Errata and Clarifications for First Printing of
Introduction to Stochastic Search and Optimization:
Estimation, Simulation, and Control
by J. C. Spall

Please provide corrections or suggestions to **James C. Spall**. To determine if a copy of the book is the first printing check the last digit at the bottom of page iv near the front.

CHAPTER 4

p. 107, lines 7 – 9 of Condition B.2: $\delta(\eta)$ and δ_0 should be constrained to be strictly positive. Hence the relevant statements should read: "....there exists a $\delta(\eta) > 0$ such that...." and "....there exists a $\delta_0 > 0$ such that...."

p. 113, line 10: "…see Section 12.3 for a discussion…" should read "…see Section 13.3 for a discussion…"

CHAPTER 5

Exercise 5.5: Q appearing on line 3 of the exercise statement should be a function of both $\boldsymbol{\theta}$ and V (not just a function of $\boldsymbol{\theta}$).

CHAPTER 6

p. 169: Because each iteration of algorithm B uses 20 loss measurements (rather than one as usual), the last sentence of p. 169 ("Because the horizontal axis….") should be replaced by: "As indicated in Exercise 6.11, algorithm B has 20 measurements being averaged for each loss evaluation. Hence, the 1000 measurements translates into $1000/(2\times 10) = 50$ iterations for FDSA and 49 iterations for algorithm B (including the initial loss evaluation)."

p. 169, line 9: The ratio of maximum eigenvalue to minimum eigenvalue is approximately 175, not 65.

CHAPTER 7

p. 196 (second line from bottom): The reference to Chapter 6 should be a reference to Chapter 5.

p. 199: In the second line of the displayed equation near the bottom of the page, a subscript j in $H_{\ell j}$ should be replaced with i according to:

$$= H_{ij}(\hat{\boldsymbol{\theta}}_k) + \sum_{\ell \neq j} H_{i\ell}(\hat{\boldsymbol{\theta}}_k)\frac{\Delta_{k\ell}}{\Delta_{kj}} + O(c_k^2)$$

(the *ij*th element of $\boldsymbol{H}$ is H_{ij}). Note that for consistency with the term before the summation, the above also interchanges the order of the subscripts—i.e., $H_{i\ell}$ instead of $H_{\ell i}$. This has no effect on the result because of the symmetry of $\boldsymbol{H}$.

CHAPTER 8

pp. 214–215 (clarification): The Metropolis criterion should also be modified when introducing τ to account for noisy loss measurements. In particular, $\exp[-\delta/(c_b T)]$ in step 2 of the SAN algorithm (p. 212) is replaced by $\exp[-(\delta-\tau)/(c_b T)]$.

Exercise 8.12 (clarification): With only 3000 measurements (as stated), it is expected that very few (< 5) of the replications will satisfy the stated convergence criteria. The second printing considers 100,000 measurements, leading to a much higher proportion of successes (although still less than 100 percent).

CHAPTER 9

p. 239, lines 10–11: The statement "…the number of possible representations for a string of length *b* bits." should be replaced by "…the maximum value in a standard nonnegative integer representation for a string of length *b* bits."

p. 239, eqn. (9.1): With the appropriate interpretation of $[a_1\ a_2\ \ldots\ a_b]$, eqn. (9.1) as stated in the text will produce a correct decoding. However, to be consistent with the decoding process in Example 9.1, it is useful to reverse the order of the summation in (9.1) according to

$$\theta = \theta_{\min} + \frac{\theta_{\max} - \theta_{\min}}{2^b - 1} \sum_{i=1}^{b} a_i 2^{b-i} \quad . \tag{9.1}$$

p. 251, line 7 of Example 9.4: The ratio of maximum eigenvalue to minimum eigenvalue is approximately 175, not 65 (same correction appears in Chapter 6, as mentioned above).

Exercise 9.6 (clarification): The variable *ident* accumulates the number of parents that are identical in the selection step. Note that when two parents are not identical, the crossover probability provides an *upper bound* to the probability of producing two offspring different from the parents. (For example, one-point crossover applied as indicated to the following two distinct chromosomes does not change either chromosome: [0 1 0|1 1 0] and [1 1 0|1 1 0].)

Exercise 9.7: Problem should ask for a *lower bound* to the probability of the chromosome passing intact to the next generation. The indicated solution on p. 554 of the "Answers to Selected Exercises" section provides this lower bound. (See **Solutions for Selected Exercises** for an explanation for this change.)

CHAPTER 10

p. 270 (paragraphs 2 and 3): Several modifications to the discussion are required to accommodate the fact that the Markov chain is *not* irreducible (because a GA with elitism cannot

move to a population whose best fitness value is lower than the current best value). However, the key conclusion discussed in the current text is unchanged: namely, the chain *does* have a unique limiting value $\bar{\boldsymbol{p}}^T$ with $\sum_{i \in J} \bar{p}_i = 1$ because the GA is always saving the best chromosome encountered.

CHAPTER 12

p. 320 (Figure 12.2): Figure legend should refer to Sects. 12.4 and 12.2, not Sects. 3.4 and 3.2.

CHAPTER 13

p. 332: The third line of eqn. (13.1) should have a minus (not plus) sign at the beginning. Hence the line should read: $-2\left[h(\boldsymbol{\theta}, \boldsymbol{x}) - E(z \mid \boldsymbol{x})\right] E\left[(z - E(z \mid \boldsymbol{x})) \middle| \boldsymbol{x}\right]$. This change does not affect the last (fourth) line of the equation.

p. 344: In the fourth line of the second full paragraph, "... n/n_T values for MSE(m)..." should be replaced by "...multiple values for MSE(m)..."

p. 353: The first integral in the second line of the equation at the top of the page should read

$$\int \frac{\partial^2 \log \ell}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \ell \, d\boldsymbol{\zeta}, \text{ not } \int \frac{\partial \log^2 \ell}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \ell \, d\boldsymbol{\zeta}.$$

CHAPTER 14

Exercise 14.13 (clarification): The second sentence should reflect that the diagonals in the covariance matrix are unity (as indicated by the first sentence where the reader is instructed to build on the setting of Exercise 14.12) while the off-diagonals are 0.5. Hence, the second sentence should read: "In the simulated indifference zone selection process, assume that $\text{cov}[y_k(\boldsymbol{\theta}_i), y_k(\boldsymbol{\theta}_j)] = 0.5$ for all k and $i \neq j$."

CHAPTER 15

pp. 429–430 (clarification to Example 15.7): Note that $Q(\theta, V)$ is a sum of a part dependent on V and a part independent of V (i.e., $Q(\theta, V) = V + b/\theta$). In this and other such cases, it is only necessary to apply the sample path averaging based on $V_0, V_1, \ldots, V_{N-1}$ to the part dependent on V. Hence, (15.16) can be re-expressed as

$$\bar{L}_N(\theta) = \frac{1}{N} \sum_{i=0}^{N-1} V_i \frac{p_V(V_i \mid \theta)}{p_V(V_i \mid \theta')} + \frac{b}{\theta},$$

from which the expression for $\partial \bar{L}_N(\theta)/\partial \theta$ in (15.17) directly follows.

CHAPTER 16

p. 440: The displayed integral should be set equal to 1 as in: $\int q(\boldsymbol{w} \mid \boldsymbol{X} = \boldsymbol{x})\, d\boldsymbol{w} = 1$. Further, in the fourth and fifth lines below this displayed integral, $f(\boldsymbol{x})$ should be $p(\boldsymbol{x})$ (two instances) and the associated $\mathbf{0}$ should be 0 (nonbold) since $p(\boldsymbol{x})$ is a scalar function. Hence, the relevant part of the lines should read: "...of points where $p(\boldsymbol{x}) \neq 0$, should be a *superset* of the set of points $\boldsymbol{x}$ where $p(\boldsymbol{x}) \neq 0$."

CHAPTER 17

pages 492 (second paragraph from bottom) and 496 (last sentence before Example 17.11 and first sentence of Example 17.11): References to Examples 17.8 and 17.9 should be to Examples 17.9 and 17.10.

APPENDIX A

Matrix relationship (xxi) (p. 512): Reference to Example A.4 should be to Example A.5.

APPENDIX B

pp. 516–517 (clarification): The basic one-sample results of Section B.1 are not restricted to only the $\mu = 0$ case (although this is the case of interest when applying these results to the two-sample setting of interest in Section B.2).