dan Simon

November 22, 2016

This document contains a list of errors in the book *Evolutionary Optimization Algorithms*, John Wiley & Sons, 2013. The main web site for the book is `http://academic.csuohio.edu/simond/EvolutionaryOptimization`. My email address is listed on my home page at `http://academic.csuohio.edu/simond`. I enthusiastically welcome feedback, comments, suggestions for improvements, and corrections. I also gratefully acknowledge the following individuals for pointing out the errata that are documented here: David vun Kannon, Rick Rarick, Armin Rashvand, George Thomas, Hans-Paul Schwefel, Simon Polstra, Ibrahim Venkat.

1. The first sentence of Section 2.6.2 on page 26 says that Example 2.7 compared a binary and continuous GA, but it actually compared four hill climbing algorithms.

2. The second-to-last sentence before Example 3.4 on page 57 should read "... *the* distribution ..." instead of "... *he* distribution ..."

3. In the first sentence of Section 4.4 on page 76, "citeDavis93" should be [Davis, 1993].

4. In Section 5.1 on the line just before Equation (5.1), $i \in [0, N]$ should instead be $i \in [1, N]$.

5. Equation (5.7) and Figure 5.7 are wrong. Here is an FSM that works for Example 5.2, and that should replace Equation (5.7):

$$S = \left[\begin{array}{cccc} 1 & 4 & 0 & 2, \\ 0 & 2 & 1 & 3, \\ 1 & 2 & 1 & 4, \\ 0 & 4 & 0 & 1 \end{array} \right].$$

   And Figure 5.7 should be changed accordingly.

6. Hans-Paul Schwefel, one of the fathers of ES, was kind enough to email me his valuable and interesting review of Chapter 6. See the appendix of this document for his review, which should be read by anyone interested in ES research.

7. In Section 6.5, on page 132, the 1st sentence of the 3rd paragraph should read, "The algorithm of Figure 6.13 *assumes* that the ..."

8. In Equation (8.10) on page 196, r should be written as

$$r = \frac{1}{2}\sqrt{\sum_{k=1}^{n}\left(\max_i x_i(k) - \min_i x_i(k)\right)^2}.$$

9. On page 212 on the 4th line of Section 8.8.7, the word "rearrangement" is misspelled.

10. On page 269 on the 2nd line of the 2nd-to-last paragraph, "call" should be "called."

11. on page 270 on the 2nd line, "the*ring* ..." should be "the *ring* ... ."

12. In Equation (14.5) on page 355, P should be n.

13. The first line of Equation (14.7) on page 356 should be

$$x(A') = \{0, -2/n, -4/n, \cdots, -2\}.$$

14. Equation (20.30) on page 529 should have $f_2(x^*) = 8w_1^2$.

15. The last sentence of the caption of Figure 21.2 on page 567 says, "... $\hat{f}(x_2) = f(y_2)$ and $\hat{f}(x_3) = f(y_3)$," but it should say, "... $\hat{f}(x_2) = f(y_1)$ and $\hat{f}(x_3) = f(y_2)$."

16. In the first paragraph on page 575, in the last sentence, the word "have" should be deleted.

17. On page 570 we want to maximize Equation (21.12). On equivalent but numerically better-conditioned problem is to maximize the log of Equation (21.12). Substituting Equation (21.16) into Equation (21.12) and removing constants gives the minimization objective function $\sigma^M|R|^{1/2}$. The log of this quantity is

$$\begin{aligned}
\log\left(\sigma^M|R|^{1/2}\right) &= \log\sigma^M + \log|R|^{1/2} \\
&= M\log\sigma + (1/2)\log\prod_{i=1}^{M}\lambda_i \\
&= M\log\sigma + (1/2)\sum_{i=1}^{M}\log\lambda_i
\end{aligned}$$

which is a much better-conditioned objective function than $\sigma^M|R|^{1/2}$.

18. The phrase "that those" should instead by "than those" on the 6th line of page 633 in Section B.2.4.

19. The word "between" is misspelled in the second-to-last line on page 636 in Section B.2.5.

20. In the Ackley function description in Section C.1.2 on page 643, the term $\sum_{i=1}^{n} x_i^2/n$ should be inside a square root operation.

21. The last sentence of Section C.1.18 on page 654 ends with "citeBersini" – it needs to be corrected to point to some reference, but I can't remember which reference – I'll have to find it later.

22. Page 683 in Section C.7.2 just before Equation (C.127) says that an orthogonal matrix is a matrix whose "transpose is equal to its determinant" – but it should be "transpose is equal to its inverse."

23. In the index, the first reference to "roulette-wheel selection" is listed as page 199, but there is also a description on page 46.

Appendix: Hans-Paul Schwefel's Review of Chapter 6

Aside of many good things to be said about your work, I feel obliged to hint to some inaccuracies and one unfortunately misleading self adaption experiment with a (10,20)-ES.

Firstly, some comments to the ES history: To my knowledge, the first publication of Ingo Rechenberg was an English translation of a hand written manuscript of a talk at the Joint DGRR/WGLR Annual Conference in Sept. 1964, the translation being done by the Royal Aircraft Establishment and first printed in August 1965. My diploma thesis (comparable to a Master thesis) was completed in March 1965 and contained already computer experiments on a Zuse Z23 machine with a (1+1)-ES and discrete variables on a 2D parabolic ridge (see: `http://Ls11-www.cs.uni-dortmund.de/people/schwefel/publications/Schw65.pdf`). Since the ES got stuck on the ridge sometimes (premature convergence, a term I don't like because convergence should mean that the goal is reached and if that happens prematurely that would be even better), we later turned to real variables, and all theory was done in that domain at the time. I had to earn money and got a job in industry (AEG) - Hence: The flashing nozzle experiments with hot water (one-component two-phase supersonic flow behind the throat) were done by me in industry in 1967/1968 and published in 1970 in the Caltech Proc. Eleventh Symp. Engineering Aspects of Magnetohydrodynamics. Peter Bienert never got a doctorate, but in 1967 a diploma degree (as Ingo Rechenberg and me together in 1965). There are some people who do not like to mention my early work in the field.

Secondly, your (10,20)-ES experiment with self-adapting mutation strength could not have been successful! You report yourself before, that the optimal strength corresponds with a success rate of about 1/5. Hence, even with already optimal sigma(s) only about 20/5=4 descendants lead to improvements; The rest, i.e., the majority of the 10 best, namely 6, will be worse than their parents. The strategy MUST diverge. To work properly, the birth surplus lambda/mu should be at least as large as the inverse of the success rate. Only better descendants can inherit suitable mutation strengths.

I always operated with a (10,100)-ES and pleaded for parallel execution of all simulations of the descendants. Trying to use an elitist (plus) ES for dynamic landscapes with moving optima, as some people have done, is also prone to fail. And, by the way, the factor tau becomes too dominant against tau prime when the number of dimensions gets larger - if you use my 1975 suggestions. At that time, by the way, many people including my colleagues found it ridiculous to use lambda > 1 and a non-elitist selection. Even now some people prefer a (1,8)-ES making use of only one best survivor. For self-adaptation to work, however, diversity of the parents is compulsory. You may like to read my recent publication in the ACM Ubiquity Symposium on EC, see: `http://ubiquity.acm.org/article.cfm?id=2369297`.

Kind regards,
Hans-Paul Schwefel
June 11, 2013